https://doi.org/10.26219/heal.aueb.5635

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

SCHOOL OF INFORMATION SCIENCES & TECHNOLOGY

DEPARTMENT OF STATISTICS

POSTGRADUATE PROGRAM

Searching Causation in earthquake time series datasets

By
SOFIA MARIA KARADIMITRIOU

A Thesis
Submitted to the Department of Statistics
of the Athens University of Economics and Business
in partial fulfilment of the requirements for
the degree of Master of Science in Statistics

Athens, Greece
May 2021

https://doi.org/10.26219/heal.aueb.5635

https://doi.org/10.26219/heal.aueb.5635

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

ΣΧΟΛΗ ΕΠΙΣΤΗΜΩΝ & ΤΕΧΝΟΛΟΓΙΑΣ ΤΗΣ ΠΛΗΡΟΦΟΡΙΑΣ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ

ΜΕΤΑΠΤΥΧΙΑΚΟ

Διερευνώντας αιτιότητα σε σεισμολογικά δεδομένα χρονολογικών σειρών

ΣΟΦΙΑ ΜΑΡΙΑ ΚΑΡΑΔΗΜΗΤΡΙΟΥ

ΔΙΑΤΡΙΒΗ
Που Υποβλήθηκε στο Τμήμα Στατιστικής
του Οικονομικού Πανεπιστημίου Αθηνών
ως μέρος των απαιτήσεων για την απόκτηση
Μεταπτυχιακού Διπλώματος Ειδίκευσης στη Στατιστική

Αθήνα
Μάϊος 2021

https://doi.org/10.26219/heal.aueb.5635

https://doi.org/10.26219/heal.aueb.5635

ACKNOWLEDGEMENTS

First, I would first like to thank my supervisor Dimitris Karlis for his guidance and patience. His understanding and strong spirit made me pursue a career in research. Whenever I was in need for questions or clarifications, Prof. Karlis would immediately set up a meeting for me. Furthermore, he was patiently guiding me while I was combining both my thesis as well as my work. Without his contribution and advices this MSc Thesis would have never been fulfilled. He is an academic and a person for whom I have an outmost respect.

I am also grateful to my family. More specifically my parents, Dimitris and Kassiani, for providing me with support throught both my undergraduate and master degree. Nonetheless I thank my brother, Nikos, for his support in my studies and life as a wonderful friend. In addition, I would like to express my gratitude to my classmates from my master degree, Michalis and Sotiris for being the best people accompanying me until the end of our master degree.

https://doi.org/10.26219/heal.aueb.5635

https://doi.org/10.26219/heal.aueb.5635

VITA

I was born on October 5th, 1990 in Athens, where I got my High School Diploma. In 2008 I entered the department of Statistics at Athens University of Economics and Business. I graduated in 2013 and I enrolled in the Master of Science in Statistics at Athens University of Economics and Business at the same year, while then I continued my academic journey by obtaining a PhD in Statistics and Probability in the School of Mathematics and Statistics, at the University of Sheffield. I then moved on for my post doctorate at the UCL Cancer Institute in Bioinformatics. Currently, I am working as a Senior Data Scientist.

https://doi.org/10.26219/heal.aueb.5635

https://doi.org/10.26219/heal.aueb.5635

ABSTRACT

Sofia Maria Karadimitriou

Searching Causation in earthquake time series datasets

December 2017

Count time series refer to data that can be represented as counts of events observed across time. In certain cases we observe more than one such series and in this case, apart from the time correlation of each series, we are interested in modelling the cross-correlation between the series. This implies that one series may influence the outcome of the other in some lagged time interval.

In many circumstances, we are interested in creating multivariate modelling approaches which explore causation between the different series. These can be for instance, in finance, the transactions between many stocks, where the transactions of one stock yesterday influence the volume of transactions of another stock today. Or, in epidemiology, the number of occurences of specific diseases in some time period influence some other disease after some period.

In the present thesis we care about a problem in seismology, where the number of seismic events in a location/area may be triggered by some earthquake in other area at some time period at the past.

Multivariate models of such nature are very limited and the analysis is difficult to come through due to high complexity of calculations or efficiency from a computational aspect.

In the present thesis, we will study a trivariate model under a basic autoregressive structure of order one. Additionally, we will derive the theoretical results for estimating the parameters of interest using maximum likelihood method estimation. Furthermore, we will provide an analysis of earthquake

https://doi.org/10.26219/heal.aueb.5635

count time series between three spatial lattices or else sub-tectonic plates in the North Aegean Sea in order to investigate the spatial causation in those seismic events. The central idea is to identify whether earthquakes in some area may trigger earthquakes in other area which is an important problem in statistical seismology. For this purpose we built a model with all possible structure and then using statistical tools we try to identify which terms are really useful.

https://doi.org/10.26219/heal.aueb.5635

https://doi.org/10.26219/heal.aueb.5635

ΠΕΡΙΛΗΨΗ

Σοφία Μαρία Καραδημητρίου

Διερευνώντας αιτιότητα σε σεισμολογικά δεδομένα χρονολογικών σειρών

Δεκέμβριος 2017

Οι χρονολογικές σειρές δεδομένων αριθμού γεγονότων είναι φαινόμενα τα οποία μπορούν να αναπαρασταθούν ως διακριτές μετρήσεις συμβάντων τα οποία εξελίσσονται στο χρόνο. Σε συγκεκριμένες περιπτώσεις παρατηρούμε περισσότερες απο μία σειρές και συγκεκριμένα, σε αυτή την περίπτωση, μας ενδιαφέρουν οι συσχετίσεις καθώς και οι επι μέρους συσχετίσεις μεταξύ των σειρών. Αυτό σημαίνει ότι μια σειρά μπορεί να επηρεάζει το αποτέλεσμα μιας άλλης σειράς με βάση διαστήματα προηγούμενων χρονικών τιμών.

Σε πολλές περιπτώσεις μας ενδιαφέρει η κατασκευή πολυμεταβλητών μοντέλων τα οποία εξηγούν την αιτιότητα μεταξύ διαφορετικών σειρών. Παραδείγματος χάριν, στα χρηματοοικονομικά θα μας ενδιέφερε η συσχέτιση ή επιρροή μεταξύ των συναλλαγών μιας μετοχής που συνέβη την προηγούμενη μέρα και στον όγκο συναλλαγών μιας άλλης την επόμενη. Ή στην επιδημιολογία, ο αριθμός κρουσμάτων συγκεκριμένων ασθενειών για κάποιο χρονικό διάστημα να επιρρεάζει τα κρούσματα σε μια άλλη ασθένεια μετά απο κάποιο χρονικό διάστημα.

Στη συγκεκριμένη διατριβή μας ενδιαφέρει ένα πρόβλημα στη σεισμολογία όπου ο αριθμός σεισμικών γεγονότων σε μια τοποθεσία/περιοχή μπορεί να προκληθεί από κάποιον σεισμό που συνέβη σε κάποιο χρονικό διάστημα σε μια άλλη περιοχή στο παρελθόν.

Πολυμεταβλητά μοντέλα τέτοιας φύσεως είναι ελάχιστα και η ανάλυση είναι δύσκολο να επιτευχθεί λόγω περιπλοκότητας από μαθηματικής απόψεως ή ακόμα και λόγω μη αποδοτικότητας από υπολογιστικής απόψεως.

https://doi.org/10.26219/heal.aueb.5635

Στη συγκεκριμένη διατριβή, θα διερευνήσουμε ένα είδος τριμεταβλητού μοντέλου κάτω από μια βασική αυτοπαλινδρομούμενη δομή τάξης ένα. Επιπλέον, θα παράγουμε θεωρητικά αποτελέσματα για την εκτίμηση των παραμέτων που ενδιαφερόμαστε χρησιμοποιόντας τη μέθοδο εκτίμησς μέγιστης πιθανοφάνειας. Επιπροσθέτως, θα προσφέρουμε μια στατιστική ανάλυση σε σεισμολογικές χρονικές σειρές αριθμού γεγονότων μεταξύ τριών περιοχών ή αλλιώς υποτεκτονικών πλακών στην περιοχή του Βορείου Αιγαίου με σκοπό την έρευεση χωρικής αιτιότητας μεταξύ αυτών των σεισμικών γεγονότων. Η κεντρική ιδέα είναι να εξακριβώσουμε εάν οι σεισμοί σε μία περιοχή θέτουν σε ενέγεια σεισμούς σε άλλη περιοχή, πρόβλημα το οποίο είναι πολύ σημαντικό στη στατιστική σεισμολογία. Για αυτό τον σκοπό κατασκευάζουμε ένα μοντέλο με όλες τις δυνατές δομές και συσχετίσεις και έπειτα με τη χρήση στατιστικών εργαλείων βρίσκουμε ποιοι όροι είναι όντως χρήσιμοι.

https://doi.org/10.26219/heal.aueb.5635

Contents

https://doi.org/10.26219/heal.aueb.5635

List of Tables

https://doi.org/10.26219/heal.aueb.5635

List of Figures

https://doi.org/10.26219/heal.aueb.5635

https://doi.org/10.26219/heal.aueb.5635

Chapter 1

Introduction

In this chapter, the field of count time series process is introduced and motivation for applying it in the area of seismology is described. The source of the data application is given and some technicalities of applying appropriate statistical techniques are discussed.

1.1 Background and Motivation

The area of analysing count time series provides techniques to enable analysis of count data that vary in time. This can be for various purposes either to detect recognisable or meaningful patterns of variation or to predict future values of one (or more) variable though the time dependence. It aims to incorporate methods and insights from the simple time-series analysis into a more complex modelling framework and therefore it is a rich area of analysis that is evolving through the years. These techniques have been applied in many fields, including financial markets for the daily number of stock transactions, monthly number of incidences of a certain disease, number of customers waiting to be served at a counter at discrete time points or number of seismic events.

https://doi.org/10.26219/heal.aueb.5635

An insightful example is in syndromic surveillance systems where the number of patients with a given symptom is recorded with the intention of detecting early an unanticipated change on this number, perhaps indicating for a threat in public health. In practice a large number of symptoms are counted creating multiple time series that are in fact correlated. Correct evaluation of such multiple series need a model that can take into account the correlation across time but at the same time the cross correlation between the different symptoms.

The context of the application within this dissertation comes from geophysical research when the number of earthquakes need to be modelled (Boudreault and Charpentier, 2011). In such data the number of earthquakes above a certain magnitude threshold and for a given time period are counted. Different series can be generated from adjacent areas, making an important scientific question the correlation between the two areas.

The common element of all the above examples is that in all cases the collected data are counts observed in different time points, while at each time point they are correlated. Hence we have two sources of correlation, serial correlation— since the data are time series— and cross correlation—since they are correlated for a given time point. More precisely, the need to account for both serial and cross correlation complicates model specification, estimation and inference.

Nonetheless, literature on multivariate time series of counts is less developed. One of the reasons is that even models not being time series are less developed due to analytical and computational problems. However, in recent years, there have been some new models to facilitate modelling approach. Additionally, even ignoring the time series correlation one can see that there are not many models for multivariate counts. Inference for multi-

https://doi.org/10.26219/heal.aueb.5635

variate counts is analytically and computationally demanding. Perhaps the case is easier and more developed when dimensions reduce to two but there are several bivariate models that cannot generalize easy to multivariate ones. This obstacles the development of flexible models to be used also in time series context.

The first attempt of introducing counted time series models was made by Jacobs and Lewis (1978) which were constructed by a simple mixture of variables and their form is similar to that of the ARMA models in the Gaussian case. However, these models are the predecessor of the Integer Valued Autoregressive models which are explained thoroughly in Chapter 2 and are based on the binomial thinning operation. The fitting of these models for a univariate series is introduced by McKenzie (1985) and Al-Osh and Alzaid (1987). Furthermore, more models that are built in this setting can be found in Latour (1997); Brannas and Nordstrom (2000); Heinen and Rengifo (2007) and Quoreshi (2006). Forecasting models for discrete-valued time series are presented in Freeland and McCabe (2004) which involve the calculation of the $k-step$ ahead forecast distribution while Jung and Tremayne (2006b) extend it by allowing a second-order Markovian dependence. Integer forecasting is also considered in Pavlopoulos and Karlis (2006). Hidden Markov models Poisson models have been used by MacDonald and Zucchini (1997), Cooper and Lipsitch (2004) and Orfanogiannaki et al. (2010). Pedeli and Karlis (2013b) introduce the bivariate version of integer valued autoregressive models with either Poisson or negative binomial innovations while they also introduce a multivariate version of this latter model. Finally, Bayesian methodology for low counts has been applied by McCabe and Martin (2005) while also a Dynamic modelling of count series via using Kalman Filtering has been applied by Brandt et al. (2000).

A newer methodology to built multivariate models is to apply copula ap-

https://doi.org/10.26219/heal.aueb.5635

proach. Copulas (see Nelsen, 2006) have found a remarkable large number of applications in finance, hydrology, biostatistics etc., since they allow the derivation and application of flexible multivariate models with given marginal distributions. A plethora of models can be derived since the marginal properties can be separated from the association properties. For the case of discrete data, copulabased modelling is less developed. For example Genest and Nešlehová (2007) provided an excellent review on the topic. Since then there are several attempts to apply copulas to discrete data with quite useful success in practice (Nikoloulopoulos and Karlis, 2009). However, some of the desirable properties of copulas are not valid when dealing with count data, as for example dependence properties which are now dependent on the marginal properties. Furthermore, the calculation of the probability mass function has drawbacks in larger dimensions.

1.2 Software

The main softwares used for this dissertation was R [R Development Core Team]. All computing was accomplished at the laboratory of the Department of Statistics, AUEB.
Despite using efficient techniques in order to minimise the complexity of the likelihood distribution, a significant number of separate R programs, each with many lines of code, were written in order to carry out the analyses in this dissertation. It is possible that coding errors still remain.

1.3 About the Application

Understanding the earthquake activity in highly seismic areas has been a long standing aim in two areas, seismology and hazard and risk assessment. Regarding the seismic safety management, the Greek Government and local authorities aim to manage to keep up with predicting seismic activities since

the whole region of Greece and the shores of Turkey is one the riskiest regions for earthquakes (Aegean tectonic plate). Papazachos (1990) was the first to notice the high seismic activity of Aegean and the surrounding area by noting a realationship between this activity and the geomorphical features of Aegean and the surrounding areas. Papadimitriou and Sykes (2001) model the earthquake counts by using a deterministic model which varies in time and space (Deng and Sykes, 1997). Efforts of earthquake prediction have been done by Varotsos and Lazaridou (1991) which are based on electric signals.
Additionally, within insurance companies and the risk assessment of compensation in case of an earthquake, a key aim is to assess the risk of claim arising from each policyholder and thus to charge appropriate premiums. Stochastic efforts for risk management in seismic datasets have been done by Ermoliev et al. (2000) and Ellingwood and Wen (2005).

The main focus of our application is to identify whether specific areas in which a seismic activity has been recorded, can trigger neighbour areas which also appear to be of high importance in terms of seismicity. Since the data that were collected are count values, we will work appropriately with the models that are introduced in the following sections and therefore choose whether there are any first order temporal dependencies in these spatial neighbouring areas.

In order to assess which area triggers what, we need to resolve to a model selection technique which results us into running 512 models. The choice of the best models was done by calculating AIC (Akaike Information Criterion, Akaike (1974)) for each model, while the final model was chosen by combining rationality and the AIC.

1.4 Format of the dissertation

The rest of this dissertation focuses on a literature review on the Integer Valued Autoregressive models in Chapter 2 and the bivariate version as well along with their properties, forecasting and predicting distributions in Chapter 3. Chapter 4 provides the trivariate version of the Integer Autoregressive models which were my individual work that my supervisor advised me to prove them in a theoretical aspect and also calculate the forecasting and predictive distributions. Chapter 5 provides an analysis of the earthquake time series data in which we work in the trivariate frame along with the model selection to choose which parameters matter. Chapter 6 summarises the work done to date, some comments further and possible avenues for future research.

https://doi.org/10.26219/heal.aueb.5635

Chapter 2

The Univariate INAR(1) Process

The simplest form of the first-order Integer-valued Autoregressive model was firstly introduced by McKenzie (1985) and Al-Osh and Alzaid (1987). The model specification will be introduced along with the first order moments and autocorrelation function. The reader is referred to McKenzie (2003) and Jung and Tremayne (2006a) for a comprehensive review of such models.

The extension of the simple INAR(1) process to the multidimensional space is interesting as it provides a general framework for multivariate count time series modelling. The model has been considered in Franke and Rao Subba (1993) and Latour (1997) but after that, a lot of time passed until some recent work on this by Pedeli and Karlis (2011) and Boudreault and Charpentier (2011).

