

Statistical Methods for Generalized Integer Autoregressive  
Processes

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree Doctor  
of Philosophy in the Graduate School of The Ohio State University

By

Pashmeen Kaur, B.Sc., M.S.

Graduate Program in Department of Statistics

The Ohio State University

2024

Dissertation Committee:

Peter F Craigmile, Co-Advisor

Christopher Hans, Co-Advisor

Yoonkyung Lee

© Copyright by

Pashmeen Kaur

2024

## Abstract

A popular and flexible time series model for counts is the generalized integer autoregressive process of order  $p$ ,  $\text{GINAR}(p)$ . These Markov processes are defined using thinning operators evaluated on past values of the process along with a discretely-valued innovation process. This class includes the commonly used  $\text{INAR}(p)$  process, defined with binomial thinning and Poisson innovations.  $\text{GINAR}$  processes can be used in a variety of settings, including modeling time series with low counts, and allow for more general mean-variance relationships, capturing both over- or under-dispersion.

While there are many thinning operators and innovation processes given in the literature, less focus has been spent on comparing statistical inference and forecasting procedures over different choices of  $\text{GINAR}$  process. We provide an extensive study of exact and approximate inference and forecasting methods that can be applied to a wide class of  $\text{GINAR}(p)$  processes with general thinning and innovation parameters. We discuss the challenges of exact estimation when  $p$  is larger. We summarize and extend asymptotic results for estimators of process parameters, and present simulations to compare small sample performance, highlighting how different methods compare.

$\text{GINAR}$  processes assume stationarity of the process, which may not be an appropriate assumption for many real-world applications. Hence, we introduce a process for non-stationary count time series called the time-varying generalized integer autoregressive process ( $\text{TV-GINAR}(p)$ ), which allows for time-varying parameters modeled via basis functions. We

introduce statistical properties, discuss estimation strategies, and statistical inference for this class of processes. We present simulation studies for the TV-GINAR( $p$ ) process and illustrate this methodology by fitting GINAR and TV-GINAR processes to a disease surveillance series and patient scores dataset.

To my parents,  
Mrs. Inderjeet Kaur and Mr. Gurcharanjit Singh

## Acknowledgments

I would like to thank my advisor, Dr. Peter Craigmile, for his patient guidance, knowledge, and kindness. His immense knowledge and passion for statistics have been inspiring throughout my journey as a PhD student. I would also like to thank my co-advisor, Dr. Christopher Hans, for his support and feedback. I am also grateful to my dissertation committee, Dr. Lee, and the Department of Statistics faculty at The Ohio State University for their invaluable teachings and mentorship.

In addition, I would like to thank my friends and family, both near and far, for their love and encouragement throughout graduate school. I am also eternally grateful for my sisters. Simran, thank you for being my pillar of strength; your love and guidance have been invaluable, and I am blessed to have an elder sister like you. Ninochka, thank you for being both a best friend and a sister; your support and encouragement have meant the world to me.

I want to thank my parents, without whom none of this would have been possible. There are not enough words to describe how thankful I am for you both and all the sacrifices you have made so that I can live my dreams. I am forever grateful for your unconditional love and support. Lastly, I want to thank my biggest cheerleader, my grandfather, whose wisdom and encouragement have been a guiding light in my life.

## Vita

2019 .....B.Sc.(Honors), Applied Mathematics,  
Economics  
University of California, Los Angeles

2021 .....M.S., Statistics  
The Ohio State University

## Fields of Study

Major Field: Statistics

# Table of Contents

	Page
Abstract . . . . .	ii
Dedication . . . . .	iv
Acknowledgments . . . . .	v
Vita . . . . .	vi
List of Tables . . . . .	x
List of Figures . . . . .	xvi
1. Introduction . . . . .	1
1.1 Research Problem and Contributions . . . . .	2
1.2 Dissertation Structure . . . . .	4
2. GINAR( $p$ ) Processes . . . . .	6
2.1 Defining Thinning Operators . . . . .	6
2.2 Defining GINAR( $p$ ) Processes . . . . .	9
2.3 Example GINAR( $p$ ) Processes . . . . .	11
2.4 Statistical Properties . . . . .	12
2.5 Transition Probabilities . . . . .	18
3. Estimation of GINAR( $p$ ) Processes . . . . .	22
3.1 Estimation Methods . . . . .	22
3.1.1 Conditional maximum likelihood . . . . .	23
3.1.2 Yule-Walker . . . . .	29
3.1.3 Conditional least squares . . . . .	31

3.1.4	Pseudo maximum likelihood . . . . .	32
3.1.5	Saddlepoint methods . . . . .	33
3.1.6	Spectral-based Whittle estimation . . . . .	34
3.1.7	Other methods . . . . .	35
3.2	Inference . . . . .	36
3.2.1	Confidence regions . . . . .	36
3.2.2	Model selection . . . . .	37
3.3	Forecasting for GINAR( $p$ ) processes . . . . .	38
3.3.1	Literature Review . . . . .	38
3.3.2	A simulation-based forecasting method . . . . .	41
4.	Simulation Studies for GINAR( $p$ ) Processes . . . . .	43
4.1	Estimating GINAR(1) process parameters . . . . .	44
4.2	Estimating GINAR(2) process parameters . . . . .	53
4.3	Estimating GINAR(4) process parameters . . . . .	62
4.4	Coverage Simulations . . . . .	71
4.5	Conclusions . . . . .	74
5.	Time Varying GINAR( $p$ ) Processes . . . . .	76
5.1	Introduction . . . . .	76
5.2	Defining TV-GINAR( $p$ ) Processes . . . . .	79
5.3	Statistical properties of TV-GINAR( $p$ ) processes . . . . .	81
6.	Estimation for TV-INAR( $p$ ) Processes . . . . .	86
6.1	Estimation Methods . . . . .	86
6.1.1	Conditional Maximum Likelihood (CML) . . . . .	86
6.1.2	Conditional Least Squares (CLS) . . . . .	87
6.1.3	Pseudo Maximum Likelihood . . . . .	88
6.2	Confidence Intervals . . . . .	88
6.3	Model Selection . . . . .	92
6.4	Simulation Studies for TV-GINAR( $p$ ) Processes . . . . .	93
6.4.1	Estimating TV-GINAR(1) process parameters . . . . .	93
6.4.2	Estimating TV-GINAR(2) process parameters . . . . .	98
6.4.3	Conclusions . . . . .	103
7.	Applications . . . . .	104
7.1	Disease Surveillance . . . . .	104
7.2	Schizophrenic Patient Data . . . . .	114
7.3	Conclusions . . . . .	121

8.	Discussion and Future Work . . . . .	123
8.1	Discussion . . . . .	123
8.2	Future Work . . . . .	125
8.2.1	Forecasting . . . . .	125
8.2.2	Time-varying parameter curves . . . . .	126
8.2.3	Multivariate Processes and Other Extensions . . . . .	128

## List of Tables

Table	Page
2.1 Examples of GINAR processes found in literature. NA values indicate that a closed form expression for the marginal distribution does not exist. . . . .	13
2.2 Examples of GINAR processes found in literature. NA values indicate that a closed form expression for the marginal distribution does not exist (continued).	14
4.1 Estimated bias, SD, and RMSE when estimating the parameters of a Po-INAR(1) process with $\alpha_1 = 0.5$ and $\mu_\epsilon = 1$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.002 for estimating $\alpha_1$ and 0.004 for estimating $\mu_\epsilon$ . . . . .	45
4.2 Estimated bias, SD, and RMSE when estimating the parameters of a NB-INAR(1) process with $\alpha_1 = 0.5$ , $r = 1$ and $\mu_\epsilon = 1$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.006 for estimating $\alpha_1$ , 0.05 for estimating $r$ , and 0.05 for estimating $\mu_\epsilon$ . . . . .	46
4.3 Estimated bias, SD, and RMSE when estimating the parameters of a Geom-INAR(1) process with $\alpha_1 = 0.5$ and $\mu_\epsilon = 1$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.002 for estimating $\alpha_1$ and 0.007 for estimating $\mu_\epsilon$ . . . . .	47
4.4 Estimated bias, SD, and RMSE when estimating the parameters of a Po-INAR(1) process with $\alpha_1 = 0.5$ and $n = 500$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.001 for estimating $\alpha_1$ and 0.031 for estimating $\mu_\epsilon$ . . . . .	50
4.5 Estimated bias, SD, and RMSE when estimating the parameters of a NB-INAR(1) process with $\alpha_1 = 0.5$ , $r = 1$ and $n = 500$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.001 for estimating $\alpha_1$ , 0.004 for estimating $r$ and 0.036 for estimating $\mu_\epsilon$ . . . . .	51

4.6	Estimated bias, SD, and RMSE when estimating the parameters of a Geom-INAR(1) process with $\alpha_1 = 0.5$ and $n = 500$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.001 for estimating $\alpha_1$ and 0.050 for estimating $\mu_\epsilon$ . . . . .	52
4.7	Estimated bias, SD, and RMSE when estimating the parameters of a Po-INAR(1) process with $\mu_\epsilon = 1$ and $n = 500$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.002 for estimating $\alpha_1$ and 0.011 for estimating $\mu_\epsilon$ . . . . .	53
4.8	Estimated bias, SD, and RMSE when estimating the parameters of a NB-INAR(1) process with $\mu_\epsilon = 1$ , $n = 500$ and $r = 1$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.005 for estimating $\alpha_1$ , and 0.012 for estimating $r$ , and 2.665 for estimating $\mu_\epsilon$ . . . . .	54
4.9	Estimated bias, SD, and RMSE when estimating the parameters of a Geom-INAR(1) process with $\mu_\epsilon = 1$ and $n = 500$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.001 for estimating $\alpha_1$ and 0.004 for estimating $\mu_\epsilon$ . . . . .	55
4.10	Estimated bias, SD, and RMSE when estimating the parameters of a Po-INAR(2) process with $\alpha_1 = 0.2$ , $\alpha_2 = 0.4$ and $\mu_\epsilon = 1$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.002 for estimating $\alpha_1$ , 0.002 for estimating $\alpha_2$ and 0.006 for estimating $\mu_\epsilon$ . . . . .	56
4.11	Estimated bias, SD, and RMSE when estimating the parameters of a NB-INAR(2) process with $\alpha_1 = 0.2$ , $\alpha_2 = 0.4$ , $r = 1$ and $\mu_\epsilon = 1$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.002 for estimating $\alpha_1$ , 0.002 for estimating $\alpha_2$ , 0.007 for estimating $\mu_\epsilon$ and 0.039 for estimating $r$ . . . . .	57
4.12	Estimated bias, SD, and RMSE when estimating the parameters of a Geom-INAR(2) process with $\alpha_1 = 0.2$ , $\alpha_2 = 0.4$ and $\mu_\epsilon = 1$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.002 for estimating $\alpha_1$ , 0.003 for estimating $\alpha_2$ and 0.008 for estimating $\mu_\epsilon$ . . . . .	58
4.13	Estimated bias, SD, and RMSE when estimating the parameters of a Po-INAR(2) process with $\alpha_1 = 0.2$ , $\alpha_2 = 0.4$ and $n = 500$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.001 for estimating $\alpha_1$ , 0.001 for estimating $\alpha_2$ and 0.024 for estimating $\mu_\epsilon$ . . . . .	59

4.14	Estimated bias, SD, and RMSE when estimating the parameters of a NB-INAR(2) process with $\alpha_1 = 0.2$ , $\alpha_2 = 0.4$ , $r = 1$ and $n = 500$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.001 for estimating $\alpha_1$ , 0.001 for estimating $\alpha_2$ , 0.135 for estimating $\mu_\epsilon$ and 0.012 for estimating $r$ . . . . .	60
4.15	Estimated bias, SD, and RMSE when estimating the parameters of a Geom-INAR(2) process with $\alpha_1 = 0.2$ , $\alpha_2 = 0.4$ and $n = 500$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.002 for estimating $\alpha_1$ , 0.006 for estimating $\alpha_2$ , and 0.104 for estimating $\mu_\epsilon$ . . . . .	61
4.16	Estimated bias, SD, and RMSE when estimating the parameters of a Geom-INAR(4) process with $\alpha_1 = 0.1$ , $\alpha_2 = 0.2$ , $\alpha_3 = 0.3$ , $\alpha_4 = 0.1$ , and $\mu_\epsilon = 1$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.002 for estimating $\alpha_1$ , 0.002 for estimating $\alpha_2$ , 0.002 for estimating $\alpha_3$ , 0.003 for estimating $\alpha_4$ and 0.014 for estimating $\mu_\epsilon$ . . . . .	63
4.17	Estimated bias, SD, and RMSE when estimating the parameters of a Po-INAR(4) process with $\alpha_1 = 0.1$ , $\alpha_2 = 0.2$ , $\alpha_3 = 0.3$ , $\alpha_4 = 0.1$ , and $\mu_\epsilon = 1$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.002 for estimating $\alpha_1$ , 0.003 for estimating $\alpha_2$ , 0.002 for estimating $\alpha_3$ , 0.003 for estimating $\alpha_4$ and 0.014 for estimating $\mu_\epsilon$ . . . . .	64
4.18	Estimated bias, SD, and RMSE when estimating the parameters of a NB-INAR(4) process with $\alpha_1 = 0.1$ , $\alpha_2 = 0.2$ , $\alpha_3 = 0.3$ , $\alpha_4 = 0.1$ , and $\mu_\epsilon = 1$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.002 for estimating $\alpha_1$ , 0.002 for estimating $\alpha_2$ , 0.002 for estimating $\alpha_3$ , 0.003 for estimating $\alpha_4$ and 0.014 for estimating $\mu_\epsilon$ , and 0.681 for estimating $r$ . . . . .	65
4.19	Estimated bias, SD, and RMSE when estimating the parameters of a Po-INAR(4) process with $\alpha_1 = 0.1$ , $\alpha_2 = 0.2$ , $\alpha_3 = 0.3$ , $\alpha_4 = 0.1$ , and $\mu_\epsilon = 1$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.001 for estimating $\alpha_1$ , 0.001 for estimating $\alpha_2$ , 0.001 for estimating $\alpha_3$ , 0.001 for estimating $\alpha_4$ and 0.004 for estimating $\mu_\epsilon$ . . . . .	66
4.20	Estimated bias, SD, and RMSE when estimating the parameters of a Geom-INAR(4) process with $\alpha_1 = 0.1$ , $\alpha_2 = 0.2$ , $\alpha_3 = 0.3$ , $\alpha_4 = 0.1$ , and $\mu_\epsilon = 1$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.001 for estimating $\alpha_1$ , 0.001 for estimating $\alpha_2$ , 0.001 for estimating $\alpha_3$ , 0.001 for estimating $\alpha_4$ and 0.004 for estimating $\mu_\epsilon$ . . . . .	67