2.1 Model Specification

A sequence of random variables $\{X_t\}$, $t = 0, 1, 2, \dots$ is an INAR(1) process if it satisfies a difference equation of the form

https://doi.org/10.26219/heal.aueb.5635

$$X_t = \alpha \circ X_{t-1} + R_t, \quad \text{t=1,2,...} \tag{2.1}$$

where $\alpha \in [0, 1]$, R_t is the innovation sequence of uncorrelated non-negative integer-valued random variables with mean μ and finite variance σ^2 (McKenzie (1985); Al-Osh and Alzaid (1987)) and X_0 represents the initial value of the process.

In (2.1) the first right term satisfies the binomial thinning operation properties. This operation is defined as

$$\alpha \circ X = \sum_{i=1}^{X} Y_i = Y$$

where $\{Y_i\}$ is a sequence of i.i.d Bernoulli random variables that satisfy $P(Y_i = 1) = \alpha = 1 - P(Y_i = 0)$ with $\alpha \in [0, 1]$. This operation is due to Steutel and Van Harn (1979) and is based on the scalar multiplication used for normal time series models so that only integer values can be derived. Note that assuming for Y_is any other distribution rather than Bernoulli, we get a generalized Steutel and van Harn operator. Many different other operators can be defined this way. In general Y_is are called the counting sequences.

Detailed theory of binomial thinning operators and their properties as well, are defined in Eduarda Da Silva and Oliveira (2004) while extensions can be found in Alzaid and Al-Osh (1988), Alzaid and Al-Osh (1993) and last but not least Al-Osh and Aly (1992). Generalizations have been proposed by Brännäs and Hellström (2001), Gourieroux and Jasiak (2003).

The INAR(1) model is consisted of two parts. The first part is based on the previous observation's survivor elements of the process at time $t-1$ with

https://doi.org/10.26219/heal.aueb.5635

a survival probability α for each element while the second part consists of the innovations of the current point t, i.e., the elements which were introduced at the system in the interval $(t-1, t]$.

2.1.1 First order moments and the autocorrelation function

Assuming that the mean and variance of the innovations of a stationary INAR(1) are finite, we can define

$$\mu_X = E(X_t) = \frac{\mu_R}{1-\alpha} \text{ and } \sigma_X^2 = Var(X_t) = \frac{\alpha\mu_R + \sigma_R^2}{1-\alpha^2}$$

where μ_R and σ_R^2 are the finite mean and variance of the i.i.d. innovations respectively. The autocovariance function of a stationary INAR(1) process $\{X_t\}_{t\in\mathbb{Z}}$ is derived through the equation

$$\gamma_X(k) = Cov(X_t, X_{t-k}) = \alpha^{|k|}\sigma_X^2,\ k \in \mathbb{Z}$$

and hence the autocorrelation function $\rho_X(k)$ can be directly calculated as

$$\rho_X(k) = \frac{\gamma_X(k)}{\gamma_X(0)} = \alpha^{|k|}$$

Apparently, the autocorrelation function dies out exponentially with lag k and α represents the autocorrelation between consecutive time points. To this aspect it resembles the typical AR model.

In practice, there are two different approaches to construct the INAR(1) models. In the first one a known and particular distribution is assumed and afterwards the required form of the distribution is identified in order to hold the stationarity principle. The latter and most popular approach starts by considering a specific form for the innovation distribution. The main focus

https://doi.org/10.26219/heal.aueb.5635

of introducing the theory and applications is by considering the Poisson case for the innovations, albeit many different forms have been introduced. The reason for this is to favour the fact that both the innovation distribution and the marginal distribution will belong to the same family (Al-Osh and Alzaid, 1987). However, by using real data can be challenging since equidispersion is implied.

2.2 Simulated Example and Extensions

The univariate INAR(1) with Poisson process is shown in Figure 2.1. Obviously, the current time point t is affected by the survivor elements of the process at time $t-1$ with probability $\alpha = 0.5$, while expected survivor elements that where introduced at the system in the interval $(t-1, t]$ is considered to be $\lambda = 1$. The autocorrelation function (ACF) is provided as well where it can be seen that the process is affected by its previous value at $t-1$ with probability $\alpha = 0.5$.

There are several extensions of the model to different directions. The model can be generalized to have p terms, (i.e. INAR(p)) but there is no unique way to do so. Also Moving Average (MA) terms can be added to the model leading to INARMA models. Different operators can be also considered giving rise to a wide range of possible models. Also covariates can be introduced to the mean of the innovation term to allow measuring the effect of additional information leading to INAR regression models. In the next chapter we will present extensions to the multivariate case. First of all we will extend the thinning operator to a matrix form and then present the multivariate INAR model.

https://doi.org/10.26219/heal.aueb.5635

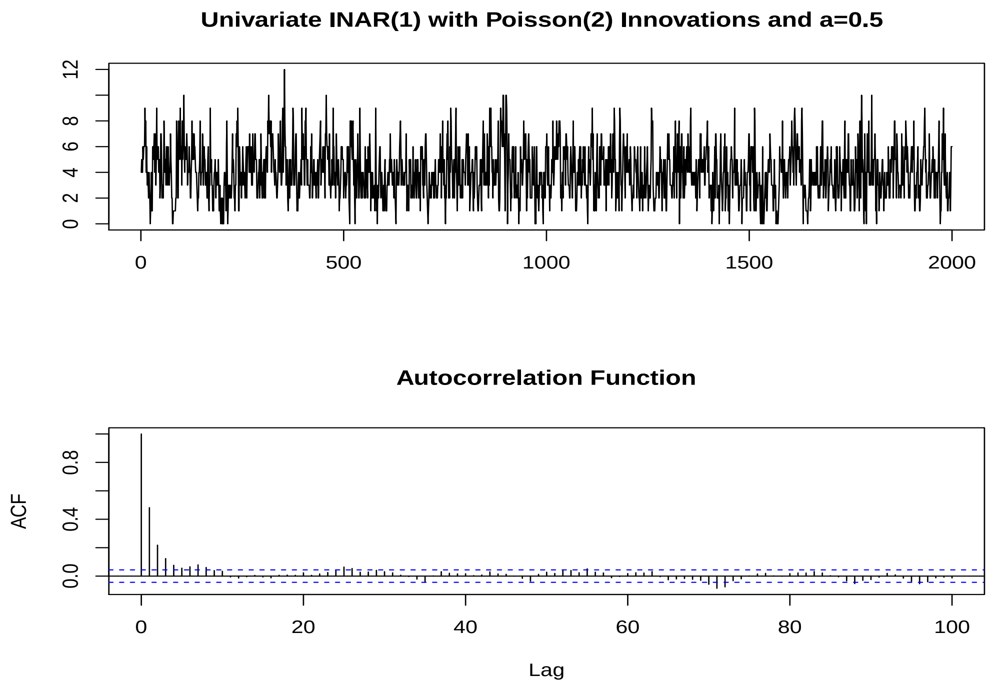

Figure 2.1: Time Series plot and ACF plot for $\alpha = 0.5$ and Poisson innovations with $\lambda = 1$

Let $\boldsymbol{A}$ to be an $r \times r$ matrix with elements α_{ij}, $i,j = 1, ..., r$ and $\boldsymbol{X}$ be a non-negative integer valued r-vector. The matrix operator $'\circ'$ is defined as

$$\boldsymbol{A} \circ \boldsymbol{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_r \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{r} \alpha_{1j} X_j \\ \vdots \\ \sum_{j=1}^{r} \alpha_{rj} X_j \end{bmatrix}$$

The univariate operations $\alpha \circ X$ and $\beta \circ Y$ are independent iff the counting processes in their definition are independent. Hence the matrix operator implies independence between the univariate operations. Properties of this operator can be found in Latour (1997).

https://doi.org/10.26219/heal.aueb.5635

Using this operator, Latour (1997) defined a multivariate generalized INAR process of order p (MGINAR(p)) by assuming that

$$\boldsymbol{X}_t = \sum_{j=1}^{p} \boldsymbol{A}_j Y_{t-j} + \epsilon_t$$

where $\boldsymbol{X}_t$ and ϵ_t are r-vectors and matrices $\boldsymbol{A}_j$ are $r \times r$ matrices. Conditions for existence and stationarity were given. A more focused presentation of the bivariate model follows.

https://doi.org/10.26219/heal.aueb.5635

Chapter 3

The Full BINAR(1) Process

Throughout the years many researchers have used a specific bivariate distribution for a bivariate vector of dependents. In Pedeli and Karlis (2013b) a bivariate Poisson distribution and a bivariate negative binomial were used. For the parametric models prediction is discussed. An interesting result is that for the bivariate Poisson innovations the univariate series have a Hermite marginal distribution. In Karlis and Pedeli (2013) a copula based bivariate innovation distribution was used allowing negative correlation as well. Furthermore, in Pedeli and Karlis (2013a)) a multivariate Poisson distribution is assumed by assuming assumes a diagonal matrix A. They focus into the evolution of innovations by assuming independence and thus introduce it into the model.

Apart from the model specification, the first moments will be discussed and both the joint and predictive distributions will be derived.

3.1 Model Specification

Let $\boldsymbol{X}_t$ and $\boldsymbol{R}_t$ be non-negative integer random vectors of length 2 and let $\boldsymbol{A}$ be a 2×2 diagonal matrix with independent elements α_{jj}, $j = 1, 2$. Then,

https://doi.org/10.26219/heal.aueb.5635

the Bivariate Integer Autoregressive model of order one is defined as

$$\boldsymbol{X}_t = \boldsymbol{A} \circ \boldsymbol{X}_t - 1 + \boldsymbol{R}_t = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} \circ \begin{bmatrix} X_{1,t-1} \\ X_{2,t-1} \end{bmatrix} + \begin{bmatrix} R_{1t} \\ R_{2t} \end{bmatrix} , t \in \mathbb{Z} \tag{3.1}$$

where now $\boldsymbol{A}\circ$ is a matrix operator which behaves the same to the usual matrix multiplication but keeps the properties of the binomial thinning operation. For each element $j = 1, 2$ and $\alpha_{12} = \alpha_{21} = 0$ the equation takes the form of the typical first order Univariate Integer Autoregressive model:

$$X_{jt} = \alpha_j \circ X_{j,t-1} + R_{jt} \tag{3.2}$$

where again the first right term represents the binomial thinning of X_j, i.e., $\alpha_j \circ X_j = \sum_{i=1}^{X_j} Y_i = Y$ with Y being a Bernoulli distributed random variable with success probability $\alpha_j \in [0, 1]$ (Steutel and Van Harn, 1979).

The non-negative integer-valued random process $\{\boldsymbol{X}_t\}$, $t \in \mathbb{Z}$ is the unique strictly stationary solution of the equation model (3.1), if the largest eigenvalue of $\boldsymbol{A}$ is less than 1. In other words, if $max(\alpha_1, \alpha_2) < 1$ whilst the expected mean of the norm of the innovations is finite, i.e., $E(\|\boldsymbol{R}_t\|) < \infty$((see also Franke and Rao Subba, 1993); Latour (1997)).

Analogously to the INAR(1) process, $\boldsymbol{X}_t$ is a composite of two parts. The first one again involves the survival elements at the previous point $t-1$, i.e. $\boldsymbol{X}_{t-1}$ with each one having a survival probability defined by the elements of matrix $\boldsymbol{A}$. The latter part encompasses the innovations which entered the system in the interval $(t-1, t]$.

Using the properties of binomial thinning (Steutel and Van Harn (1979))

 https://doi.org/10.26219/heal.aueb.5635

it is obvious that

$$E(\boldsymbol{A} \circ \boldsymbol{X}) = \boldsymbol{A}E(\boldsymbol{X}) \quad \text{and}$$
$$E\big((\boldsymbol{A} \circ \boldsymbol{X})(\boldsymbol{A} \circ \boldsymbol{X})^\top\big) = \boldsymbol{A}(\boldsymbol{I} - \boldsymbol{A})E(\boldsymbol{X}) + \boldsymbol{A}E(\boldsymbol{X}\boldsymbol{X}^\top)\boldsymbol{A}^\top$$

Considering independence between and within the thinning operations and that the innovation sequence $\{R_t\}$ is i.i.d. with finite mean λ_j, $j = 1, 2$ and variance $\sigma_k^2 = u_j\lambda_j$, $u_j > 0$, then the vector of means takes the form:

$$\mu_1 = \frac{(1 - a_{22})\lambda_1 + \alpha_{12}\lambda_2}{(1 - \alpha_{11})(1 - \alpha_{22}) - \alpha_{12}\alpha_{21}}$$

$$\mu_2 = \frac{(1 - a_{11})\lambda_2 + \alpha_{21}\lambda_1}{(1 - \alpha_{11})(1 - \alpha_{22}) - \alpha_{12}\alpha_{21}} \tag{3.3}$$

and each variance of the two processes and their covariance takes the form:

$$\begin{aligned}\gamma_{11}(0) &= Var(X_t)\\ &= \frac{1}{(1 - \alpha_{11}^2)}\{\alpha_{12}^2 Var(Y_t) + 2\alpha_{11}\alpha_{12}Cov(X_{1t}, X_{2t}) + \alpha_{11}(1 - \alpha_{11})\mu_1 + \alpha_{12}(1 - \alpha_{12})\mu_2 + u_1\lambda_1\}\end{aligned}$$

$$\begin{aligned}\gamma_{22}(0) &= Var(Y_t)\\ &= \frac{1}{(1 - \alpha_{22}^2)}\{\alpha_{21}^2 Var(X_t) + 2\alpha_{21}\alpha_{22}Cov(X_{1t}, X_{2t}) + \alpha_{21}(1 - \alpha_{21})\mu_1 + \alpha_{22}(1 - \alpha_{22})\mu_2 + u_2\lambda_2\}\end{aligned}$$

$$\gamma_{21}(0) = Cov(X_t, Y_t) = \frac{\alpha_{11}\alpha_{21}Var(X_t) + \alpha_{22}\alpha_{12}Var(Y_t) + \phi_{12}}{1 - \alpha_{11}\alpha_{22} - \alpha_{12}\alpha_{21}} \tag{3.4}$$

Since λ_j, σ_j, and α_j take positive values, it is implicated that also the mean, variance and covariance are all positive valued. Relying on u_j, if $u_j > 1$ then the variance may be larger than the mean, if it resides in the interval $(0, 1)$ then the variance is may be smaller than the mean and it is equal to one then the mean and variance are considered equal. These phenomena are called overdispersion, underdispersion and equidispersion respectively.

https://doi.org/10.26219/heal.aueb.5635

In order to introduce dependence for the two series of the BINAR(1), dependence in the innovation R_{1t} and R_{2t} must be considered while preserving the previous assumptions fixed. No matter what type of distribution is hidden in the innovations, the covariance ϕ_{12} of R_{1t}, R_{2t} at time t completely determines the covariance between the current value of the process and the innovations of the other process at the same time and vice versa Pedeli and Karlis (2011), i.e.,

$$Cov(X_{1t}, R_t) = Cov(R_{1t}, R_{2t}) \tag{3.5}$$

The equation (3.5) assumes that $\{\boldsymbol{X}_t\}$ is strictly stationary, which means that the joint distribution of $\{\boldsymbol{X}_t\}$ is the same as $\{\boldsymbol{X}_{t+h}\}$.

3.2 Model Distribution

In order to derive the distribution of the process, under stationarity, the joint probability generating function (pgf) $G_X(\boldsymbol{s})$ satisfies the difference equation

$$G_X(s) = G_X(A^T s) G_R(s) \tag{3.6}$$

for further details see also Steutel and Van Harn (1986). By considering that $X_{jt} = \sum_{i=0}^{\infty} A_j^i \circ R_{j,t-i} = \sum_{i=0}^{\infty} B_i$ we would like to focus on the search of the unconditional distribution of $\boldsymbol{B}$ which is a simplified notation form.