4.21	Estimated bias, SD, and RMSE when estimating the parameters of a NB-INAR(4) process with $\alpha_1 = 0.1$ , $\alpha_2 = 0.2$ , $\alpha_3 = 0.3$ , $\alpha_4 = 0.1$ , and $\mu_\epsilon = 1$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.001 for estimating $\alpha_1$ , 0.001 for estimating $\alpha_2$ , 0.001 for estimating $\alpha_3$ , 0.001 for estimating $\alpha_4$ , 0.005 for estimating $\mu_\epsilon$ and 0.015 for estimating $r$ . . . . .	68
4.22	Estimated bias, SD, and RMSE when estimating the parameters of a Po-INAR(4) process with $\alpha_1 = 0.1$ , $\alpha_2 = 0.2$ , $\alpha_3 = 0.3$ and $\alpha_4 = 0.1$ and $\mu_\epsilon = 10$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.002 for estimating $\alpha_1$ , 0.003 for estimating $\alpha_2$ , 0.001 for estimating $\alpha_3$ , 0.006 for estimating $\alpha_4$ , and 0.102 for estimating $\mu_\epsilon$ . . . . .	69
4.23	Estimated bias, SD, and RMSE when estimating the parameters of a Geom-INAR(4) process with $\alpha_1 = 0.1$ , $\alpha_2 = 0.2$ , $\alpha_3 = 0.3$ and $\alpha_4 = 0.1$ and $\mu_\epsilon = 10$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.001 for estimating $\alpha_1$ , 0.003 for estimating $\alpha_2$ , 0.001 for estimating $\alpha_3$ , 0.006 for estimating $\alpha_4$ , and 0.102 for estimating $\mu_\epsilon$ . . . . .	70
4.24	Estimated bias, SD, and RMSE when estimating the parameters of a NB-INAR(4) process with $\alpha_1 = 0.1$ , $\alpha_2 = 0.2$ , $\alpha_3 = 0.3$ and $\alpha_4 = 0.1$ and $\mu_\epsilon = 10$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.001 for estimating $\alpha_1$ , 0.001 for estimating $\alpha_2$ , 0.001 for estimating $\alpha_3$ , 0.001 for estimating $\alpha_4$ , 0.042 for estimating $\mu_\epsilon$ and 0.009 for estimation $r$ . . . . .	71
4.25	A comparison of the estimated coverages for a 95% confidence interval for $\mu_\epsilon$ and $\alpha_1$ in the Po-INAR(1) model, as $n$ is varied, for select estimation methods. The number of replicates is 10,000. The maximum standard error for $\mu_\epsilon$ and $\alpha_1$ is 0.43 and 0.45 respectively. The bold values indicate coverage probabilities no different than 0.95, on the basis of a significance test with level 0.05. . . . .	72
4.26	A comparison of the estimated coverages for a 95% confidence interval for $\mu_\epsilon$ and $\alpha_1$ in the NB-INAR(1) model, as $n$ is varied, for select estimation methods. The number of replicates is 10,000. The maximum standard error for $\mu_\epsilon$ and $\alpha_1$ is 0.43 and 0.45 respectively. The bold values indicate coverage probabilities no different than 0.95, on the basis of a significance test with level 0.05. . . . .	72

4.27	A comparison of the estimated coverages for a 95% confidence interval for $\mu_\epsilon$ and $\alpha_1$ in the Geom-INAR(1) model, as $n$ is varied, for select estimation methods. The number of replicates is 10,000. The maximum standard error for $\mu_\epsilon$ and $\alpha_1$ is 0.43 and 0.45 respectively. The bold values indicate coverage probabilities no different than 0.95, on the basis of a significance test with level 0.05. . . . .	72
6.1	Estimated bias, SD, and RMSE when estimating the parameters of a TV-PoINAR(1) process with $\eta_{1,0} = 0.1$ , $\eta_{1,1} = 0.8$ and $\mu_\xi = 1$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.002 for estimating $\mu_\xi$ , 0.005 for estimating $\eta_{1,0}$ , and 0.006 for estimating $\eta_{1,1}$ . . . . .	96
6.2	Estimated bias, SD, and RMSE when estimating the parameters of a TV-PoINAR(1) process with $\eta_{1,0} = 0.1$ , $\eta_{1,1} = 0.8$ , $\eta_{2,0} = 0$ and $\eta_{2,1} = 0.5$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.004 for estimating $\hat{\eta}_{2,0}$ , 0.007 for estimating $\hat{\eta}_{2,1}$ , 0.006 for estimating $\eta_{1,0}$ and 0.011 for estimating $\eta_{1,1}$ . . . . .	96
6.3	Estimated bias, SD, and RMSE when estimating the parameters of a TV-PoINAR(2) process with $\eta_{1,0} = -0.1$ , $\eta_{1,1} = -0.8$ , $\eta_{2,0} = 0.1$ , $\eta_{2,1} = 0.6$ and $\mu_\xi = 1$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.003 for estimating $\mu_\xi$ , 0.152 for estimating $\eta_{1,0}$ and 1.613 for estimating $\eta_{1,1}$ , 0.006 for estimating $\eta_{2,0}$ and 0.009 for estimating $\eta_{2,1}$ . . . . .	100
6.4	Estimated bias, SD, and RMSE when estimating the parameters of a TV-PoINAR(2) process with $\eta_{1,0} = -0.1$ , $\eta_{1,1} = -0.8$ , $\eta_{2,0} = 0.1$ , $\eta_{2,1} = 0.6$ and $\mu_\xi = 1$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.002 for estimating $\mu_\xi$ , 0.006 for estimating $\eta_{1,0}$ and 0.011 for estimating $\eta_{1,1}$ , 0.005 for estimating $\eta_{2,0}$ and 0.005 for estimating $\eta_{2,1}$ . . . . .	100
6.5	Estimated bias, SD, and RMSE when estimating the parameters of a TV-PoINAR(2) process with $\eta_{1,0} = -0.1$ , $\eta_{1,1} = -0.8$ , $\eta_{2,0} = 0.1$ , $\eta_{2,1} = 0.6$ , $\eta_{3,0} = 0$ and $\eta_{3,1} = 0.5$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.006 for estimating $\eta_{3,0}$ , 0.010 for estimating $\eta_{3,1}$ , 0.082 for estimating $\eta_{1,0}$ and 1.615 for estimating $\eta_{1,1}$ , 0.008 for estimating $\eta_{2,0}$ and 0.015 for estimating $\eta_{2,1}$ . . . . .	101