We can start now investigating by defining A^i as

$$A^i = \begin{bmatrix} \alpha_i & \beta_i \\ \gamma_i & \delta_i \end{bmatrix}$$

https://doi.org/10.26219/heal.aueb.5635

thus the vector $\boldsymbol{B}$ takes the form

$$B^i = \begin{bmatrix} \alpha_i & \beta_i \\ \gamma_i & \delta_i \end{bmatrix} \circ \begin{bmatrix} R_1 \\ R_2 \end{bmatrix} = \begin{bmatrix} \alpha_i \circ R_1 + \beta_i \circ R_2 \\ \gamma_i \circ R_1 + \delta_i \circ R_2 \end{bmatrix}$$

By definition it is known that the binomial thinning is independent and so the conditional distribution $B_i|R$ has pgf

$$G_{B_i|R}(s_1, s_2) = (1-\alpha_i+\alpha_i s_1)^{R_1}(1-\beta_i+\beta_i s_2)^{R_2}(1-\gamma_i+\gamma_i s_1)^{R_1}(1-\delta_i+\delta_i s_2)^{R_2}$$

and by integrating $\boldsymbol{R}$, the unconditional pgf becomes

$$\begin{aligned} G_{B_i}(s_1, s_2) &= E[(1-\alpha_i+\alpha_i s_1)^{R_1}(1-\beta_i+\beta_i s_2)^{R_2}(1-\gamma_i+\gamma_i s_1)^{R_1}(1-\delta_i+\delta_i s_2)^{R_2}] \\ &= G_R((1-\alpha_i+\alpha_i s_1)(1-\gamma_i+\gamma_i s_1), (1-\beta_i+\beta_i s_2)(1-\delta_i+\delta_i s_2)) \end{aligned} \tag{3.7}$$

and finally the joint pgf $G_X(s)$ takes the form

$$\begin{aligned} G_{\mathbf{X}}(s) &= G_{X_t,Y_t}(s_1, s_2) \\ &= \prod_{i=0}^{\infty} G_{\mathbf{R}}((1-\alpha_i+\alpha_i s_1)(1-\gamma_i+\gamma_i s_1), (1-\beta_i+\beta_i s_2)(1-\delta_i+\delta_i s_2)) \end{aligned}$$

where $a_0 = \beta_0 = \gamma_0 = \delta_0 = 1$. By taking a decomposition of $\mathbf{A}$, we can write it as

$$A^i = PD^iP^{-1}$$

where $\mathbf{P}$ is the matrix with the eigenvectors and $\mathbf{D}$ the diagonal matrix with the eigenvalues. Since in order to achieve stationarity, the eigenvalues must be smaller than 1, then as $i \to \infty$, D^i tends to zero and through that A^i as

https://doi.org/10.26219/heal.aueb.5635

well. Pedeli and Karlis (2013c) has proved that $\boldsymbol{A}^i$ has elements

$$\alpha_i = \frac{2^{-1-i}(\alpha_{11}+\alpha_{22}-C)^i(-\alpha_{11}+\alpha_{22}+C)}{C} - \frac{2^{-1-i}(-\alpha_{11}+\alpha_{22}-C)(\alpha_{11}+\alpha_{22}+C)^i}{C}$$

$$\beta_i = -\frac{2^{-2-i}(\alpha_{11}+\alpha_{22}-C)^i(-\alpha_{11}-\alpha_{22}+C)(-\alpha_{11}+\alpha_{22}+C)}{\alpha_{21}C} - \frac{2^{-2-i}(-\alpha_{11}+\alpha_{22}-C)(-\alpha_{11}+\alpha_{22}+C)(\alpha_{11}+\alpha_{22}+C)^i}{\alpha_{21}C}$$

$$\gamma_i = -\frac{2^{-i}\alpha_{21}(\alpha_{11}+\alpha_{22}-C)^i}{C} + \frac{2^{-i}\alpha_{21}(\alpha_{11}+\alpha_{22}+C)^i}{C}$$

$$\delta_i = \frac{2^{-1-i}(\alpha_{11}+\alpha_{22}-C)^i(\alpha_{11}-\alpha_{22}+C)}{C} + \frac{2^{-1-i}(-\alpha_{11}+\alpha_{22}+C)(\alpha_{11}+\alpha_{22}+C)^i}{C}$$

with $C = \sqrt{\alpha_{11}^2 + 4\alpha_{12}\alpha_{21} - 2\alpha_{11}\alpha_{22} + \alpha_{22}^2}$.

3.3 Estimation of full BINAR(1) Model

A general least squares estimate was discussed in Latour (1997). However if parametric assumptions are used for the innovations then more detailed estimation can be done. Parametric models offer also more exibility for predictions.

In order to estimate the parameters of the full BINAR(1) model, the conditional maximum likelihood will be considered. The conditional distribution of the full BINAR(1) model can be interpreted as a convolution of two bino-

https://doi.org/10.26219/heal.aueb.5635

mial variates and a bivariate distribution of the innovations, i.e.

$$f_1(k) = \sum_{j_X=0}^{k} \binom{X_{t-1}}{j_X} \binom{Y_{t-1}}{k-j_X} \alpha_{11}^{j_X} (1-\alpha_{11})^{X_{t-1}-j_X} \alpha_{12}^{k-j_X} (1-\alpha_{12})^{Y_{t-1}-k+j_X}$$
$$f_2(s) = \sum_{j_Y=0}^{s} \binom{Y_{t-1}}{j_Y} \binom{X_{t-1}}{s-j_Y} \alpha_{22}^{j_Y} (1-\alpha_{22})^{Y_{t-1}-j_Y} \alpha_{21}^{s-j_Y} (1-\alpha_{21})^{X_{t-1}-s+j_Y}$$

and with the bivariate distribution $f_3(r_1, r_2) = P(R_{1t} = r_1, R_{2t} = r_2)$, the conditional distribution takes the form

$$f(x_{1t}|x_{1,t-1}) = \sum_{k=0}^{g_1} \sum_{s=0}^{g_2} f_1(k) f_2(s) f_3(x_{1t}-k, x_{2t}-s) \tag{3.8}$$

where $g_1 = min(x_{1t}, x_{1,t-1})$ and $g_2 = min(x_{2t}, x_{2,t-1})$ and thus the conditional likelihood takes the form

$$L(\theta|x) = \prod_{t=1}^{T} f(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \theta) \tag{3.9}$$

where θ is the vector of the unknown parameters and can be estimated via maximasing (3.9). It is known that asymptotically the estimate $\hat{\theta}$ is normally distributed after appointing a set of regularity conditions and applying results of Billingsley (1961) for estimation of Markov processes.

Numerical maximization of (3.9) is straightforward with R. The binomial convolution implies finite summation and hence it is easy. Note also that since the pgf is polynomial calculation of the convolution is easy via polynomial multiplications. Packages in R exist that allows quick polynomial multiplication and hence easy and cheap convolution calculation.

https://doi.org/10.26219/heal.aueb.5635

3.4 Predictive Distribution for the full BINAR(1) Model

The derivation of the predictive distribution of a full BINAR(1) model is based on the statement of Al-Osh and Alzaid (1987) that the conditional distribution of the INAR(1) model is given by

$$(X_t, X_{t-h}) \quad = \quad \big(\alpha^h \circ X_{t-h} + \sum_{i=0}^{h-1} \alpha^i \circ R_{t-i}, X_{t-h}\big) \tag{3.10}$$

where R_t is the sequence of uncorrelated non-negative integer-valued random variables with finite mean and variance. Now, the conditional distribution of the full BINAR(1) model can be written in a similar form with the help of binomial thinning (Steutel and Van Harn, 1979) properties.

$$(\boldsymbol{X}_t, \boldsymbol{X}_{t-h}) \quad = \quad \big(\boldsymbol{A}^h \circ \boldsymbol{X}_{t-h} + \sum_{i=0}^{h-1} \boldsymbol{A}^i \circ \boldsymbol{R}_{t-i}, \boldsymbol{X}_{t-h}\big) \tag{3.11}$$

where now $\boldsymbol{R}_t$ are interpreted as correlated non-negative integer values random vectors of length *2*. By defining $\boldsymbol{B} = \sum_{i=0}^{h-1} \boldsymbol{B}_i = \sum_{i=0}^{h-1} \boldsymbol{A}^i \circ \boldsymbol{R}t - i$ and

$$\boldsymbol{A}^i = \begin{bmatrix} \alpha_i & \beta_i \\ \gamma_i & \delta_i \end{bmatrix}, \ \boldsymbol{A}^h = \begin{bmatrix} \alpha_h & \beta_h \\ \gamma_h & \delta_h \end{bmatrix}$$

with the help of (3.7) in the same way with the joint distribution, the predictive distribution can be derived,

$$G_B(s) \quad = \quad \prod_{i=0}^{h-1} G_R((1-\alpha_i+\alpha_i s_1)(1-\gamma_i+\gamma_i s_1), (1-\beta_i+\beta_i s_2)(1-\delta_i+\delta_i s_2)) \tag{3.12}$$

https://doi.org/10.26219/heal.aueb.5635

Thus, the joint pgf $X_{1,T+h}, X_{2,T+h}|X_{1T}, Y_{2T}$ takes the form

$$G_{X_{T+h}}(s|\boldsymbol{x}_T) = [(1-\alpha_i+\alpha_i s_1)(1-\gamma_i+\gamma_i s_1)]^{X_{1T}}[(1-\beta_i+\beta_i s_2)(1-\delta_i+\delta_i s_2)]^{X_{2T}}G_B(s)$$

Hence, the predictive density distribution is the following

$$\begin{aligned}
P(\boldsymbol{X}_{t+h}=x|\boldsymbol{X}_T) = \sum_{k=0}^{m_{X_1}}\sum_{s=0}^{m_{X_2}}\Bigg[\sum_{j_{X_1}}^{k}\binom{X_{1T}}{j_{X_1}}\binom{X_{2T}}{k-j_{X_1}}\alpha_h^{j_{X_1}}(1-\alpha_h)^{X_{1T}-j_{X_1}}\beta_h^{k-j_{X_1}}(1-\beta_h)^{X_{2T}-k+j_{X_1}} \\
\times\sum_{j_{X_2}}^{s}\binom{X_{2T}}{j_{X_2}}\binom{X_{1T}}{s-j_{X_2}}\delta_h^{j_{X_2}}(1-\delta_h)^{X_{2T}-j_{X_2}}\gamma_h^{s-j_{X_2}}(1-\gamma_h)^{X_{2T}-s+j_{X_1}}\Bigg] \\
\times P\Bigg(\sum_{i=0}^{h-1}(\alpha_i\circ R_{1,T+h-i}+\beta_i\circ R_{2,T+h-i})=x_1-k, \\
\sum_{i=0}^{h-1}(\gamma_i\circ R_{1,T+h-i}+\delta_i\circ R_{2,T+h-i})=x_2-s\Bigg)
\end{aligned}$$

where $m_{X_1} = min(x_1, x_{1T})$ and $m_2 = min(x, x_{2T})$ and the elements of the matrices $\boldsymbol{A}^i$, $\boldsymbol{A}^h$ can be computed numerically.

3.5 The Full BINAR(1) Process with Bivariate Poisson Innovations

While different distributional assumptions can be introduced for the evolution of innovations, we will focus on the Bivariate Poisson case's properties, estimation and illustration.

3.5.1 Model Specification

Assuming that the innovation process $\{R_{1t}, R_{2t}\}$ follows a bivariate Poisson distribution, we can write their probability mass function as BP(λ_1,λ_2,ϕ), i.e.

$$P(R_{1t} = x_1, R_{2t} = x_2) = \exp -(\lambda_1 + \lambda_2 - \phi)\frac{(\lambda_1 - \phi)^x_1}{x_1!}\frac{(\lambda_2 - \phi)^x_2}{x_2!} \times \sum_{i=0}^{min(x_1,x_2)} \binom{x_1}{i}\binom{x_2}{i} i! \Big(\frac{\phi}{(\lambda_1 - \phi)(\lambda_2 - \phi)}\Big)^i \tag{3.13}$$

where λ_1,$\lambda_2 > 0$ and $\phi \in [0, min(\lambda_1,\lambda_2))$ which is the covariance between the two random variables and marginally these two variables are Poisson distributed with λ_1 and λ_2 respectively.

Obviously, if the two variables are independent, and conclusively $\phi = 0$, then the pmf takes the form of the product of two independent Poisson distributions, i.e.

$$P(R_{1t} = x, R_{2t} = y) = \exp -(\lambda_1 + \lambda_2)\frac{\lambda_1^{x_1}}{x_1!}\frac{\lambda_2^{x_2}}{x_2!} \tag{3.14}$$

Therefore, $\boldsymbol{R}_t$ are equidispersed with $\sigma_{X1} = \lambda_1$ and $\sigma_{X2} = \lambda_2$. Now the variances and covariance take form

$$\begin{aligned} \gamma_{11}(0) &= Var(X_{1t}) \\ &= \Big(\big[(1-\alpha_{11}^2)(1-\alpha_{22}^2)(1-\alpha_{11}\alpha_{22}) - \alpha_{12}\alpha_{21}(1-\alpha_{11}^2) - \alpha_{12}^2\alpha_{21}^2(1+\alpha_{11}\alpha_{22}-\alpha_{12}\alpha_{21})\big]\mu_1 \\ &+ 2\alpha_{12}(\alpha_{11}+\alpha_{12}\alpha_{21}\alpha_{22}-\alpha_{11}\alpha_{22}^2)(\alpha_{12}\alpha_{22}\mu_2+\phi)\Big)/D_\gamma \end{aligned} \tag{3.15}$$

$$\begin{aligned} \gamma_{22}(0) &= Var(X_{2t}) \\ &= \Big(\big[(1-\alpha_{11}^2)(1-\alpha_{22}^2)(1-\alpha_{11}\alpha_{22}) - \alpha_{12}\alpha_{21}(1-\alpha_{11}^2) - \alpha_{12}^2\alpha_{21}^2(1+\alpha_{11}\alpha_{22}-\alpha_{12}\alpha_{21})\big]\mu_2 \\ &+ 2\alpha_{21}(\alpha_{22}+\alpha_{11}\alpha_{12}\alpha_{21}-\alpha_{11}^2\alpha_{22})(\alpha_{11}\alpha_{21}\mu_1+\phi)\Big)/D_\gamma \end{aligned} \tag{3.16}$$

 https://doi.org/10.26219/heal.aueb.5635

$$\begin{aligned}\gamma_{21}(0) &= Cov(X_{1t}, X_{2t}) \\ &= \left[(1-\alpha_{11}^2)(1-\alpha_{22}^2) - \alpha_{12}^2\alpha 21^2\right](\alpha_{11}\alpha_{21}\mu_1 + \alpha_{12}\alpha_{22}\mu_2 + \phi)/D_\gamma \quad (3.17)\end{aligned}$$

where $D_\gamma = (1-\alpha_{11}^2)(1-\alpha_{22}^2)(1-\alpha_{11}\alpha_{22} - \alpha_{12}\alpha_{21}(1+\alpha_{11}^2+\alpha_{11}^2 - 3\alpha_{11}^2\alpha_{22}^2) - \alpha_{12}^2\alpha_{21}^2(1-\alpha_{12}\alpha_{21}+3\alpha_{11}\alpha_{22})$. However, the process is overdispersed and its magnitude is dependent on the values of the parameters Pedeli and Karlis (2013c).

3.5.2 Estimation

Considering that the innovation series are Bivariate Poisson distributed as shown in (3.13) and using convolutions as shown in the previous section, $f_3(.)$ takes the form

$$f_3(x_{1t}-k, x_{2t}-s) = \exp -(\lambda_1+\lambda_2-\phi)\sum_{m=0}^{b}\frac{(\lambda_1-\phi)^{x_{1t}-k-m}((\lambda_2-\phi)^{x_{2t}-s-m})\phi^m}{(x_{1t}-k-m)!(x_{2t}-s-m)!m!} \quad (3.18)$$

where $b = min(x_{1t}-k, x_{2t}-s)$. Thus (3.8) with the help of $f_1(.), f_2(.), f_3(.)$ takes the form

$$\begin{aligned}f(\boldsymbol{x}_t|\boldsymbol{x}_{t-1},\boldsymbol{\theta}) = \sum_{k=0}^{g_1}\sum_{s=0}^{g_2}&\left[\sum_{j_{X_1}=0}^{k}\binom{x_{1,t-1}}{j_{X_1}}\binom{x_{2,t-1}}{k-j_{X_1}}\alpha_{11}^{j_{X_1}}(1-\alpha_{11})^{x_{t-1}-j_{X_1}}\alpha_{12}^{k-j_{X_1}}(1-\alpha_{12})^{x_{2,t-1}-k+j_{X_1}}\right] \\ &\times\left[\sum_{j_{X_2}=0}^{s}\binom{x_{2,t-1}}{j_{X_2}}\binom{x_{1,t-1}}{s-j_{X_2}}\alpha_{22}^{j_{X_2}}(1-\alpha_{22})^{x_{2,t-1}-j_{X_2}}\alpha_{21}^{s-j_{X_2}}(1-\alpha_{21})^{x_{1,t-1}-s+j_{X_2}}\right] \\ &\times\exp -(\lambda_1+\lambda_2-\phi)\sum_{m=0}^{b}\frac{(\lambda_1-\phi)^{x_{1t}-k-m}((\lambda_2-\phi)^{x_{2t}-s-m})\phi^m}{(x_{1t}-k-m)!(x_{2t}-s-m)!m!}\end{aligned} \quad (3.19)$$

with $\theta = \alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22}, \lambda_1, \lambda_2, \phi$, $g_1 = min(x_{1t}, x_{1,t-1})$,$g_2 = min(x_{2t}, x_{2,t-1})$ and $b = min(k, s)$. Therefore, $L(\theta|x) = \prod_{t=1}^{T} f(\boldsymbol{x}_t|\boldsymbol{x}_{t-1},\boldsymbol{\theta})$ for some initial

https://doi.org/10.26219/heal.aueb.5635

value $\boldsymbol{x}_0$ can be maximised in order to derive the ML estimators of the vector θ.