6.6	Estimated bias, SD, and RMSE when estimating the parameters of a TV-PoINAR(1) process with $\eta_{1,0} = 0.1$ , $\eta_{1,1} = 0.8$ and $\mu_\xi = 1$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.004 for estimating $\eta_{3,0}$ , 0.007 for estimating $\eta_{3,1}$ , 0.013 for estimating $\eta_{1,0}$ , 0.017 for estimating $\eta_{1,1}$ , 0.006 for estimating $\eta_{2,0}$ and 0.010 for estimating $\eta_{2,1}$ . . . . .	101
7.1	A comparison of different model selection and diagnostic criteria for different GINAR processes fit to the meningococcal disease series. All fitted processes contain a yearly seasonality component. Values in bold indicate the processes for which the AIC and RMSE values are the smallest. . . . .	106
7.2	Parameter estimates and 95% confidence intervals for the NB-INAR(4) model shown in Table 7.1. . . . .	107
7.3	A comparison of different model selection and diagnostic criteria for different GINAR processes fit to the meningococcal disease series. All fitted processes contain a yearly seasonality component and trend. Values in bold indicate the processes for which the AIC and RMSE values are the smallest. . . . .	110
7.4	A comparison of different model selection and diagnostic criteria for different GINAR processes fit to the perceptual score series. Values in bold indicate the processes for which the AIC and RMSE values are the smallest. . . . .	117

## List of Figures

Figure	Page
6.1 (a) Marginal mean as a function of time for the 10,000 realizations in gray and the smoothed result over time, using loess smoothing, shown in black. (b) Marginal variance as a function of time for the 10,000 realizations in gray and the smoothed result over time shown in black. . . . .	94
6.2 (a) Parametrization of $\{\alpha_1(u) : u \in [0, 1]\}$ for TV-GINAR(1) process simulation. (b) Parametrization of $\{\mu_\xi(u) : u \in [0, 1]\}$ for TV-GINAR(1) process simulation.	95
6.3 Parametrization of $\{\alpha_1(u) : u \in [0, 1]\}$ and $\{\alpha_2(u) : u \in [0, 1]\}$ for TV-GINAR(2) process simulation. . . . .	99
6.4 Spectral density for specified $t$ values for TV-GINAR(2) process when innovation parameter is constant. . . . .	102
7.1 (a) Time series plot of the weekly meningococcal disease cases; (b) sample ACF and (c) sample PACF of residuals after removing a yearly sinusoidal seasonality term. . . . .	105
7.2 (a) Time series plot of the weekly meningococcal disease cases; (b) sample ACF and (c) sample PACF of residuals after removing a yearly sinusoidal seasonality and trend term. . . . .	109
7.3 The estimated dependence parameter curves, with their 95% confidence intervals, for the TV-NBINAR(3)-Poly model fit to the meningococcal disease cases in Germany from 2001–2017. The dashed line represents the constant dependence parameter estimates from the NB-INAR(3) model. . . . .	112
7.4 Spectral density for specified $t$ values for TV-NBINAR(3)-Poly process not accounting for trend and seasonality. . . . .	113

7.5	The estimated innovation mean parameter curve, with their 95% confidence intervals, for the TV-NBINAR(3)-Poly model fit to the meningococcal disease cases in Germany from 2001–2017. . . . .	114
7.6	(a) Time series plot of the daily score achieved by the patient with a dashed line at day 61, when the patient received the treatment; (b) sample ACF and (c) sample PACF of residuals after removing the polynomial trend. . . . .	115
7.7	(a), (b) The estimated $\alpha_1(u)$ and $\mu_\xi(u)$ parameter curve for the TV-NBINAR(1)-Basis model respectively; (c), (d) The estimated $\alpha_1(u)$ and $\mu_\xi(u)$ parameter curve for the TV-NBINAR(1)-Poly model respectively ; (e), (f) The estimated $\alpha_1(u)$ and $\mu_\xi(u)$ parameter curve for the TV-NBINAR(1)-2 model respectively; (g), (h) The estimated $\alpha_1(u)$ and $\mu_\xi(u)$ parameter curve for the TV-NBINAR(1)-Ind model respectively. . . . .	119
7.8	(a), (b) The estimated $\alpha_1(u)$ and $\mu_\xi(u)$ parameter curve for the TV-NBINAR(1)-Basis-6 model respectively. . . . .	120
7.9	Spectral density for specified days values for TV-NBINAR(1)-Basis process not accounting for polynomial trend. . . . .	121

## Chapter 1: Introduction

Discrete-valued time series appear in many settings such as epidemiology, social science, and finance. Example count series include the number of patients admitted in a hospital every day, the daily cases of a rare infectious disease, the monthly number of car accidents in a region, and yearly insurance claims. While traditional time series models such as the autoregressive integrated moving average (ARIMA) processes can approximately explain such data, they do not preserve discreteness and accurately fail to capture different mean-variance relationships. Also, when counts are low (e.g., when dealing with incidence of rare diseases and events) or depict zero-inflation, ARIMA processes provide a poor approximation to the distribution of the data. Further, processes for continuous-valued time series fail to produce integer forecasts, leading to the creation of arbitrary forecasting methods for discrete-valued series.

One popular class of processes for non-negative discrete-valued series, the integer autoregressive process of order one (INAR(1)), was introduced independently by [McKenzie \[1985\]](#) and [Al-Osh and Alzaid \[1987\]](#). The INAR(1) process can be considered the integer-valued analogue of the AR(1) process because they have a matching autocorrelation structure. This process was later extended to the order  $p$  case by [Alzaid and Al-Osh \[1990\]](#) and [Jin-Guan and Yuan \[1991\]](#), who defined the process using different dependence structures for the counting

series.  $\text{INAR}(p)$  processes are defined using a binomial thinning operation and a discretely-valued usually Poisson distributed innovation sequence, which guarantees that the series is discretely-valued. There are various other models for modeling time series of counts, for example, generalized state space models, observation driven models like, and hidden Markov models. For a review refer to [Davis et al. \[2021\]](#).

While it is possible to generalize  $\text{INAR}(p)$  processes to capture other marginal distributions (see e.g., [McKenzie \[1985\]](#), [Al-Osh and Aly \[1992\]](#), and [Ristić et al. \[2012\]](#) for extensions to negative binomial marginal distributions), we can generalize further. In this dissertation we focus our attention on generalized integer autoregressive models ( $\text{GINAR}(p)$ ), introduced by [Latour \[1998\]](#). This class of models extends the  $\text{INAR}(p)$  process by allowing for different thinning operators. The inclusion of these thinning operators allows for more general time series structure: we can further vary the choice of marginal distributions, and can vary over- and under-dispersion structures (e.g., [Jung et al. \[2005\]](#), [Weiß \[2013\]](#), [Bourguignon and Vasconcellos \[2015b\]](#), [Huang and Zhu \[2021\]](#)). Estimation of these  $\text{GINAR}(p)$  processes can be carried out in a variety of ways, including conditional maximum likelihood, conditional least squares and Yule-Walker, to name a few. We discuss estimation, forecasting and inference for these processes in later chapters.