3.6 Simulated Examples

In Figures 3.1 and 3.2 the overdispersion and the almost equal means can be seen. Through ACF and cross correlation plots the contribution from the one series to another is obvious. For lower and higher lags as well series 1 triggers series 2 and vice versa while for lag$= \pm 1$ the triggering happens with probabilities equal to 0.2. Analogously, for the series themselves,i.e., the previous value at time $t-1$ for both series, triggers the one at time t with probability 0.5.

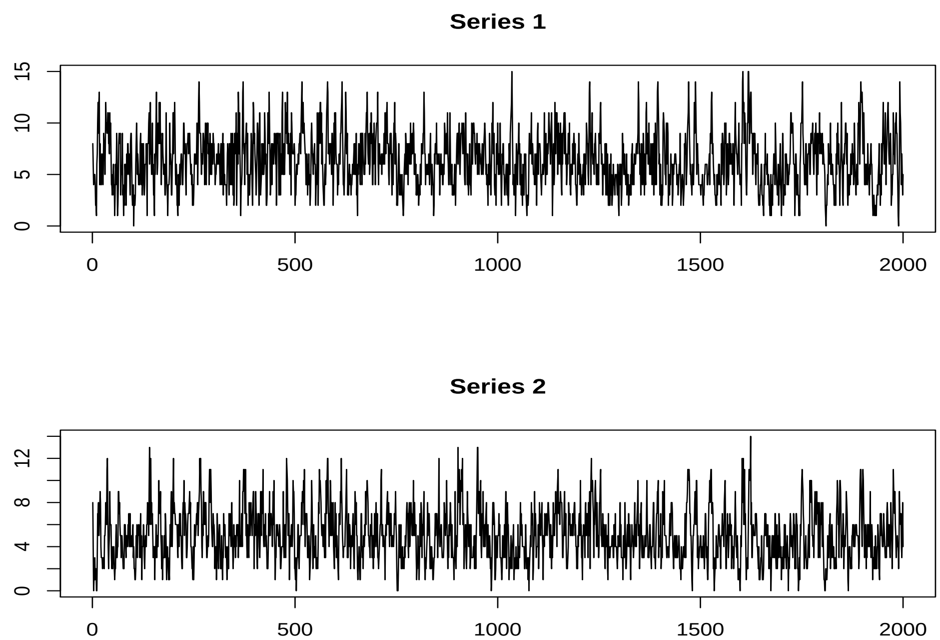

Figure 3.1: Time Series plot for $\alpha_{11} = 0.5$,$\alpha_{12} = 0.2$, $\alpha_{21} = 0.2$, $\alpha_{22} = 0.5$ and Poisson innovations with $\lambda_1 = 2$, $\lambda_2 = 1$, $\phi = 0.2$

https://doi.org/10.26219/heal.aueb.5635

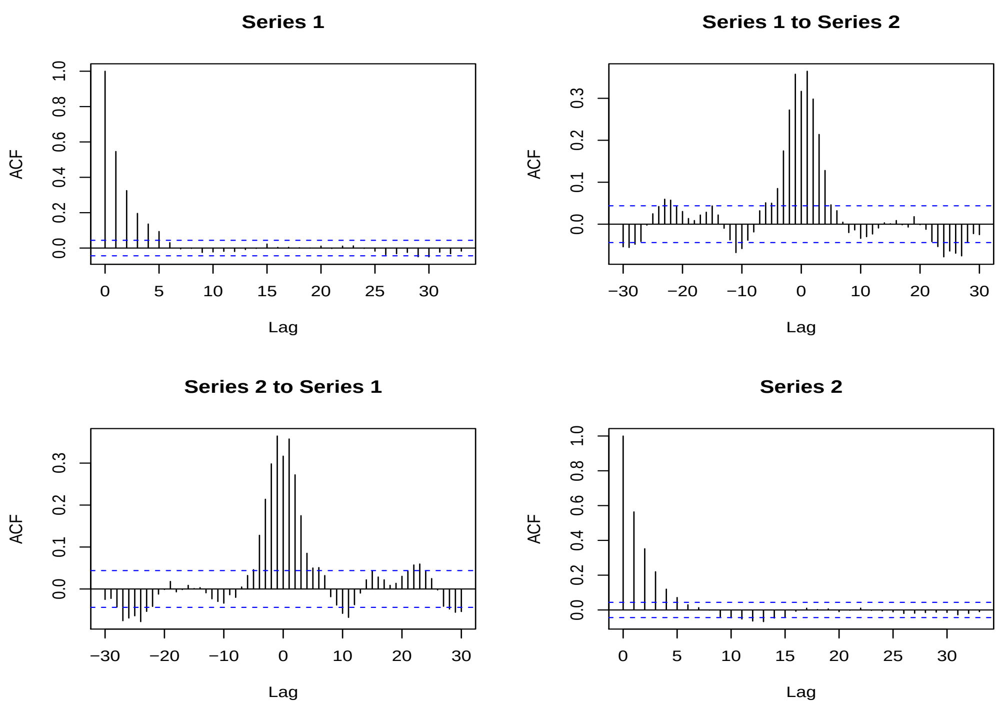

Figure 3.2: ACF and CCF plot for $\alpha_{11} = 0.5$,$\alpha_{12} = 0.2$, $\alpha_{21} = 0.2$, $\alpha_{22} = 0.5$ and Poisson innovations with $\lambda_1 = 2$, $\lambda_2 = 1$, $\phi = 0.2$

https://doi.org/10.26219/heal.aueb.5635

Chapter 4

The Full TRINAR(1) Process with trivariate Poisson Innovations

In this chapter the trivariate version of integer-valued autoregressive models of order one for uncorrelated innovations is described. Since moving higher in dimension it is computationally demanding, the exclusion of dependence parameter vector $\boldsymbol{\phi}$ was considered in the simulations and application.

Even in the case that the pmf of the multivariate Poisson distribution is easily written, the computational power is very demanding. For instance, if $\boldsymbol{A}$ is considered diagonal, the decrease in parameters to be estimated produces a staggering difference in efficiency. The conditional likelihood can be derived as in the bivariate case but now this is a convolution of several binomials and a multivariate discrete distribution.

In the next sections apart from the model specification, the first moments will be proved while the joint and predictive distributions will be derived for uncorrelated innovations.

 https://doi.org/10.26219/heal.aueb.5635

4.1 Model Specification

Let $\boldsymbol{X}_t$ and $\boldsymbol{R}_t$ be non-negative integer random vectors of length 3 and let $\boldsymbol{A}$ be a 3×3 diagonal matrix with independent elements α_{jj}, $j = 1, 2, 3$. Then the full trivariate autoregressive model of order one is defined as

$$\boldsymbol{X}_t = \boldsymbol{A} \circ \boldsymbol{X}_t - 1 + \boldsymbol{R}_t = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \end{bmatrix} \circ \begin{bmatrix} X_{1,t-1} \\ X_{2,t-1} \\ X_{3,t-1} \end{bmatrix} + \begin{bmatrix} R_{1t} \\ R_{2t} \\ R_{3t} \end{bmatrix} t \in \mathbb{Z} \quad (4.1)$$

where again $\boldsymbol{A}\circ$ is an operator matrix which behaves the same to the usual matrix multiplication but keeps the properties of the binomial thinning operation. For each element $j = 1, 2, 3$ the equation takes the form of the typical first order Univariate Integer Autoregressive model:

$$X_{1t} = \alpha_{11} \circ X_{1,t-1} + \alpha_{12} \circ X_{2,t-1} + \alpha_{13} \circ X_{3,t-1} + R_{1t} \quad (4.2)$$

where again the first right term represents the binomial thinning of X_j, i.e., $\alpha_j \circ X_j = \sum_{i=1}^{X_j} Y_i = Y$ with Y being a Bernoulli distributed random variable with success probability $\alpha_j \in [0, 1]$ Steutel and Van Harn (1979). The unique strictly stationary solution, analogously to the bivariate case is the integer-valued random process $\{\boldsymbol{X}_t\}$, $t \in \mathbb{Z}$ of the equation model (4.1), i.e., if $max(\alpha_1, \alpha_2, \alpha_3) < 1$ with $E(\|\boldsymbol{R}_t\|)$ finite.

Considering the composition of the TRINAR(1) series $\boldsymbol{X}_t$, $\boldsymbol{A} \circ \boldsymbol{X}_{t-1}$ is referred to the survivors of the elements of the process at the previous time point $t - 1$ as well as by relaxing the diagonality assumption of $\boldsymbol{A}$, we can conclude that not only each series is connected, or in other words, dependent to its own survivors, but also to the survivors of the elements of the other series at the previous point $t - 1$. Additionally, the second component R_t are the innovations which can be assumed either dependent or independent.

It is straightforward to obtain the moments of the trivariate process by using the properties of binomial thinning from Latour (1997) and Franke and Rao Subba (1993):

$$\boldsymbol{\mu} = E(\boldsymbol{X}_t) = (\boldsymbol{I} - \boldsymbol{A})^{-1} E(\boldsymbol{R}_t) \tag{4.3}$$

the vector of means is calculated via the formula

$$\boldsymbol{\mu} = \frac{1}{det(\boldsymbol{I} - \boldsymbol{A})} adj(\boldsymbol{I} - \boldsymbol{A}) \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix}$$

Thus the means take the form

$$\mu_1 = \frac{(1 - \alpha_{33} - \alpha_{22} + \alpha_{22}\alpha 33 - \alpha 23\alpha 32)\lambda 1 + (\alpha_{13}\alpha_{32} + \alpha_{12} - \alpha_{12}\alpha_{33})\lambda_2 + (\alpha_{12}\alpha_{23} + \alpha_{13} - \alpha_{13}\alpha_{22}}{C}$$

$$\mu_2 = \frac{(\alpha_{23}\alpha_{31} + \alpha_{21} - \alpha_{21}\alpha_{33})\lambda_1 + (1 - \alpha_{33} - \alpha_{11} + \alpha_{11}\alpha_{33} - \alpha_{13}\alpha_{31})\lambda_2 + (\alpha_{13}\alpha_{21} + \alpha_{23} - \alpha_{23}\alpha_{11})\lambda_3}{C}$$

$$\mu_3 = \frac{(\alpha_{21}\alpha_{32} + \alpha_{31} - \alpha_{31}\alpha_{22})\lambda_1 + (\alpha_{12}\alpha_{31} + \alpha_{32} - \alpha_{32}\alpha_{11})\lambda_2 + (1 - \alpha_{22} - \alpha_{11} + \alpha_{11}\alpha_{22} - \alpha_{21}\alpha_{12})\lambda_3}{C}$$

with C being the determinant of $(\boldsymbol{I} - \boldsymbol{A})$ which was calculated as

$$\begin{aligned} C = det(\boldsymbol{I} - \boldsymbol{A}) = 1 &- \alpha_{33} - \alpha_{22} + \alpha_{22}\alpha_{33} - \alpha_{32}\alpha_{23} - \alpha_{11} + \alpha_{11}\alpha_{33} + \alpha_{11}\alpha_{22} - \alpha_{11}\alpha_{22}\alpha_{33} \\ &+ \alpha_{11}\alpha_{32}\alpha_{23} - \alpha_{12}\alpha_{21} + \alpha_{12}\alpha_{21}\alpha_{33} - \alpha_{12}\alpha_{31}\alpha_{23} - \alpha_{31}\alpha_{21}\alpha_{32} - \alpha_{13}\alpha_{31} + \alpha_{13}\alpha_{31}\alpha_{22} \end{aligned}$$

The variance-covariance matrix satisfies the equation

$$\gamma(0) = Var(\boldsymbol{X}_t) = \boldsymbol{A}\gamma(0)\boldsymbol{A}^T + diag(\boldsymbol{B}\boldsymbol{\mu}) + Var(\boldsymbol{R}_t) \tag{4.4}$$

where $\boldsymbol{B}$ is the 3×3 variance matrix of the Bernoulli random variables $[B_{ij}] = a_{ij}(1 - a_{ij})$, while the innovations $\boldsymbol{R}_t$ are identically distributed sequences $\{R_{it}\}_{i=1}^3$ with mean $E(\boldsymbol{R}_t) = \lambda = (\lambda_1, \lambda_2, \lambda_3)'$ and with variance-covariance

https://doi.org/10.26219/heal.aueb.5635

matrix

$$V(\boldsymbol{R}_t) = \begin{bmatrix} u_1\lambda_1 & \phi_{12} & \phi_{13} \\ \phi_{12} & u_2\lambda_2 & \phi_{23} \\ \phi_{31} & \phi_{32} & u_3\lambda_3 \end{bmatrix}$$

Depending on u_i we could say that whether $u_i < 1$, $u_i > 1$ or $u_i = 1$, we have underdispersion, overdispersion or equidispersion respectively and if all $\phi_{ij} = 0$ then the innovations are independent. Furthermore, it is known that the difference equation Cov(X_{it}, R_{jt})=Cov(R_{it}, R_{jt}), $i, j = 1, 2, 3$, $i \neq j$, is satisfied for the full TRINAR(1) model. This means that the covariance of one process and the innovations of the other at the same point t equals to the covariance of the innovations of these two processes for the same point t.

The first term of (4.4) is

$$\boldsymbol{A}\gamma(0)\boldsymbol{A}^T = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$$

with

$$a = (\alpha_{11}\gamma_{11}+\alpha_{12}\gamma_{21}+\alpha_{13}\gamma_{31})\alpha_{11}+(\alpha_{11}\gamma_{12}+\alpha_{12}\gamma_{22}+\alpha_{13}\gamma_{32})\alpha_{12}+(\alpha_{11}\gamma_{13}+\alpha_{12}\gamma_{23}+\alpha_{13}\gamma_{33})\alpha_{13}$$

$$b = (\alpha_{11}\gamma_{11}+\alpha_{12}\gamma_{21}+\alpha_{13}\gamma_{31})\alpha_{21}+(\alpha_{11}\gamma_{12}+\alpha_{12}\gamma_{22}+\alpha_{13}\gamma_{32})\alpha_{12}+(\alpha_{11}\gamma_{13}+\alpha_{12}\gamma_{23}+\alpha_{13}\gamma_{33})\alpha_{23}$$

$$c = (\alpha_{11}\gamma_{11}+\alpha_{12}\gamma_{21}+\alpha_{13}\gamma_{31})\alpha_{31}+(\alpha_{11}\gamma_{12}+\alpha_{12}\gamma_{22}+\alpha_{13}\gamma_{32})\alpha_{32}+(\alpha_{11}\gamma_{13}+\alpha_{12}\gamma_{23}+\alpha_{11}\gamma_{33})\alpha_{33}$$

$$d = (\alpha_{21}\gamma_{11}+\alpha_{22}\gamma_{21}+\alpha_{23}\gamma_{31})\alpha_{11}+(\alpha_{21}\gamma_{12}+\alpha_{22}\gamma_{22}+\alpha_{23}\gamma_{32})\alpha_{12}+(\alpha_{21}\gamma_{13}+\alpha_{22}\gamma_{23}+\alpha_{23}\gamma_{33})\alpha_{13}$$

$$e = (\alpha_{21}\gamma_{11}+\alpha_{22}\gamma_{21}+\alpha_{23}\gamma_{31})\alpha_{21}+(\alpha_{21}\gamma_{12}+\alpha_{22}\gamma_{22}+\alpha_{23}\gamma_{32})\alpha_{22}+(\alpha_{21}\gamma_{13}+\alpha_{22}\gamma_{23}+\alpha_{23}\gamma_{33})\alpha_{23}$$

$$f = (\alpha_{21}\gamma_{11}+\alpha_{22}\gamma_{21}+\alpha_{23}\gamma_{31})\alpha_{31}+(\alpha_{21}\gamma_{12}+\alpha_{22}\gamma_{22}+\alpha_{23}\gamma_{32})\alpha_{32}+(\alpha_{21}\gamma_{13}+\alpha_{22}\gamma_{23}+\alpha_{23}\gamma_{33})\alpha_{33}$$

$$g = (\alpha_{31}\gamma_{11}+\alpha_{32}\gamma_{21}+\alpha_{33}\gamma_{31})\alpha_{11}+(\alpha_{31}\gamma_{12}+\alpha_{32}\gamma_{22}+\alpha_{33}\gamma_{32})\alpha_{12}+(\alpha_{31}\gamma_{13}+\alpha_{32}\gamma_{23}+\alpha_{33}\gamma_{33})\alpha_{13}$$

$$h = (\alpha_{31}\gamma_{11}+\alpha_{32}\gamma_{21}+\alpha_{33}\gamma_{31})\alpha_{21}+(\alpha_{31}\gamma_{12}+\alpha_{32}\gamma_{22}+\alpha_{33}\gamma_{32})\alpha_{22}+(\alpha_{31}\gamma_{13}+\alpha_{32}\gamma_{23}+\alpha_{33}\gamma_{33})\alpha_{23}$$

$$i = (\alpha_{31}\gamma_{11}+\alpha_{32}\gamma_{21}+\alpha_{33}\gamma_{31})\alpha_{31}+(\alpha_{31}\gamma_{12}+\alpha_{32}\gamma_{22}+\alpha_{33}\gamma_{32})\alpha_{32}+(\alpha_{31}\gamma_{13}+\alpha_{32}\gamma_{23}+\alpha_{33}\gamma_{33})\alpha_{33}$$

and the second element takes the form

$$diag(\boldsymbol{B}\mu) = \begin{bmatrix} b_{11} & 0 & 0 \\ 0 & b_{22} & 0 \\ 0 & 0 & b_{33} \end{bmatrix}$$

with

$$b_{11} = \alpha_{11}(1-\alpha_{11})\mu_1 + \alpha_{12}(1-\alpha_{12})\mu_2 + \alpha_{13}(1-\alpha_{13})\mu_3$$

 https://doi.org/10.26219/heal.aueb.5635

$$b_{22} = \alpha_{21}(1-\alpha_{21})\mu_1 + \alpha_{22}(1-\alpha_{22})\mu_2 + \alpha_{23}(1-\alpha_{23})\mu_3$$

$$b_{33} = \alpha_{31}(1-\alpha_{31})\mu_1 + \alpha_{32}(1-\alpha_{32})\mu_2 + \alpha_{33}(1-\alpha_{33})\mu_3$$

The covariance function $\gamma(h)$ for $h > 0$, is computed through iterative procedures and finally takes the form:

$$\gamma(h) = E\{(\boldsymbol{X}_{t+h} - \boldsymbol{\mu})(\boldsymbol{X}_{t+h} - \boldsymbol{\mu})^T\} = \boldsymbol{A}\gamma(h-1) = \boldsymbol{A}^h\gamma(0), \quad h \geq 1 \quad (4.5)$$

The marginal correlation process is one of ARMA(3,4) structure (McKenzie (1988); Dewald et al. (1989)) and by using the Cayley Hamilton Theorem to (4.5), then there exist constants ξ_1,ξ_2,ξ_3 such that $\boldsymbol{A}^3 - \xi_1\boldsymbol{A}^2 - \xi_3\boldsymbol{I}_3 = 0$ and thus $\gamma(h)$ satisfies

$$\gamma(h) - \xi_1\gamma(h-1) - \cdots - \xi_n\gamma(h-n) = 0, \quad h \geq n \quad (4.6)$$

Through (4.5) and (4.6) the autocorrelation function satisfies the equation:

$$\rho_{jj}(h) - \sum_{i=1}^{3} \xi_i \rho_{jj}(h-i) = 0, \quad h \geq 1$$

If we set the non-diagonal elements of $\boldsymbol{A}$ equal to 0 then we have the case of the constrained TRINAR(1) process. As the univariate marginal distribution of the simple INAR(1) model in terms of the innovation sequence $\{R_t\}$ was expressed i.e., $X_t = \sum_{i=1}^{\infty} \alpha^i \circ R_{t-i}$.

This result can be extended to the case of a constrained multivariate-or trivariate-INAR(1) (Pedeli and Karlis (2011),Pedeli and Karlis (2013c), Pedeli

https://doi.org/10.26219/heal.aueb.5635

and Karlis (2013a)).

$$X_{jt} = \sum_{i=0}^{\infty} \alpha_j^i \circ R_{j,t-i}$$

and the distribution of the TRINAR(1) a process can also be expressed in terms of the multivariate innovation sequence $\boldsymbol{R_t}$, i.e. as

$$X_{jt} = \sum_{i=0}^{\infty} A_j^i \circ R_{j,t-i}$$

This can be useful in order to compute the joint pgf of the multivariate process and through that derive the distribution of the process.

4.2 Model Distribution

The derivation of the joint distribution of a full TRIVAR(1) model is calculated with the same reasoning as in the BINAR(1) by using the statement of Al-Osh and Alzaid (1987). The conditional distribution of the full TRINAR(1) model can be written in a similar form of the equation. but now $\boldsymbol{R}_t$ are interpreted as correlated non-negative integer values random vectors of length *3*.

By defining again $\boldsymbol{B} = \sum_{i=0}^{h-1} \boldsymbol{B}_i = \sum_{i=0}^{h-1} \boldsymbol{A}^i \circ \boldsymbol{R}_{t-i}$ and

$$\boldsymbol{A}^i = \begin{bmatrix} \alpha_i & \beta_i & \gamma_i \\ \delta_i & \epsilon_i & \zeta_i \\ \eta_i & \theta_i & \kappa_i \end{bmatrix},$$

https://doi.org/10.26219/heal.aueb.5635

the matrix $\boldsymbol{B}$ takes the form

$$B^i = \begin{bmatrix} \alpha_i & \beta_i & \gamma_i \\ \delta_i & \epsilon_i & \zeta_i \\ \eta_i & \theta_i & \kappa_i \end{bmatrix} \circ \begin{bmatrix} R_1 \\ R_2 \\ R_3 \end{bmatrix} = \begin{bmatrix} \alpha_i \circ R_1 + \beta_i \circ R_2 + \gamma_i \circ R_3 \\ \delta_i \circ R_1 + \epsilon_i \circ R_2 + \zeta_i \circ R_3 \\ \eta_i \circ R_1 + \theta_i \circ R_2 + \kappa_i \circ R_3 \end{bmatrix}$$

We would like to find the unconditional distribution of $\boldsymbol{B}_i$. By using the binomial thinning property of independence, conditionally on $\boldsymbol{R}$ the following pgf

$$\begin{aligned} G_{\boldsymbol{B}_i|\boldsymbol{R}}(s) = (1-\alpha_i+\alpha_i s_1)^{R_1}(1-\delta+\delta s_1)^{R_1}(1-\eta_i+\eta_i s_1)^{R_1} \\ \times(1-\beta_i+\beta_i s_2)^{R_2}(1-\epsilon_i+\epsilon_i s_2)^{R_2}(1-\theta_i+\theta_i s_2)^{R_2} \\ \times(1-\gamma_i+\gamma_i s_3)^{R_3}(1-\zeta_i+\zeta_i s_3)^{R_3}(1-\kappa_i+\kappa_i s_3)^{R_3} \end{aligned} \tag{4.7}$$

By taking expectation to drop the terms of R, we have that

$$\begin{aligned} G_{\boldsymbol{B}_i}(s_1,s_2,s_3) = G_{\boldsymbol{R}}[(1-\alpha_i+\alpha_i s_1)(1-\delta+\delta s_1)(1-\eta_i+\eta_i s_1), \\ (1-\beta_i+\beta_i s_2)(1-\epsilon_i+\epsilon_i s_2)(1-\theta_i+\theta_i s_2),(1-\gamma_i+\gamma_i s_3)(1-\zeta_i+\zeta_i s_3)(1-\kappa_i+\kappa_i s_3)] \end{aligned}$$

Thus, the joint pgf $X_{1,T+h}, X_{2,T+h}|X_{1T}, Y_{2T}$ takes the form

$$\begin{aligned} G_X(\boldsymbol{s}) = G_{X_{1t},X_{2t},X3t}(s_1,s_2,s_3) = \prod_{i=0}^{\infty} G_{\boldsymbol{R}}[(1-\alpha_i+\alpha_i s_1)(1-\delta+\delta s_1)(1-\eta_i+\eta_i s_1), \\ (1-\beta_i+\beta_i s_2)(1-\epsilon_i+\epsilon_i s_2)(1-\theta_i+\theta_i s_2),(1-\gamma_i+\gamma_i s_3)(1-\zeta_i+\zeta_i s_3)(1-\kappa_i+\kappa_i s_3)] \end{aligned} \tag{4.8}$$

where $\alpha_0 = \beta_0 = \gamma_0 = \delta_0 = \epsilon_0 = \zeta_0 = \eta_0 = \theta_0 = \kappa_0 = 1$. The elements of $\boldsymbol{A}$ can be numerically calculated however the complexity increases strongly.

https://doi.org/10.26219/heal.aueb.5635

4.3 Estimation of full TRINAR(1) Model

Analogously to the bivariate case, the conditional maximum likelihood will be considered. The conditional distribution of the TRINAR(1) model can be described as the convolution of three binomial variables and a trivariate distribution of the innovations, i.e.,

$$\begin{aligned} f_1(k) &= \sum_{i_1=0}^{k}\sum_{j_1=0}^{k-i_1}\sum_{w_1=0}^{k-(i_1+j_1)} \binom{X_{1,t-1}}{i_1}\binom{X_{2,t-1}}{j_1}\binom{X_{3,t-1}}{w_1}\alpha_{11}^{i_1}(1-\alpha_{11})^{X_{1,t-1}-i_1} \\ &\times\alpha_{12}^{j_1}(1-\alpha_{12})^{X_{2,t-1}-j_1}\alpha_{13}^{w_1}(1-\alpha_{13})^{X_{3,t-1}-w_1} \\ f_2(s) &= \sum_{i_2=0}^{s}\sum_{j_2=0}^{s-i_2}\sum_{w_2=0}^{s-(i_2+j_2)} \binom{X_{2,t-1}}{i_2}\binom{X_{1,t-1}}{j_2}\binom{X_{3,t-1}}{w_2}\alpha_{22}^{i_2}(1-\alpha_{22})^{X_{2,t-1}-i_2} \\ &\times\alpha_{21}^{j_2}(1-\alpha_{21})^{X_{1,t-1}-j_2}\alpha_{23}^{w_2}(1-\alpha_{23})^{X_{3,t-1}-w_2} \\ f_3(h) &= \sum_{i_3=0}^{h}\sum_{j_3=0}^{h-i_3}\sum_{w_3=0}^{k-(i_3+j_3)} \binom{X_{3,t-1}}{i_3}\binom{X_{1,t-1}}{j_3}\binom{X_{2,t-1}}{w_3}\alpha_{33}^{i_3}(1-\alpha_{33})^{X_{3,t-1}-i_3} \\ &\times\alpha_{31}^{j_3}(1-\alpha_{31})^{X_{1,t-1}-j_3}\alpha_{32}^{w_3}(1-\alpha_{23})^{X_{2,t-1}-w_3} \end{aligned} \tag{4.9}$$

and with $f_4(r_1, r_2, r_3) = P(R_{1t} = r_1, R_{2t} = r_2, R_{3t} = r_3)$, the conditional distribution takes the form

$$f(\boldsymbol{x_t}|\boldsymbol{x_{t-1}}) = \sum_{k=0}^{g_X}\sum_{s=0}^{g_Y}\sum_{s=0}^{g_Z} f_1(k)f_2(s)f_3(h)f_4(x_t-k, y_t-s, z_t-h) \tag{4.10}$$

where $g_X = min(x_t, x_{t-1})$ $g_Y = min(y_t, y_{t-1})$ and $g_Z = min(z_t, z_{t-1})$ and thus the conditional likelihood takes the form

$$L(\boldsymbol{\theta}|\boldsymbol{x}) = \prod_{t=1}^{T} f(\boldsymbol{x_t}|\boldsymbol{x_{t-1}}, \boldsymbol{\theta}) \tag{4.11}$$

https://doi.org/10.26219/heal.aueb.5635

where $\boldsymbol{\theta}$ is the vector of the unknown parameters and can be estimated via maximasing (4.11).

The trivariate Poisson distribution for the three innovation processes (R_{1t}, R_{2t}, R_{3t}) $TP(\lambda_1, \lambda_2, \lambda_3, \boldsymbol{\phi})$ can be written as

$$\begin{aligned} P(x_{1t}-k, x_{2t}-s, z_{3t}-h) &= \exp\left(-\left(\lambda_1+\lambda_2+\lambda_3+\phi_1+\phi_2+\phi_3\right)\right) \\ &\times \frac{\lambda_1^{x_{1t}-k-m}\lambda_2^{x_{2t}-s-m}\lambda_3^{x_{3t}-h}}{(x_{1t}-k)!(x_{2t}-s)!(x_{3t}-h)!}\frac{\phi_{12}^k}{\phi_{12}!}\frac{\phi_{13}^s}{\phi_{13}!}\frac{\phi_{23}^h}{\phi_{23}!} \end{aligned}$$

where the summation is over the set $C \subset N^3$ defined as

$$C = \left[(\phi_{12}, \phi_{13}, \phi_{23}) \in N^3 : \{\phi_{12}+\phi_{13} \le x_{1t} \cup \{\phi_{12}+\phi_{23} \le x_{2t} \cup \{\phi_{13}+\phi_{23} \le x_{3t} \neq \varnothing\}\right]$$

while marginally each random variable follows a Poisson distribution with parameters λ_1, λ_2, λ_3 respectively. The parameters ϕ_12, ϕ_13 and ϕ_23 are the covariances between R_{1t} with R_{2t}, R_{1t} with R_{3t} and R_{2t} with R_{3t} respectively.

Therefore, in the case that ϕ_{12}, ϕ_{13}, ϕ_{23} the variables are independent and the trivariate Poisson distribution reduces to the product of three independent Poisson distributions, i.e.,

$$f_4(x_{1t}-k, x_{2t}-s, z_{3t}-h) = \exp\left(-(\lambda_1+\lambda_2+\lambda_3)\right)\frac{\lambda_1^{x_{1t}-k-m}\lambda_2^{x_{2t}-s-m}\lambda_3^{x_{3t}-h}}{(x_{1t}-k)!(x_{2t}-s)!(x_{3t}-h)!} \tag{4.12}$$

From now on we will consider independent innovations for the calculation of the likelihood and predictive distribution. However, the full dependence case can be derived analogously to the following results of the independent case.

https://doi.org/10.26219/heal.aueb.5635

Hence, from equations (4.9), (4.10) and (4.12), the conditional likelihood (4.11) takes the form

$$
\begin{aligned}
f(\boldsymbol{x_t}|\boldsymbol{x_{t-1}},\boldsymbol{\theta}) \quad = \quad & \sum_{k=0}^{g_{X_1}}\sum_{s=0}^{g_{X_2}}\sum_{h=0}^{g_{X_3}}\Bigg[\sum_{i_1=0}^{k}\sum_{j_1=0}^{k-i_1}\sum_{w_1=0}^{k-(i_1+j_1)}\binom{X_{1,t-1}}{i_1}\binom{X_{2,t-1}}{j_1}\binom{X_{3,t-1}}{w_1}\alpha_{11}^{i_1}(1-\alpha_{11})^{X_{1,t-1}-i_1} \\
& \times\alpha_{12}^{j_1}(1-\alpha_{12})^{X_{2,t-1}-j_1}\alpha_{13}^{w_1}(1-\alpha_{13})^{X_{3,t-1}-w_1} \\
& \times\sum_{i_2=0}^{s}\sum_{j_2=0}^{s-i_1}\sum_{w_2=0}^{s-(i_2+j_2)}\binom{X_{2,t-1}}{i_2}\binom{X_{1,t-1}}{j_2}\binom{X_{3,t-1}}{w_2}\alpha_{22}^{i_2}(1-\alpha_{22})^{X_{2,t-1}-i_2} \\
& \times\alpha_{21}^{j_2}(1-\alpha_{21})^{X_{1,t-1}-j_2}\alpha_{23}^{w_2}(1-\alpha_{23})^{X_{3,t-1}-w_2} \\
& \times\sum_{i_3=0}^{h}\sum_{j_3=0}^{h-i_3}\sum_{w_3=0}^{h-(i_3+j_3)}\binom{X_{3,t-1}}{i_3}\binom{X_{1,t-1}}{j_3}\binom{X_{2,t-1}}{w_3}\alpha_{33}^{i_3}(1-\alpha_{33})^{X_{3,t-1}-i_3} \\
& \times\alpha_{31}^{j_3}(1-\alpha_{31})^{X_{1,t-1}-j_3}\alpha_{32}^{w_3}(1-\alpha_{23})^{X_{2,t-1}-w_3}\Bigg] \\
& \times\exp-(\lambda_1+\lambda_2+\lambda_3)\frac{\lambda_1^{x_{1t}-k}\lambda_2^{x_{2t}-s}\lambda_3^{x_{3t}-h}}{(x_{1t}-k)!(x_{2t}-s)!(x_{3t}-h)!}
\end{aligned}
\tag{4.13}
$$

with $\boldsymbol{\theta}=\alpha_{11},\alpha_{12},\alpha_{13},\alpha_{21},\alpha_{22},\alpha_{23},\alpha_{31},\alpha_{32},\alpha_{33},\lambda_1,\lambda_2,\lambda_3$ and $g_{X_1}=min(x_{1t},x_{2,t-1})$, $g_{X_2}=min(x_{2t},x_{2,t-1})$ and $g_{X_3}=min(x_{3t},x_{3,t-1})$ as previously. Thus $L(\theta|x)$ can again be maximised in order to derive the ML estimators of $\boldsymbol{\theta}$.