## 1.1 Research Problem and Contributions

In this section we discuss the contributions of this dissertation. In the literature for integer autoregressive time series processes, many variations of the process have been introduced by varying the two main components of the process: the thinning operation and the innovation distribution. Subsequently, estimation methods, asymptotic theory and applications are presented for these individual processes. In this dissertation, our focus is on the generalized

INAR( $p$ ) process that can allow for general thinning and innovation distributions. We present statistical properties of this process, and computationally efficient methods for obtaining transition probabilities. We also study a variety of different estimation methods and show how to extend those to the GINAR( $p$ ) process. We also present asymptotic theory for the conditional maximum likelihood estimator of a GINAR( $p$ ) process.

Another contribution of the dissertation is that we provide extensive simulation studies comparing the various estimation methods for larger model orders and different thinning and innovation distributions. In the literature we did not find such an extensive comparison. Our goal for these simulation studies is to understand any patterns in the performance of estimation methods for the GINAR( $p$ ) process parameters. For instance, does the small sample performance of an estimation method stay consistent when we change the process order or change the innovation distribution? These simulations provide a clearer understanding of the advantages and disadvantages of different estimation methods and help us understand cases where they perform well or not. We also provide R code that can be used to easily implement these estimation methods for GINAR( $p$ ) processes. R code that accompanies this dissertation can be downloaded from <https://github.com/petercraigmile/GINAR/> [R Core Team, 2024]. Furthermore, we also present a discussion of methodologies for forecasting GINAR( $p$ ) processes and discuss methods for model selection.

We further study a new class of GINAR( $p$ ) processes, called the time-varying GINAR( $p$ ) process (TV-GINAR( $p$ )) to model non-stationary processes. The GINAR( $p$ ) processes discussed thus far assume stationarity, however in many time series applications this assumption is not valid. Research in modeling non-stationary count time series using GINAR processes has involved allowing for a time-varying innovation mean parameter, varying the thinning distribution and accommodating structural breaks. However, we want to focus on allowing

the dependence parameter to change with time as well via generalized linear functions of basis vectors. The use of these basis vectors gives the TV-GINAR( $p$ ) process additional flexibility to model a wide variety of non-stationary processes. We present statistical properties for this time-varying class of stochastic processes, discuss estimation methods, and provide simulations to study small sample performance. We also discuss techniques for forecasting and model selection.

To illustrate these time series processes and statistical inference methodology we present applications in disease surveillance and a schizophrenic patient scores dataset.

## 1.2 Dissertation Structure

The outline of the dissertation is as follows. In Chapter 2 we begin by defining the thinning operator and introduce GINAR( $p$ ) processes. We present a literature review of processes that fit in this GINAR( $p$ ) framework. Next, we present statistical properties of this process along with defining the transition probabilities. In particular, we present an efficient way for the calculation of transitional probabilities.

In Chapter 3, we discuss the estimation and inference for GINAR( $p$ ) processes. We present the following estimation methods: conditional maximum likelihood (CML), Yule-Walker (Y-W), conditional least squares (CLS), pseudo maximum likelihood, spectral-based Whittle estimation and saddlepoint methods. We show how to extend these methods to estimating the parameters of GINAR( $p$ ) processes and present asymptotic theory for the CML and pseudo maximum likelihood estimator. We also discuss inference methodologies like building confidence intervals and model selection.

In Chapter 4, we present extensive simulation studies for the GINAR(1), GINAR(2) and GINAR(4) processes. Our contribution lies in the extensive range of simulations we conduct,

along with the diverse array of estimation methods we consider. We consider different thinning and innovation distributions to understand how the aforementioned estimation methods perform under changing parameters, model order, and model specification. We also present coverage simulations using the confidence interval methods proposed in Chapter 3.

In Chapter 5, we provide a literature review of current methods for modeling non-stationary discrete-valued time series and introduce the TV-GINAR( $p$ ) process. We present statistical properties and transition probabilities for the TV-GINAR( $p$ ) process. In Chapter 6 we discuss estimation of parameters and statistical inference for these processes. In particular, we present the following estimation methods: CML, pseudo maximum likelihood, and CLS. We further show how to use the Delta method to build confidence intervals for the process parameters and present simulation studies for the TV-GINAR(1) and TV-GINAR(2) processes. We consider different series lengths and consider two cases. In the first case, we allow for time-varying dependence parameters while the innovation sequence parameters are constant over time. In the second case, we allow both, the dependence parameter and innovation sequence parameters, to vary with time. Lastly, we present model selection simulations for these two processes.

In Chapter 7 we apply these class of processes (GINAR( $p$ ) and TV-GINAR( $p$ )) to modeling a disease surveillance series and a schizophrenic patient scores dataset. Lastly, in Chapter 8 we summarize the findings of this dissertation and outline possible future work.

## Chapter 2: GINAR( $p$ ) Processes

In this chapter we give a formal definition of generalized integer autoregressive processes of order  $p$ , GINAR( $p$ ) for short. We begin by defining the thinning operator and providing examples of commonly used thinning operators. We then define the GINAR( $p$ ) process and provide a literature review of the different examples of processes that fit under this GINAR( $p$ ) framework. We further derive statistical properties of the thinning operator and GINAR( $p$ ) process and discuss two methods for calculating transition probabilities.

### 2.1 Defining Thinning Operators

We begin by defining the **generalized thinning operator** which is an immediate extension of the binomial thinning operator introduced by [Steutel and van Harn \[1979\]](#) (we introduce binomial thinning as a special case below). A thinning operation is a probabilistic operations that can be applied to random integer values, that always lead to integer values. They circumvent the scalar multiplication of the ARIMA models, ensuring we always have integer values. As we will see later, these thinning operators allow for autoregressive (AR) correlation structures but generate count time series, while allowing for different distributional relationships and possible over-dispersion or under-dispersion.

**Definition 2.1.1** Let  $X$  be a non-negative integer-valued random variable (RV) with  $\alpha \in [0, 1]$ . Then the generalized thinning operator,  $\odot$ , is defined by

$$\alpha \odot X = \sum_{k=1}^X Y_k, \quad (2.1)$$

where  $\{Y_k : k \in \mathbb{N}\}$  are a set of independent and identically distributed (IID) RVs from some count distribution with mean  $\alpha$  and variance  $\beta$  [Latour, 1998]. In (2.1) when  $X = 0$ , the sum is assumed to be 0.

Before we discuss other thinning operators, we define the probability generating function of a discrete RV, which will be used at various points throughout this dissertation.

**Definition 2.1.2** Let  $X$  be a discrete RV taking values in  $\mathbb{N}_0$ , the set of non-negative integers. Then the **probability generating function (pgf)** of  $X$  is defined as,

$$\Phi_X(z) = E(z^X) = \sum_{x=0}^{\infty} P(X = x)z^x.$$

Using Definition 2.1.2, we can further define the general form of the pgf for the generalized thinning operator as follows,

**Lemma 2.1.3** The pgf of  $\alpha \odot X$ , as defined in Definition 2.1.1, is

$$\Phi_{\alpha \odot X}(z) = \Phi_X(\Phi_Y(z)),$$

where  $Y$  represents the distribution of  $Y_k$  in (2.1) which are IID RVs.