Numerical maximization of (4.13) is straightforward with R as in the bivariate case. The binomial convolution implies finite summation and hence it is easy. The pgf is a polynomial calculation of the convolution as well and thus it is easy derived via polynomial multiplications. Packages in R exist that allows quick polynomial multiplication and hence easy and cheap convolution calculation. However, if we increase the parameters of the model, for instance if we consider the full trivariate model with innovation dependecies, then the computational inefficiency increases.

The EM-Algorithm has been used for a multivariate Poisson by Karlis and

https://doi.org/10.26219/heal.aueb.5635

Meligkotsidou (2005) as well. The idea is the same as with ML, i.e., the model can be considered as the convolution of random variables but only their sum is observed instead of the seperate variables that incorporate the sum.

4.4 Predictive Distribution of the Full TRIVAR(1) model with Poisson Innovations

The derivation of the predictive distribution of a full TRIVAR(1) model is calculated with the same reasoning as in the BINAR(1) by using the statement of Al-Osh and Alzaid (1987). The conditional distribution of the full TRINAR(1) model can be written in a similar form of the equation but now $\boldsymbol{R}_t$ are interpreted as correlated non-negative integer values random vectors of length *3*.

By defining $\boldsymbol{B} = \sum_{i=0}^{h-1} \boldsymbol{B}_i = \sum_{i=0}^{h-1} \boldsymbol{A}^i \circ \boldsymbol{R}_{t-i}$ and

$$\boldsymbol{A}^i = \begin{bmatrix} \alpha_i & \beta_i & \gamma_i \\ \delta_i & \epsilon_i & \zeta_i \\ \eta_i & \theta_i & \kappa_i \end{bmatrix}, \quad \boldsymbol{A}^h = \begin{bmatrix} \alpha_h & \beta_h & \gamma_h \\ \delta_h & \epsilon_h & \zeta_h \\ \eta_h & \theta_h & \kappa_h \end{bmatrix}$$

We would like to find the unconditional distribution of $\boldsymbol{B}_i$. By using the binomial thinning property of independence the pgf of the matrix $\boldsymbol{B}$ is

$$G_{\boldsymbol{B}}(s_1, s_2, s_3) = \prod_{i=0}^{h-1} G_{\boldsymbol{R}}[(1-\alpha_i+\alpha_i s_1)(1-\delta_i+\delta_i s_1)(1-\eta_i+\eta_i s_1),$$
$$(1-\beta_i+\beta_i s_2)(1-\epsilon_i+\epsilon_i s_2)(1-\theta_i+\theta_i s_2), (1-\gamma_i+\gamma_i s_3)(1-\zeta_i+\zeta_i s_3)(1-\kappa_i+\kappa_i s_3)]$$

Thus, the joint pgf $X_{1,T+h}, X_{2,T+h}, X_{3,T+h}|X_{1T}, X_{2T}, X_{3T}$ takes the form

$$G_{X_{t+h}}(\boldsymbol{s}) = \{(1-\alpha_i+\alpha_i s_1)(1-\delta_i+\delta_i s_1)(1-\eta_i+\eta_i s_1)\}^{X_{1T}}$$
$$\{(1-\beta_i+\beta_i s_2)(1-\epsilon_i+\epsilon_i s_2)(1-\theta_i+\theta_i s_2)\}^{X_{1T}}\{(1-\gamma_i+\gamma_i s_3)(1-\zeta_i+\zeta_i s_3)(1-\kappa_i+\kappa_i s_3)\}^{X_{3T}} \tag{4.14}$$

where $\alpha_0 = \beta_0 = \gamma_0 = \delta_0 = \epsilon_0 = \zeta_0 = \eta_0 = \theta_0 = \kappa_0 = 1$. Hence, in order to predict up to a time point $t+h$,

$$G_{X_{T+h}}(s|\boldsymbol{x}_T) = \{(1-\alpha_i+\alpha_i s_1)(1-\delta_i+\delta_i s_1)(1-\eta_i+\eta_i s_1)\}^{X_{1T}}$$
$$\{(1-\beta_i+\beta_i s_2)(1-\epsilon_i+\epsilon_i s_2)(1-\theta_i+\theta_i s_2)\}^{X_{2T}}\{(1-\gamma_i+\gamma_i s_3)(1-\zeta_i+\zeta_i s_3)(1-\kappa_i+\kappa_i s_3)\}^{X_{3T}} G_{\boldsymbol{B}}(s)$$

Hence, the predictive density distribution is the following

$$\begin{aligned} P(\boldsymbol{X}_{t+h} = x|\boldsymbol{X}_T) = \sum_{k=0}^{m_{X_1}} \sum_{s=0}^{m_{X_2}} \Bigg[\sum_{j_{X_1}}^{k} \binom{X_{1T}}{j_{X_1}} \binom{X_{2T}}{k-j_{X_1}} \alpha_h^{j_{X_1}} (1-\alpha_h)^{X_{1T}-j_{X_1}} \beta_h^{k-j_{X_1}} (1-\beta_h)^{X_{2T}-k+j_{X_1}} \\ \times \sum_{j_{X_2}}^{s} \binom{X_{2T}}{j_{X_2}} \binom{X_{1T}}{s-j_{X_2}} \delta_h^{j_{X_2}} (1-\delta_h)^{X_{2T}-j_{X_2}} \gamma_h^{s-j_{X_2}} (1-\gamma_h)^{X_{2T}-s+j_{X_1}} \Bigg] \\ \times P\Bigg(\sum_{i=0}^{h-1} (\alpha_i \circ R_{1,T+h-i} + \beta_i \circ R_{2,T+h-i}) = x_1 - k, \\ \sum_{i=0}^{h-1} (\gamma_i \circ R_{1,T+h-i} + \delta_i \circ R_{2,T+h-i}) = x_2 - s \Bigg) \end{aligned}$$

where $m_{X_1} = min(x_1, x_{1T})$ and $m_2 = min(x, x_{2T})$ and the elements of the matrices $\boldsymbol{A}^i$, $\boldsymbol{A}^h$ can be computed numerically.

4.5 Simulated Example

For the three simulated series, the ratio of their mean to their variance are higher than one which means the series are overdispersed for Poisson data. It seems from the ACF plot (4.2) that the series trigger each other in a high magnitude. Furthermore, in Figure 4.1 it seems that the series progress in the same manner around different means.

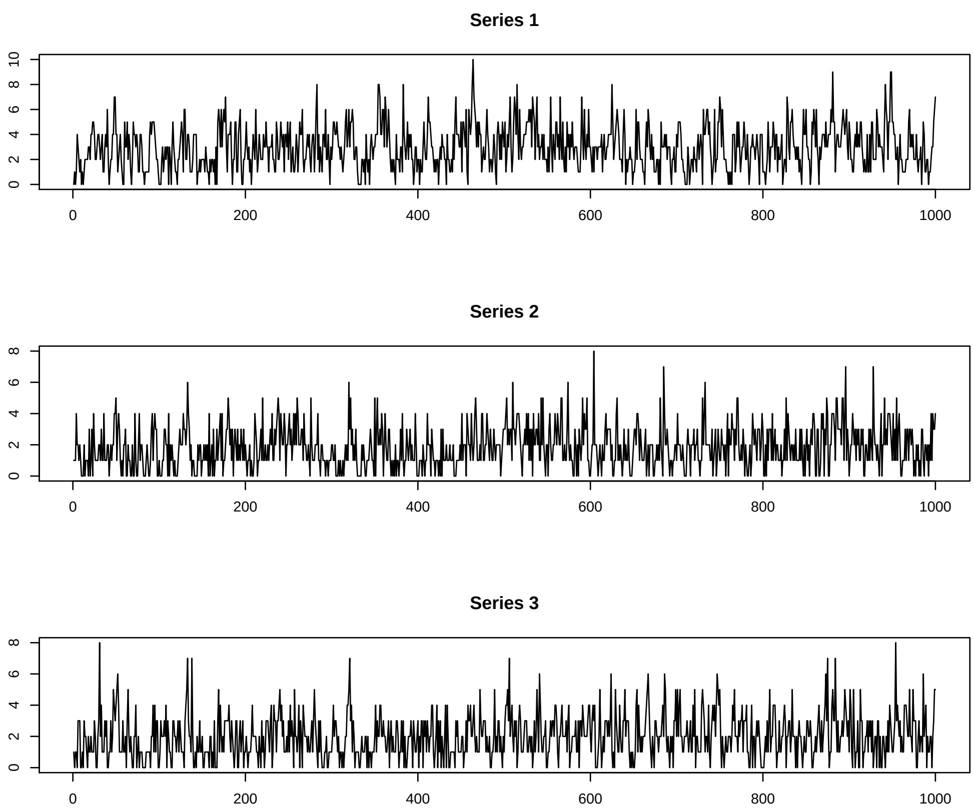

Figure 4.1: Time Series plot for $\alpha_{11} = 0.4$, $\alpha_{12} = 0.1$, $\alpha_{13} = 0.1$, $\alpha_{21} = 0.3$, $\alpha_{22} = 0.2$, $\alpha_{23} = 0.1$,$\alpha_{31} = 0.1$,$\alpha_{32} = 0.2$, $\alpha_{33} = 0.3$ and Poisson innovations with $\lambda_1 = 1$, $\lambda_2 = 0.8$, $\lambda_3 = 0.9$. The mean (variance) of series 1, 2, 3 is 2.973(3.09), 1.844(1.87), 1.955(2.04) respectively.

https://doi.org/10.26219/heal.aueb.5635

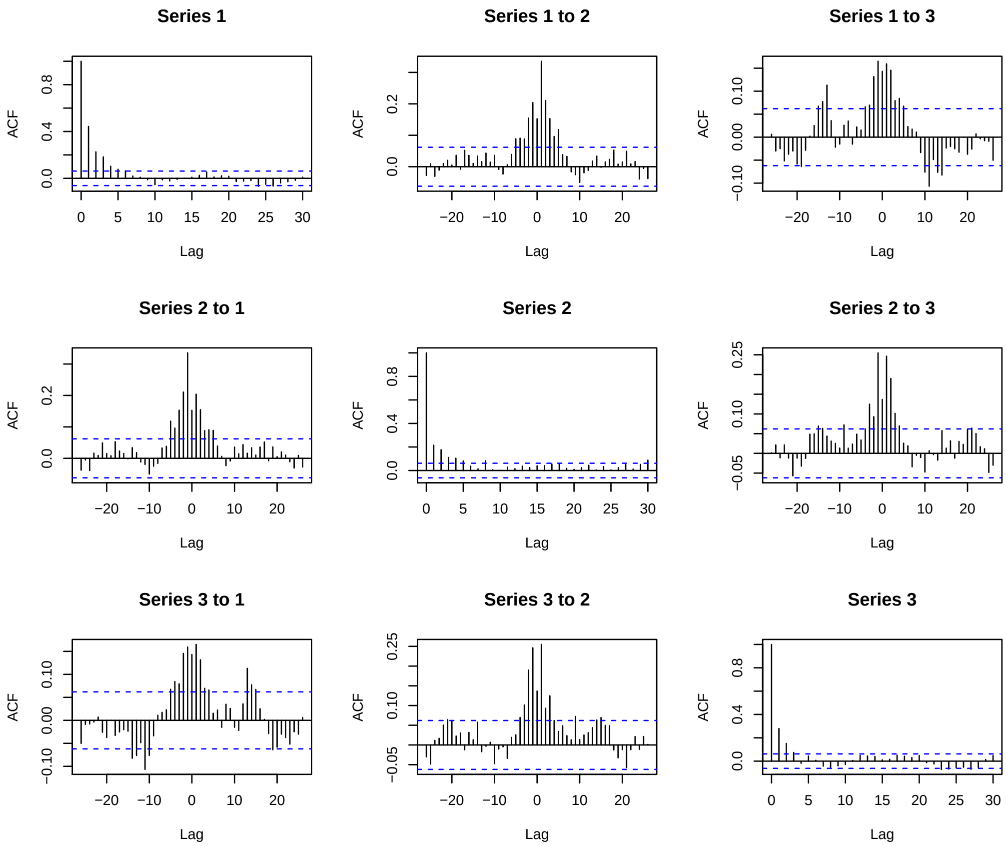

Figure 4.2: ACF and CCF plots for $\alpha_{11} = 0.4$, $\alpha_{12} = 0.1$, $\alpha_{13} = 0.1$, $\alpha_{21} = 0.3$, $\alpha_{22} = 0.2$, $\alpha_{23} = 0.1$,$\alpha_{31} = 0.1$,$\alpha_{32} = 0.2$, $\alpha_{33} = 0.3$ and Poisson innovations with $\lambda_1 = 1$, $\lambda_2 = 0.8$, $\lambda_3 = 0.9$

https://doi.org/10.26219/heal.aueb.5635

Chapter 5

Application on high Seismicity Data

5.1 Introduction

The area of the North Aegean Sea (NAS), Western Greece, is characterized by very high seismicity due to the westward propagation of the North Anatolian Fault (NAF) into the Aegean. It comprises the North Aegean Tough (NAT) and its parallel branches that are the dominant tectonic features. The epicentres of all the strong earthquakes that occurred in the area belong to one of these tectonic settings.

In this study the NAS is defined as the area bounded by the rectangle with latitude range $38.4^o - 40.5^o$N and longitude range $23.5^o - 27.0^o$E. The complexity and the extent of the NAS has led many authors to divide it into sub-areas based on their particular features (Leptokaropoulos et al. (2012), Rhoades et al. (2010)) and study the migration of seismicity from one sub-area to another (Papadopoulos et al. (2002)).

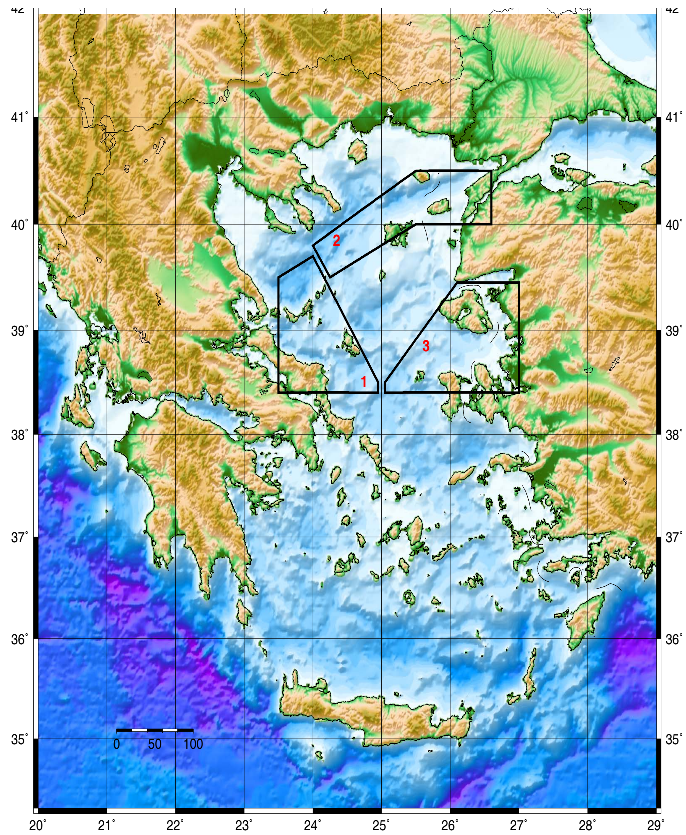

Figure 5.1

In our study 3 sub-areas were selected (Figure 5.1) that do not overlap and correspond to different tectonic settings. A common Local completeness magnitude, based on the Gutenberg and Richter relationship Gutenberg and Richter (1944), equal or greater to $Ml \geq 3.5$, was considered for the three sub-areas, for the time period from 1964 to 2008. Only shallow events, with depth less than $60Km$ were selected from the catalogue of the Institute of Geodymanics, National Observatory of Athens (www.gein.noa.gr).

The joint modelling of the monthly number of earthquakes that occurred

https://doi.org/10.26219/heal.aueb.5635

in the 3 sub-areas are the data used in this application. The fact that the 3 sub-areas are part of a common, more complex system imply that the 3 time series are correlated.

	Mean	*Variance*	*Variance/Mean*	*1st order autocorrelation*
Area 1	1.090	17.741	16.276	0.246
Area 2	0.879	12.136	13.805	0.159
Area 3	1.790	19.149	10.697	0.264

Table 5.1: Mean, Variance, Ratio of Variance to the Mean and 1st lag autocorrelation for the three areas

In addition, data collected in successive time intervals tend to be dependent. The time series of earthquake counts, for the 3 sub-areas are shown in Figure 5.2. The mean number (variances) of earthquakes per month is 1.09 (17.75) for sub-area 1, 0.88 (12.14) for sub-area 2 and 1.79 (19.15) for sub-area 3. Area 3 is the most active one, with seismicity rate almost equal to the sum of the seismicity rates of areas 2 and 3.

The correlation between series 1 and 2 is 0.02, between series 1 and 3 is 0.088, while between series 2 and 3 is -0.013. Moreover, it is obvious from Figure (5.3) that the relationship between the sub-Area 1 and sub-Area 3 may be related to their past lags of each other (cross autocorrelation up to 0.492 for the 1st lag) while the sub areas 2 with 1 and 3 (from 0.02 to 0.073 and from 0.008 to 0.01 respectively) does not seem to be related to the early lags of each other. Additionally, all areas seem to be correlated to the 1st lag of themselves with 0.246, 0.159, 0.264 for Areas 1,2 and 3 respectively.

https://doi.org/10.26219/heal.aueb.5635

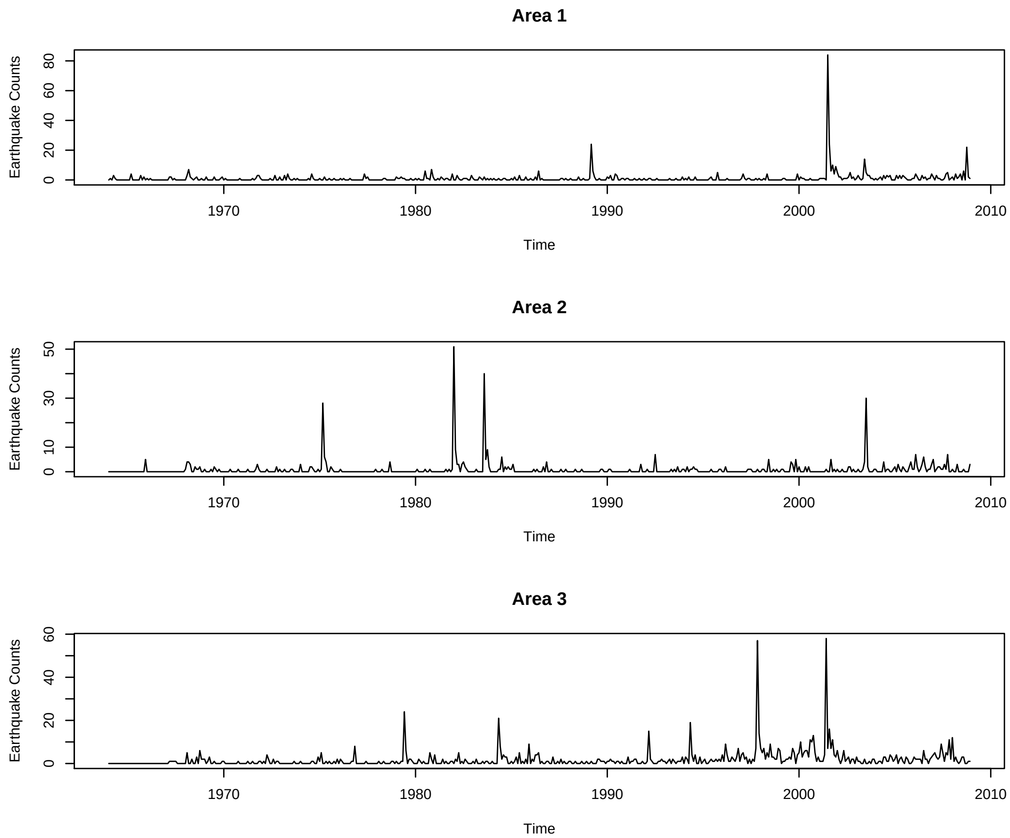

Figure 5.2: Time series plots for the earthquake counts occurring monthly in the 3 sub-areas from January 1964 to December 2008

The high variance values that appear in all three time series imply for overdispersion. The ability of the TRINAR(1) models to describe the serial correlation that appear in the univariate series of earthquake counts as also to describe the serial correlation between them, and the fact that the data of the 3 time series were selected from areas that share the same seismotectonic features leads to the joint modelling of the 3 time series and thus to a Trivariate INAR(1) model. However, the constrained Poisson TRINAR(1) is unable to capture overdispersion in the data.

https://doi.org/10.26219/heal.aueb.5635

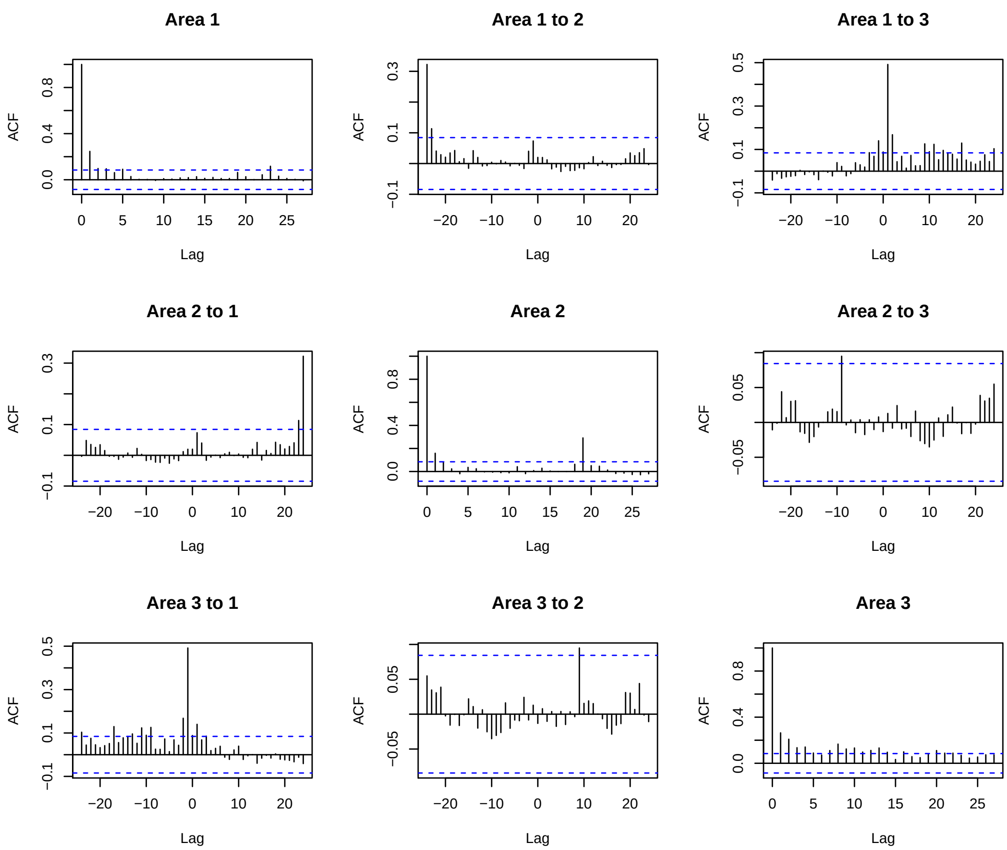

Figure 5.3: ACF and CCF plots for the earthquake counts occurring monthly in the 3 sub-areas from January 1964 to December 2008

5.2 Model Fitting

For the analysis the Trivariate INAR(1) model with independent Poisson innovations was selected. This means that we examined

https://doi.org/10.26219/heal.aueb.5635

$$
\begin{aligned}
X_t &= \alpha_{11}X_{t-1} + \alpha_{12}Y_{t-1} + \alpha_{13}Z_{t-1} + R_{1t} \\
Y_t &= \alpha_{21}X_{t-1} + \alpha_{22}Y_{t-1} + \alpha_{23}Z_{t-1} + R_{2t} \\
Z_t &= \alpha_{31}X_{t-1} + \alpha_{32}Y_{t-1} + \alpha_{33}Z_{t-1} + R_{3t}
\end{aligned}
$$

where X_t, Y_t, Z_t are the monthly earthquake counts in Area 1, 2 and 3 respectively. If we set an element of the matrix $\boldsymbol{A}$, as $\alpha_{ij} = 0$, $i \neq j$, e.g. $\alpha_{31} = 0$, then we can interpret it as X_t is being triggered by Z_t but not vice versa.

In order to conduct a primal analysis considering which areas trigger each other but also themselves, the full model was examined and estimated. This can be translated as considering that all the three sub-areas trigger each other at time t according to time $t-1$.

Moreover, with probability α_{12} and α_{13} Area 1 at time $t-1$ will trigger at time t Area 2 and Area 3 respectively while with probability α_{11} will Area 1 will trigger itself. Additionally, with probability α_{21} and α_{23} Area 2 at time $t-1$ will trigger at time t Area 1 and Area 3 respectively while with probability α_{11} will Area 2 will trigger itself. Finally, with probability α_{31} and α_{32} Area 3 at time $t-1$ will trigger at time t Area 1 and Area 2 respectively while with probability α_{11} will Area 3 will trigger itself. When the estimation gives non-zero probabilities α_{ij} for the triggering effect between two different areas, that means that these two areas are tectonic neighbours. In the case that a triggering probability α_{ij} for two different areas is zero, there is an indication that these two areas are not considered tectonic neighbours.

Afterwards, an all subsets model selection was used in order to resort to

https://doi.org/10.26219/heal.aueb.5635

the best model that can represent the data in the best way. Akaike Information Criterion (AIC), (Akaike, 1974) was considered for the choice of the best model but also one might consider whether all the estimated parameters have such a great impact to be chosen in favour of a simplest model.

5.2.1 Estimation of the Full Model

For the estimation of the Full Model the conditional likelihood maximasation technique for a Trivariate INAR(1) with Poisson Innovations was considered. The estimated parameters of the full model are

$$\hat{\boldsymbol{A}} = \begin{bmatrix} 0.209 & 0.070 & 0.003 \\ 0.004 & 0.200 & 0.025 \\ 0.158 & 0.094 & 0.251 \end{bmatrix}$$

,

$$\hat{\boldsymbol{\lambda}} = \quad (0.588, 0.420, 0.919)$$

with standard errors

$$s.e.\hat{(}\boldsymbol{A}) = \begin{bmatrix} 0.040 & 0.044 & 0.014 \\ 0.022 & 0.044 & 0.023 \\ 0.054 & 0.057 & 0.042 \end{bmatrix}$$

,

$$s.e.(\hat{\boldsymbol{\lambda}}) = \quad (0.055, 0.048, 0.078)$$

The estimations imply that

1. The first Area at time point t will be triggered from the the earthquakes that happened at time point $t-1$, by the same Area, Area 2 and Area 3 with probability 0.209, 0.070 and 0.003 respectively.

2. The second Area at time point t will be triggered from the earthquakes that were counted at time point $t-1$ by itself, Area 1 and Area 3 with

https://doi.org/10.26219/heal.aueb.5635

probability 0.158, 0.200, 0.025 respectively and

3. Area 3 will be triggered from the earthquakes that happened at time point $t-1$ by itself, Area 1 and Area 2 with probability 0.158, 0.094, 0.251 respectively.

Obviously the diagonal elements of $\boldsymbol{A}$ should be kept in the model due to their high values of probability. However, elements α_{13} and α_{21} present lower values close to 1% and thus, intuitively it could be assumed that they are not needed in the model. Additionally, α_{31} shows a significant amount of probability (close to 16%) and should be kept in the model considering that the third Area has a moderate contribution to the earthquakes of Area 1.

The model equation of the full estimated model is provided bellow:

$$\begin{aligned} X_t &= 0.209X_{t-1} + 0.070Y_{t-1} + 0.003Z_{t-1} \\ Y_t &= 0.004X_{t-1} + 0.200Y_{t-1} + 0.025Z_{t-1} \\ Z_t &= 0.158X_{t-1} + 0.094Y_{t-1} + 0.251Z_{t-1} \end{aligned}$$

The log-likelihood and AIC values for the model are evaluated as $log(L(\boldsymbol{\theta}|\boldsymbol{X})) = -2876.592$ and $AIC = 5777.184$.

5.2.2 Model Selection

The all subsets model selection method was considered. Therefore all the possible 512 models were fitted and the ones with the best combination of AIC and Likelihood were considered.

In Figure 5.4 grey indicates whether a parameter is included -or considered as

https://doi.org/10.26219/heal.aueb.5635

Model	α_{11}	α_{12}	α_{13}	α_{21}	α_{22}	α_{23}	α_{31}	α_{32}	α_{33}	λ_1	λ_2	λ_3
First	0.209	0.070	0.014	-	0.200	0.037	0.157	-	0.251	0.569	0.399	0.924
Second	0.209	-	0.021	0.017	0.200	0.025	0.157	0.094	0.251	0.602	0.396	0.919
Full	0.209	0.070	0.003	0.004	0.200	0.025	0.158	0.094	0.251	0.558	0.420	0.919
Proposed	0.209	0.044	-	-	0.200	0.025	0.158	0.094	0.251	0.588	0.421	0.919

Table 5.2: Parameter estimates for the best three models

non-zero- in the model, while white indicates the latter. The model containing all the variables maximises the Log-Likelihood, with $\ell(\boldsymbol{\theta}) = -2876.592$ while the model without α_{31} and α_{32} minimises the likelihood among the best ones with a $\ell(\boldsymbol{\theta}) = 2894.269$. The top models that were also considered with a small difference from the full one, where the ones with $\alpha_{12} = 0$ and $\alpha_{21} = \alpha_{32} = 0$ with $\ell(\boldsymbol{\theta}) = -2876.864$ and $\ell(\boldsymbol{\theta}) = -2877.111$.

Furthermore, the triggering probabilities of each area that affect themselves are considered in every model, which makes sense since in the full model estimation the probabilities were of high value. However, when it was expected that α_{31} would be kept in every model since it was estimated at the level of 0.158 probability, the one that is chosen in every model apart from the diagonal elements was α_{23} which was estimated with a triggering probability of 0.025 in the full model. One reason for this type of diversity is that since the triggering probabilities are not of a high value, they are considered not to have a high impact in terms of contribution.

Nonetheless, the addition of parameters and thus the increased complexity of the model should be considered when choosing the right model. Hence, considering Figure 5.5 which incorporates the complexity of a model by putting a penalty for the number of parameters, the best model considers $\alpha_{21} = \alpha_{32} = 0$ with $AIC = 5774.222$, which was also in the top three mod-

Model	α_{11}	α_{12}	α_{13}	α_{21}	α_{22}	α_{23}	α_{31}	α_{32}	α_{33}	λ_1	λ_2	λ_3
First	0.044	0.049	0.02	-	0.043	0.022	0.054	-	0.039	0.053	0.048	0.079
Second	0.043	-	0.022	0.023	0.053	0.022	0.060	0.058	0.037	0.057	0.047	0.079
Full	0.040	0.044	0.014	0.022	0.044	0.023	0.054	0.057	0.042	0.055	0.048	0.078
Proposed	0.045	0.044	-	-	0.040	0.021	0.053	0.055	0.035	0.051	0.046	0.090

Table 5.3: Standard errors of the transformed estimates $logit(a_{ij})$ and λ_i for the best three models.

els according to log-likelihood. The second and third best models are with $\alpha_{12} = 0$ ($AIC = 5775.708$) and the full model ($AIC = 5777.184$) respectively.

The best models' estimations are given in Table 5.2.2.The model equation for the best model according to AIC is given bellow:

$$\begin{aligned} X_t &= 0.209X_{t-1} + 0.070Y_{t-1} + 0.014Z_{t-1} \\ Y_t &= 0.200Y_{t-1} + 0.025\ Z_{t-1} \\ Z_t &= 0.157X_{t-1} + 0.251\ Z_{t-1} \end{aligned}$$

By taking into consideration the first model and assigning $\alpha_{21} = \alpha_{32} = 0$ the estimated triggering probability from Area 1 to Area 3 is of very low value. Similarly, by fitting the second model the equation that is derived is:

$$\begin{aligned} X_t &= 0.209\ X_{t-1} + 0.021Z_{t-1} \\ Y_t &= 0.017X_{t-1} + 0.200Y_{t-1} + 0.025Z_{t-1} \\ Z_t &= 0.157\ X_{t-1} + 0.094Y_{t-1} + 0.251\ Z_{t-1} \end{aligned}$$

https://doi.org/10.26219/heal.aueb.5635

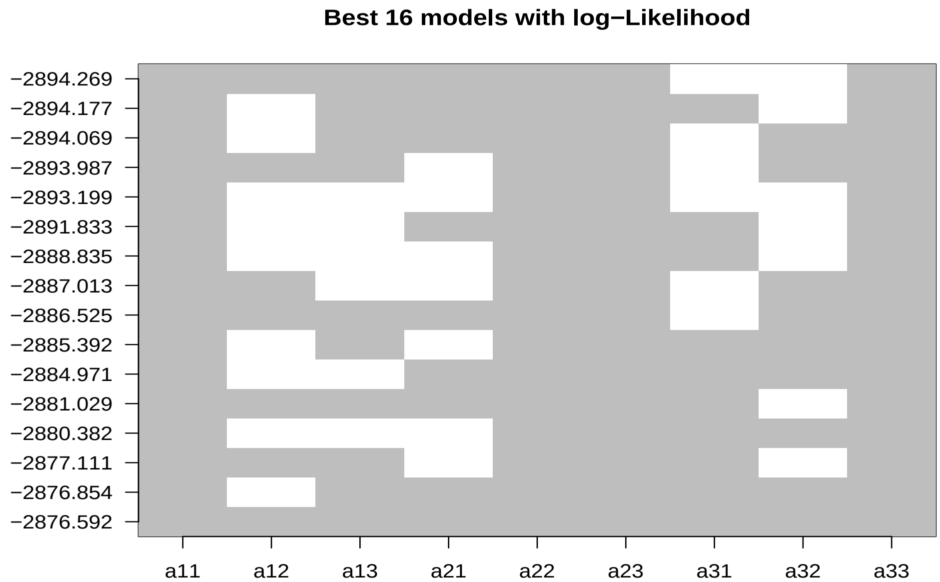

Figure 5.4: Best 16 ranked models with ascending order from up tp the bottom. The y-axis represents the log-likelihood of each model and the x-axis represents the parameters that are considered for that model. Gray boxes indicate when a parameter is considered as non-zero.