Zhu and Joe [2010b] prove this lemma using the law of iterated expectations with

$$\begin{aligned} \Phi_{\alpha \odot X}(z) &= E \left[ E \left( z^{\sum_{k=1}^X Y_k} \middle| X \right) \right] \\ &= E \left[ (\Phi_Y(z))^X \right] \\ &= \Phi_X(\Phi_Y(z)). \end{aligned}$$

The second step follows from the property of pgfs; i.e.,  $\Phi_{X_1+\dots+X_N}(z) = \prod_{j=1}^N \Phi_{X_j}(z)$ .  $\square$

Some thinning operators used in the literature that follow Definition 2.1.1 are discussed below.

**Binomial thinning:** Suppose  $Z = \{Y_k : k \in \mathbb{N}\}$  is a set of IID Bernoulli RVs with success probability  $\alpha$ ,  $\text{Bern}(\alpha)$ . Then  $\alpha \odot X$  given  $X = x$  follows a binomial distribution with size  $x$  and success probability  $\alpha$ ,  $\text{Binomial}(x, \alpha)$ , with a probability mass function (pmf) evaluated at  $z$  of

$$\binom{x}{z} \alpha^z (1 - \alpha)^{(x-z)} I\{0 \leq z \leq x\}.$$

**Negative binomial thinning:** Let  $\{Y_k : k \in \mathbb{N}\}$  be a set of IID Geometric( $1/(1 + \alpha)$ ) RVs, with pmf given by  $P(Y_k = y) = \alpha^y / (1 + \alpha)^{y+1} I\{y \geq 0\}$ . Then  $\alpha \odot X$  given  $X = x$  has a negative binomial distribution with size  $x$  and success probability  $1/(1 + \alpha)$  when  $x > 0$ , and is 0 with probability 1 when  $x = 0$  [Ristić et al., 2009].

**$\rho$ -Binomial thinning:** In this thinning operation, the  $\{Y_k : k \in \mathbb{N}\}$  are a set of IID  $\rho$ -Bernoulli RVs, with pmf given as follows

$$P(Y_k = y) = \begin{cases} 1 - \alpha & \text{if } y = 0; \\ \alpha \left(\frac{\rho}{1+\rho}\right)^{y-1} \left(\frac{1}{1+\rho}\right) & \text{if } y = 1, 2, \dots \end{cases} \quad (2.2)$$

where  $\alpha \in [0, 1]$  is the success probability and  $\rho \in [0, 1)$  is the dispersion parameter [Kolev et al., 2000].

**Modified negative binomial thinning:** As compared with negative binomial thinning, this thinning operator has slightly modified bounds given by

$$\alpha \odot X = \sum_{k=1}^{X+1} Y_k,$$

where  $Y_k \sim \text{Geometric}(1/(1+\alpha))$ . This modified definition was introduced to circumvent the problem of having constant zeroes over time since  $\alpha \odot 0 = 0$ , and is used in the definition of

minifaction INAR(1) models which are not themselves a GINAR process [Aleksić and Ristić, 2021]. Regardless we can still use the modified negative binomial thinning operator to define GINAR processes.

There are other thinning operators defined in the literature that do not fit into the form given by Definition 2.1.1. See, for example, Zhang et al. [2010], Ristić et al. [2013], and Weiß [2008] for a survey.

## 2.2 Defining GINAR( $p$ ) Processes

We can now define the class of GINAR( $p$ ) processes.

**Definition 2.2.1** *The GINAR( $p$ ) process [Latour, 1998]  $\{X_t : t \in \mathbb{Z}\}$  is the non-negative integer-valued stationary and ergodic process defined by*

$$X_t = \sum_{j=1}^p \alpha_j \odot X_{t-j} + \epsilon_t, \quad t \in \mathbb{Z}. \quad (2.3)$$

*In the above definition,  $\alpha_j \in [0, 1)$  for each  $j = 1, \dots, p$  with  $\sum_{j=1}^p \alpha_j < 1$ , and the innovation process  $\{\epsilon_t : t \in \mathbb{Z}\}$  is a set of IID non-negative integer-valued RVs with  $\mu_\epsilon > 0$  and variance  $\sigma_\epsilon^2 > 0$ . Also, the  $\{Y_k\}$  associated with each thinning operation (see (2.1)) are mutually independent and independent of  $\{\epsilon_t\}$ , and  $\epsilon_t$  is independent of  $X_{t-j}$  for all  $t \in \mathbb{Z}$  and  $j \geq 1$ .*

When we use binomial thinning in (2.3), we obtain the INAR( $p$ ) process of Jin-Guan and Yuan [1991] with the INAR(1) process of Al-Osh and Alzaid [1987] and McKenzie [1985] when  $p = 1$ . The less commonly used INAR( $p$ ) process of Al-Osh and Aly [1992] cannot be written in this form.

Note that in the Jin-Guan and Yuan [1991] definition of the GINAR( $p$ ) model with binomial thinning and Poisson innovations, all counting series are mutually independent and independent of  $\epsilon_t$ ; i.e., the variables  $\alpha_j \odot X_{t-j}, j = 1, 2, \dots, p$ , conditional on  $X_{t-j}, j =$

$1, 2, \dots, p$ , are mutually independent; and  $\epsilon_t$  is independent of  $X_{t-j}$  for all  $j \geq 1$ . This is not the case in [Alzaid and Al-Osh \[1990\]](#) definition where the assumption is that the conditional distribution of  $(\alpha_1 \circ X_t, \alpha_2 \circ X_t, \dots, \alpha_p \circ X_t)^T$  given  $X_t = x_t$  is multinomial with parameters  $(\alpha_1, \alpha_2, \dots, \alpha_p, x_n)$ . Unsurprisingly the [Al-Osh and Alzaid \[1987\]](#) definition of GINAR( $p$ ) processes leads to processes with very different properties to those in [Definition 2.2.1](#). We do not consider this process further in this dissertation.

As we will see in [Section 2.4](#), the GINAR( $p$ ) process defined by [Definition 2.2.1](#) has the same the correlation structure as the stationary AR( $p$ ) process. Recall, the definition of the AR( $p$ ) process as follows,

$$X_t = \sum_{j=1}^p \phi_j X_{t-j} + \epsilon_t, \quad t \in \mathbb{Z},$$

where  $\{\epsilon_t\}$  is a white noise process; i.e.,  $\{\epsilon_t\} \sim WN(0, \sigma^2)$ , that is uncorrelated with  $X_s$  for  $s < t$ . In the AR( $p$ ) process,  $\phi_j, j = 1, 2, \dots, p$  are the model coefficients with the constraint that all roots of the characteristic polynomial lie outside the unit circle. The characteristic polynomial for the AR( $p$ ) process is defined as

$$\Phi(B) = 1 - \sum_{j=1}^p \phi_j B^j,$$

where  $B$  is the backshift operator [[Brockwell and Davis, 2016](#)]. Note that a key difference in the GINAR( $p$ ) and AR( $p$ ) models is that the AR( $p$ ) model has one random component in the model (white noise), whereas the GINAR( $p$ ) has added randomness components due to the thinning operations.

One way of obtaining the marginal distribution of the process is through calculations involving pgfs. This method has been used for deriving marginal distributions of GINAR(1) processes; see, for example, [Al-Osh and Aly \[1992\]](#) and [Aleksić and Ristić \[2021\]](#). For higher order models obtaining marginal distributions may not be feasible. [Joe \[2019\]](#) shows that

the stationary marginal distribution of a GINAR( $p$ ) process can be obtained through

$$\Phi_X(z) = \prod_{j=1}^p \Phi_X\{\Phi_{Y_j}(z)\}\Phi_{\epsilon_t}(z).$$

### 2.3 Example GINAR( $p$ ) Processes

This is a more focused literature review of the different variations of the INAR process that fit under the GINAR( $p$ ) framework. The most widely studied is the INAR(1) process with Poisson innovations and binomial thinning introduced by [Al-Osh and Alzaid \[1987\]](#).

Research in the area has looked at extensively at varying either one of the components of the GINAR( $p$ ) process; i.e., the thinning operator and innovation sequence. A list of different innovation distribution and thinning operators is shown in [Table 2.1](#). Varying the thinning operator allows to capture different dependence structure in the process. [Zhu and Joe \[2010a\]](#) provide three new thinning operators that generate a negative binomial marginal model, but have different conditional heteroskedasticities. [Ristić et al. \[2012\]](#) introduce a negative-binomial thinning operator where the counting series follows a geometric distribution, this was also presented in [Wang et al. \[2021\]](#). [Weiß \[2015\]](#) introduce a binomial-Poisson thinning operator to capture different dependencies in data and [Ristić et al. \[2013\]](#) introduce the generalized binomial thinning operator where the counting series is a sequence of dependent Bernoulli counting series. They also derive the distributional properties of the innovation sequence that give geometric marginals. [Aleksić and Ristić \[2021\]](#) introduce a modified negative binomial thinning operator to avoid issues with constant zero behavior over time. [Borges et al. \[2016\]](#) introduce the  $\rho$ -binomial thinning operator, where the counting series follows a  $\rho$ -Bernoulli distribution described in [Section 2.1](#). [Weiß \[2008\]](#) further provides an extensive review of different thinning operators used for GINAR models.

The other component of the model is the innovation sequence, which can easily be varied. Bourguignon and Vasconcellos [2015b] introduces the power series INAR(1) (PSINAR(1)) model with power series innovations, which are a flexible innovation distribution to model under dispersion, equidispersion, and over dispersion. Recall that under dispersion is when observed variability is lower than expected variability, overdispersion is when observed variability is greater than expected, and equidispersion is when they are equal [Kokonendji, 2014]. Some distributions in the power series distributional family include Bernoulli, binomial, geometric, Poisson, negative binomial and logarithmic. Other examples of innovation distributions include, double Poisson and generalized Poisson innovations [Bourguignon et al., 2019], Bell [Huang and Zhu, 2021], Poisson-Lindley [Lívio et al., 2018], Lerch, good, weighted Poisson (WP), power-law WP, Poisson polynomial [Weiß, 2013], Katz family [Kim and Lee, 2017], zero-inflated Poisson [Jazi et al., 2012], geometric [Aleksić and Ristić, 2021]. Hence, modifying the thinning and/or innovation distribution gives us flexible and simple processes for count time series that can account for different features that we observe in the data. All these variations fit into the GINAR( $p$ ) process framework.

## 2.4 Statistical Properties

In this section we present important statistical properties of the GINAR( $p$ ) process and the generalized thinning operator, that will be useful throughout this dissertation.

**Lemma 2.4.1** *The thinning operator defined by Definition 2.1.1 has the following properties.*

- a)  $0 \odot X = 0$ ;
- b)  $1 \odot X = X$ ;
- c)  $E(\alpha \odot X) = \alpha E(X)$ ;
- d)  $E(\alpha \odot X)^2 = \alpha^2 E(X^2) + \beta E(X)$ .

Table 2.1: Examples of GINAR processes found in literature. NA values indicate that a closed form expression for the marginal distribution does not exist.

<b>Process</b>	<b>Thinning Operator</b>	<b>Innovation Dist.</b>	<b>Marginal Dist.</b>	<b>Reference</b>
<b>PSINAR(1)</b>	Binomial	Power Series	NA	<a href="#">Bourguignon and Vasconcelos [2015b]</a>
<b>DP INAR(1)</b>	Binomial	double Poisson	NA	<a href="#">Bourguignon et al. [2019]</a>
<b>GP INAR(1)</b>	Binomial	Generalized Poisson	NA	<a href="#">Bourguignon et al. [2019]</a>
<b>BL-INAR(1)</b>	Binomial	Bell	NA	<a href="#">Huang and Zhu [2021]</a>
<b>PL-INAR(1)</b>	Binomial	Poisson-Lindley	NA	<a href="#">Lívio et al. [2018]</a>
<b>INARKF(1)</b>	Binomial	Katz family	NA	<a href="#">Kim and Lee [2017]</a>
<b>ZP-INAR(1)</b>	Binomial	zero-inflated Poisson	NA	<a href="#">Kim and Lee [2017]</a>
<b>INAR(1)</b>	Binomial	Lerch, good, weighted Poisson (WP), power-law WP, Poisson polynomial	NA	<a href="#">Weiß [2013]</a>
<b>DCINAR(1)</b>	Generalized Binomial	Mixture of zero and Geometric RVs	Geometric	<a href="#">Ristić et al. [2013]</a>
<b><math>\rho</math>-GINAR(1)</b>	$\rho$ -Binomial	Derived to satisfy marginal	Geometric	<a href="#">Borges et al. [2016]</a>
<b>Geometric Minifaction INAR(1)</b>	Neg. Binomial	Geometric	Geometric	<a href="#">Aleksić and Ristić [2021]</a>

Table 2.2: Examples of GINAR processes found in literature. NA values indicate that a closed form expression for the marginal distribution does not exist (continued).

<b>Process</b>	<b>Thinning Operator</b>	<b>Innovation Dist.</b>	<b>Marginal Dist.</b>	<b>Reference</b>
<b>NBTINAR(1)</b>	Neg. Binomial	non-negative RV	NA	<a href="#">Wang et al. [2021]</a>
<b>NonLINAR(1)</b>	Geometric	non-negative RV	NA	<a href="#">Barreto-Souza et al. [2023]</a>
<b>Extended Poisson INAR(1)</b>	Bernoulli-Poisson Convolution	Poisson	NA	<a href="#">Weiß [2015]</a>
<b>GINARS(<math>p</math>)</b>	Non-negative RV	Signed generalized power series	NA	<a href="#">Zhang et al. [2010]</a>

Proof: For property (c) the result follows from iterated expectations and the mean property of the thinning operator where

$$\begin{aligned}
 E(\alpha \odot X) &= E\{E(\alpha \odot X|X)\} \\
 &= E\{\alpha X\} \\
 &= \alpha E(X).
 \end{aligned}$$

Property (d) also follows from the law of iterated expectation and the mean and variance properties of the thinning operator given by

$$\begin{aligned}
 E(\alpha \odot X)^2 &= \text{var}(\alpha \odot X) + (E(\alpha \odot X))^2 \\
 &= E(\text{var}(\alpha \odot X|X)) + \text{var}(E(\alpha \odot X|X)) + \alpha^2(E(X))^2 \\
 &= E(X\beta) + \text{var}(\alpha X) + \alpha^2(E(X))^2 \\
 &= \beta E(X) + \alpha^2 E(X^2) - \alpha^2(E(X))^2 + \alpha^2(E(X))^2 \\
 &= \alpha^2 E(X^2) + \beta E(X).
 \end{aligned}$$