The second best model assumes that Area 1 does not trigger Area 2, the estimated probability for triggering Area 1 by Area 2 is also of very low value. In the full model both α_{13} and α_{21} are estimated with probabilities less than 0.01 while the elements that we are removing with the best models seemed to have more contribution to the model.

Consequently, we decided to choose the model which removes the triggering probabilities α_{13} and α_{21} and keep inside the rest of the parameters. We will propose the model that although it is not in the best 16 models in terms of likelihood or AIC, it is a more rational choice according to the estimations

https://doi.org/10.26219/heal.aueb.5635

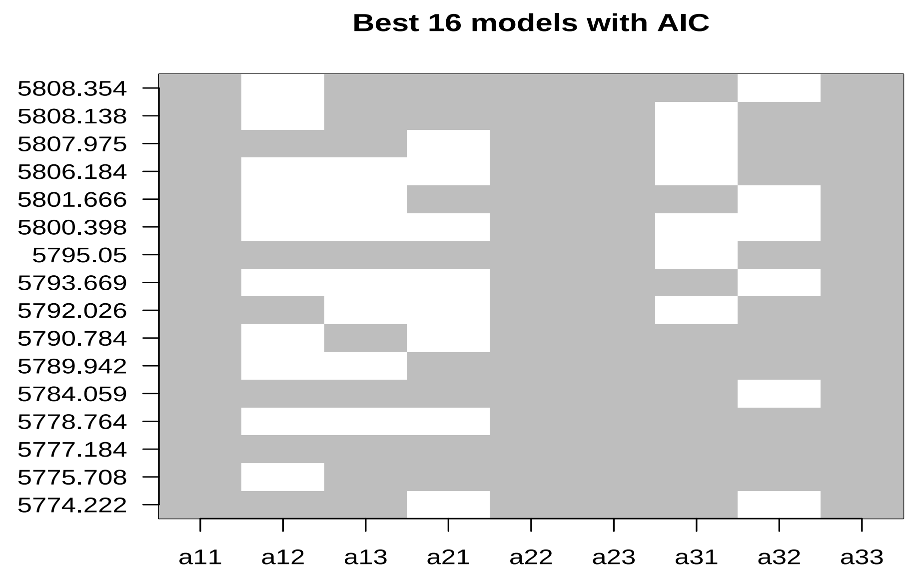

Figure 5.5: Best 16 ranked models with ascending order from up to the bottom. The y-axis represents the AIC value of each model and x-axis represents the parameters that are considered for that model. Gray boxes indicate when a parameter is considered as non-zero.

from the rest of the models. Thus, the equation of the proposed and final model is:

$$\begin{aligned} X_t &= 0.209X_{t-1} + 0.072Y_{t-1} \\ Y_t &= 0.200Y_{t-1} + 0.025Z_{t-1} \\ Z_t &= 0.158X_{t-1} + 0.094Y_{t-1} + 0.251Z_{t-1} \end{aligned}$$

The proposed model with $AIC = 5774.222$ implies that there is a triggering

https://doi.org/10.26219/heal.aueb.5635

probability from time point $t-1$ to time point t from Areas 1, 2 and 3 to themselves with probabilities 0.209,0.200 and 0.251 respectively. There is an indication that Area 3 is a moderate tectonic neighbour of Area 1 since it triggers it with probability 0.123. Area 3 only triggers area 2 with a small probability of 0.013 while Area 2 affect only Area 3 with a probability of 0.076. Finally, Area 2 affects also Area 1 with probability 0.059.

Summing up the results, Area 3 seems to be the tectonic neighbour which has the highest impact among the three. Furthermore, Area 2 and 3 seem to affect more the next month's earthquakes happening in the same location compared to Area 1 but still Area 1 has also has a high contribution.

https://doi.org/10.26219/heal.aueb.5635

Chapter 6

Conclusion

The results of the model are easy to communicate. However, it is not so easy to think ways to summarise the results which can then be discussed with safety professionals and experts. Whether or not it is correct to define an area as risky after an earthquake occurred on its neighbour, it requires further analysis and local knowledge. With the appearance of an earthquake of a small magnitude of one area, there is a really small probability of having an appearance of a catastrophic earthquake in the other area. Hence, for the safety professionals this may be false alarm. Although our analyses indicates the triggering effects from the one area to the other, in terms of how one should use this result to avoid deaths of civilians or to predict whether a house will be destroyed is limited. If it is deemed correct to do so, then the statistical method is of great use because it identifies potential risk before earthquakes happen.

The method chosen for this dissertation is one of many and other approaches may yield interesting results. Furthermore, this problem should be examined also spatially, in the sense of trying to show spatial association of the neighbourhood structure and either use them for smoothing, inference and predictions for the possible outcome of a seismic activity. The spatial de-

https://doi.org/10.26219/heal.aueb.5635

pendence could help us recognise a pattern that is not only temporal between the areas as we showed but also a spatial one. Thus, one could also use spatio-temporal point processes to find new triggering probabilities of spatio-temporal dependence and predict possible earthquakes.

Fitting the models was computationally very slow. For that reason the innovation series was considered uncorrelated in order to increase the computational efficiency. One major problem was the choice of the initial values for the maximisation of the conditional likelihood. We also had to transform the parameters' initial values so that the algorithm could converge. These hurdles make the all subsets model selection computationally expensive while we account only on likelihood based methods. Likelihood-based criteria approaches with respect to their asymptotic correspondence with cross-validation, may not always hold. However, there has been a limited amount of work on appropriate model selection techniques on Vector Autoregressive models which mostly are based on regularisation techniques, such as in Ren and Zhang (2013) and Ren et al. (2013) where adaptive lasso and two-step shrinkage methods are used respectively. Additionally, one could also consider as well MCMC methods where we can take advantage of conjugate priors. Therefore, one would not need to conduct a thorough model selection but only infer on the parameters. Finally, a further combined methodology that could be explore is by either placing a Laplace prior to the logistic transformation of the parameters, or a spike and slab prior on their variance, since the operator matrix itself is sparse.

The literature review skimmed the surface of what is an absolutely a newly upcoming field. Analysis of temporal correlation did finally result in a usable conclusion but left many questions unanswered. Much future work could also be done in Bayesian computational techniques, such as using Approximate Bayesian Computation Inference since a multivariate Poisson with depen-

https://doi.org/10.26219/heal.aueb.5635

dencies can be considered a non-closed form or apply sequential methods for online estimation.

One major omission is the lack of any model validation. An easy method would simply have been to leave out the last year of data and use it to perform out-of-sample model validation. Whichever method is used, validation is needed in order to understand not only whether the model has correctly estimated the triggering probabilities, but also to investigate its predictive capability.

https://doi.org/10.26219/heal.aueb.5635

Bibliography

Akaike, H. (1974). A new look at the statistical model identification, *Automatic Control, IEEE Transactions on*, **19 (6)**: 716–723.

Al-Osh, M. A. and Aly, E.-E. A. (1992). First order autoregressive time series with negative binomial and geometric marginals, *Communications in Statistics-Theory and Methods*, **21 (9)**: 2483–2492.

Al-Osh, M. A. and Alzaid, A. A. (1987). First-order integer-valued autoregressive (inar (1)) process, *Journal of Time Series Analysis*, **8 (3)**: 261–275.

Alzaid, A. and Al-Osh, M. (1988). First-order integer-valued autoregressive (inar (1)) process: Distributional and regression properties, *Statistica Neerlandica*, **42 (1)**: 53–61.

Alzaid, A. and Al-Osh, M. (1993). Some autoregressive moving average processes with generalized poisson marginal distributions, *Annals of the Institute of Statistical Mathematics*, **45 (2)**: 223–232.

Billingsley, P. (1961). *Statistical inference for Markov processes*, vol. 2, University of Chicago Press Chicago.

Boudreault, M. and Charpentier, A. (2011). Multivariate integer-valued autoregressive models applied to earthquake counts, *arXiv preprint arXiv:1112.0929*.

https://doi.org/10.26219/heal.aueb.5635

Brandt, P. T., Williams, J. T., Fordham, B. O. and Pollins, B. (2000). Dynamic modeling for persistent event-count time series, *American Journal of Political Science*, pp. 823–843.

Brännäs, K. and Hellström, J. (2001). Generalized integer-valued autoregression, *Econometric Reviews*, **20 (4)**: 425–443.

Brannas, K. and Nordstrom, J. (2000). A bivariate integer valued allocation model for guest nights in hotels and cottages.

Cooper, B. and Lipsitch, M. (2004). The analysis of hospital infection data using hidden markov models, *Biostatistics*, **5 (2)**: 223–237.

Deng, J. and Sykes, L. R. (1997). Evolution of the stress field in southern california and triggering of moderate-size earthquakes: A 200-year perspective, *Journal of Geophysical Research: Solid Earth (1978–2012)*, **102 (B5)**: 9859–9886.

Dewald, L. S., Lewis, P. A. and Mckenzie, E. (1989). A bivariate first-order autoregressive time series model in exponential variables (bear (1)), *Management science*, **35 (10)**: 1236–1246.

Eduarda Da Silva, M. and Oliveira, V. L. (2004). Difference equations for the higher-order moments and cumulants of the inar (1) model, *Journal of Time Series Analysis*, **25 (3)**: 317–333.

Ellingwood, B. R. and Wen, Y.-K. (2005). Risk-benefit-based design decisions for low-probability/high consequence earthquake events in mid-america, *Progress in Structural Engineering and Materials*, **7 (2)**: 56–70.

Ermoliev, Y. M., Ermolieva, T. Y., MacDonald, G. and Norkin, V. (2000). Stochastic optimization of insurance portfolios for managing exposure to catastrophic risks, *Annals of Operations Research*, **99 (1-4)**: 207–225.

https://doi.org/10.26219/heal.aueb.5635

Franke, J. and Rao Subba, T. (1993). Multivariate first-order integer-valued autoregressions.

Freeland, R. K. and McCabe, B. P. (2004). Forecasting discrete valued low count time series, *International Journal of Forecasting*, **20 (3)**: 427–434.

Genest, C. and Nešlehová, J. (2007). A primer on copulas for count data, *Astin Bulletin*, **37 (02)**: 475–515.

Gourieroux, C. and Jasiak, J. (2003). Nonlinear innovations and impulse responses with application to var sensitivity, *Les Cahiers du CREF of HEC Montréal Working Paper*, **(03-08)**.

Gutenberg, B. and Richter, C. F. (1944). Frequency of earthquakes in california, *Bulletin of the Seismological Society of America*, **34 (4)**: 185–188.

Heinen, A. and Rengifo, E. (2007). Multivariate autoregressive modeling of time series count data using copulas, *Journal of Empirical Finance*, **14 (4)**: 564–583.

Jacobs, P. A. and Lewis, P. A. (1978). Discrete time series generated by mixtures ii: asymptotic properties, *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 222–228.

Jung, R. C. and Tremayne, A. (2006a). Binomial thinning models for integer time series, *Statistical Modelling*, **6 (2)**: 81–96.

Jung, R. C. and Tremayne, A. R. (2006b). Coherent forecasting in integer time series models, *International Journal of Forecasting*, **22 (2)**: 223–238.

Karlis, D. and Meligkotsidou, L. (2005). Multivariate poisson regression with covariance structure, *Statistics and Computing*, **15 (4)**: 255–265.

Karlis, D. and Pedeli, X. (2013). Flexible bivariate inar (1) processes using copulas, *Communications in Statistics-Theory and Methods*, **42 (4)**: 723–740.

https://doi.org/10.26219/heal.aueb.5635

Latour, A. (1997). The multivariate ginar (p) process, *Advances in Applied Probability*, pp. 228–248.

Leptokaropoulos, K., Papadimitriou, E., Orlecka-Sikora, B. and Karakostas, V. (2012). Seismicity rate changes in association with the evolution of the stress field in northern aegean sea, greece, *Geophysical Journal International*, **188 (3)**: 1322–1338.

MacDonald, I. L. and Zucchini, W. (1997). *Hidden Markov and other models for discrete-valued time series*, vol. 110, CRC Press.

McCabe, B. and Martin, G. M. (2005). Bayesian predictions of low count time series, *International Journal of Forecasting*, **21 (2)**: 315–330.

McKenzie, E. (1985). Some simple models for discrete variate time series1.

McKenzie, E. (1988). Some arma models for dependent sequences of poisson counts, *Advances in Applied Probability*, pp. 822–835.

McKenzie, E. (2003). Ch. 16. discrete variate time series, *Handbook of statistics*, **21**: 573–606.

Nelsen, R. (2006). An introduction to copulas, 2nd, *New York: Springer-Science Business Media*.

Nikoloulopoulos, A. K. and Karlis, D. (2009). Modeling multivariate count data using copulas, *Communications in Statistics-Simulation and Computation*, **39 (1)**: 172–187.

Orfanogiannaki, K., Karlis, D. and Papadopoulos, G. (2010). Identifying seismicity levels via poisson hidden markov models, *Pure and applied geophysics*, **167 (8-9)**: 919–931.

Papadimitriou, E. E. and Sykes, L. R. (2001). Evolution of the stress field in the northern aegean sea (greece), *Geophysical Journal International*, **146 (3)**: 747–759.

https://doi.org/10.26219/heal.aueb.5635

Papadopoulos, G., Ganas, A. and Plessa, A. (2002). The skyros earthquake (mw 6.5) of 26 july 2001 and precursory seismicity patterns in the north aegean sea, *Bulletin of the Seismological Society of America*, **92 (3)**: 1141–1145.

Papazachos, B. (1990). Seismicity of the aegean and surrounding area, *Tectonophysics*, **178 (2)**: 287–308.

Pavlopoulos, H. and Karlis, D. (2006). On time series of overdispersed count: Modelling, inference, simulation and prediction, Tech. rep., Technical Report, Department of Statistics Athens University of Economics and Business, 221, 1-38.

Pedeli, X. and Karlis, D. (2011). A bivariate inar (1) process with application, *Statistical modelling*, **11 (4)**: 325–349.

Pedeli, X. and Karlis, D. (2013a). On composite likelihood estimation of a multivariate inar (1) model, *Journal of Time Series Analysis*, **34 (2)**: 206–220.

Pedeli, X. and Karlis, D. (2013b). On estimation of the bivariate poisson inar process, *Communications in Statistics-Simulation and Computation*, **42 (3)**: 514–533.

Pedeli, X. and Karlis, D. (2013c). Some properties of multivariate inar (1) processes, *Computational Statistics & Data Analysis*, **67**: 213–225.

Quoreshi, A. S. (2006). Bivariate time series modeling of financial count data, *Communications in Statistics-Theory and Methods*, **35 (7)**: 1343–1358.

Ren, Y., Xiao, Z. and Zhang, X. (2013). Two-step adaptive model selection for vector autoregressive processes, *Journal of Multivariate Analysis*, **116**: 349–364.

https://doi.org/10.26219/heal.aueb.5635

Ren, Y. and Zhang, X. (2013). Model selection for vector autoregressive processes via adaptive lasso, *Communications in Statistics-Theory and Methods*, **42 (13)**: 2423–2436.

Rhoades, D., Papadimitriou, E., Karakostas, V., Console, R. and Murru, M. (2010). Correlation of static stress changes and earthquake occurrence in the north aegean region, *Pure and applied geophysics*, **167 (8-9)**: 1049–1066.

Steutel, F. and Van Harn, K. (1979). Discrete analogues of self-decomposability and stability, *The Annals of Probability*, pp. 893–899.

Steutel, F. and Van Harn, K. (1986). Discrete operator-selfdecomposabiuty and queueing networks, *Stochastic Models*, **2 (2)**: 161–169.

Varotsos, P. and Lazaridou, M. (1991). Latest aspects of earthquake prediction in greece based on seismic electric signals, *Tectonophysics*, **188 (3)**: 321–347.