It can be shown that for binomial and negative binomial thinning that  $\beta = \alpha(1 - \alpha)$  and  $\beta = \alpha(1 + \alpha)$ , respectively.  $\square$

**Lemma 2.4.2** *Let  $\{X_t : t \in \mathbb{Z}\}$  be a GINAR( $p$ ) process of Definition 2.2.1. Then for each  $t \in \mathbb{Z}$ , the conditional mean is*

$$\mu_{X_t|X_{t-1}, \dots, X_{t-p}} = E(X_t|X_{t-1}, \dots, X_{t-p}) = \sum_{j=1}^p \alpha_j X_{t-j} + \mu_\epsilon,$$

and the conditional variance is

$$\sigma_{X_t|X_{t-1}, \dots, X_{t-p}}^2 = \text{var}(X_t|X_{t-1}, \dots, X_{t-p}) = \sum_{j=1}^p \beta_j X_{t-j} + \sigma_\epsilon^2.$$

Further, the marginal mean and variance are respectively

$$\mu_X = \frac{\mu_\epsilon}{1 - \sum_{j=1}^p \alpha_j} \quad \text{and} \quad \text{var}(X_t) = \sum_{j=1}^p \beta_j \mu_X + \frac{\sigma_\epsilon^2}{1 - \sum_{j=1}^p \alpha_j \rho_X(j)},$$

where  $\{\rho_X(k) : k \in \mathbb{Z}\}$  is the autocorrelation sequence for  $\{X_t\}$ . The autocovariance sequence  $\{\gamma_X(k) : k \in \mathbb{Z}\}$  satisfies the following for all lags  $k \neq 0$ :

$$\gamma_X(k) = \alpha_1 \gamma_X(|k| - 1) + \alpha_2 \gamma_X(|k| - 2) + \dots + \alpha_p \gamma_X(|k| - p).$$

Similarly, the autocorrelation sequence satisfies for all lags  $k \neq 0$ :

$$\rho_X(k) = \alpha_1 \rho_X(|k| - 1) + \alpha_2 \rho_X(|k| - 2) + \dots + \alpha_p \rho_X(|k| - p).$$

The spectral density function,  $f_X(\cdot)$ , of the process is

$$f_X(\nu) = \frac{\sigma_\epsilon^2 + \mu_X \sum_{j=1}^p \beta_j}{2\pi |\alpha(e^{-i\nu})|^2}, \quad \nu \in [-\pi, \pi],$$

where  $\alpha(B) = 1 - \sum_{j=1}^p \alpha_j B^j$  is a transfer function and  $B$  is the backshift operator.

Proof: The conditional expectation and variance property can be derived as follows:

$$\begin{aligned}
E(X_t|X_{t-1}, \dots, X_{t-p}) &= \sum_{j=1}^p E(\alpha_j \odot X_{t-j}|X_{t-1}, \dots, X_{t-p}) + E(\epsilon_t|X_{t-1}, \dots, X_{t-p}) \\
&= \sum_{j=1}^p E(\alpha_j \odot X_{t-j}|X_{t-j}) + E(\epsilon_t) \\
&= \sum_{j=1}^p \alpha_j X_{t-j} + \mu_\epsilon,
\end{aligned}$$

and

$$\begin{aligned}
\text{var}(X_t|X_{t-1}, \dots, X_{t-p}) &= \sum_{j=1}^p \text{var}(\alpha_j \odot X_{t-j}|X_{t-1}, \dots, X_{t-p}) + \text{var}(\epsilon_t|X_{t-1}, \dots, X_{t-p}) \\
&= \sum_{j=1}^p \beta_j X_{t-j} + \sigma_\epsilon^2.
\end{aligned}$$

The marginal mean is obtained through the law of iterated expectations and stationarity of the process as follows:

$$\begin{aligned}
\mu_X &= E(E(X_t|X_{t-1}, \dots, X_{t-p})) \\
&= E\left(\sum_{j=1}^p \alpha_j X_{t-j} + \mu_\epsilon\right) \\
&= \mu_X \sum_{j=1}^p \alpha_j + \mu_\epsilon.
\end{aligned}$$

Rearranging we get

$$\mu_X = \frac{\mu_\epsilon}{1 - \sum_{j=1}^p \alpha_j}.$$

The marginal variance of the process is obtained as follows:

$$\begin{aligned}
\text{var}(X_t) &= \text{var}(E(X_t|X_{t-1}, \dots, X_{t-p})) + E(\text{var}(X_t|X_{t-1}, \dots, X_{t-p})) \\
&= E\left(\sum_{j=1}^p \beta_j X_{t-j} + \sigma_\epsilon^2\right) + \text{var}\left(\sum_{j=1}^p \alpha_j X_{t-j} + \mu_\epsilon\right) \\
&= \mu_X \sum_{j=1}^p \beta_j + \sigma_\epsilon^2 + \sum_{j=1}^p \sum_{k=1}^p \alpha_k \alpha_j \rho_X(j-k) \text{var}(X_t) \\
&= \mu_X \sum_{j=1}^p \beta_j + \sigma_\epsilon^2 + \sum_{j=1}^p \alpha_j \rho_X(j) \text{var}(X_t).
\end{aligned}$$

Note that the last step follows from:

$$\begin{aligned}
\text{var}(X_t) \sum_{j=1}^p \alpha_j \alpha_k \rho_X(j) &= \text{var}(\boldsymbol{\alpha}^T \mathbf{X}) \\
&= \text{var}(X_t) \boldsymbol{\alpha}^T \mathbf{R} \boldsymbol{\alpha} \\
&= \text{var}(X_t) \boldsymbol{\alpha}^T \mathbf{r},
\end{aligned}$$

where  $\mathbf{R} = [\rho_X(|i-j|)]_{p \times p}$ ,  $\mathbf{r} = (\rho_X(1) \dots \rho_X(p))^T$ , and  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_p)^T$ . Thus

$$\text{var}(X_t) = \frac{\mu_X \sum_{j=1}^p \beta_j + \sigma_\epsilon^2}{1 - \sum_{j=1}^p \alpha_j \rho_X(j)}.$$

Next, we present a proof for the autocorrelation and autocovariance properties of the process. To derive the autocovariance structure we present the GINAR( $p$ ) process in a matrix form similar to [Jin-Guan and Yuan \[1991\]](#). Let  $\mathbf{X}_t = (X_t, X_{t-1}, \dots, X_{t-p+1})^T$  and let  $\mathbf{A}$  and  $\boldsymbol{\epsilon}_t$  be defined as

$$\mathbf{A} = \begin{bmatrix} \alpha_1 & \alpha_2 & \dots & \alpha_{p-1} & \alpha_p \\ 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix} \text{ with } \boldsymbol{\epsilon}_t = \begin{bmatrix} \epsilon_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad t \in \mathbb{Z}.$$

Then the GINAR( $p$ ) process is given by

$$\mathbf{X}_t = \mathbf{A} \odot \mathbf{X}_{t-1} + \boldsymbol{\epsilon}_t, \quad t \in \mathbb{Z}.$$

We define the multivariate autocovariance sequence  $\{\mathbf{\Gamma}(k) : k \in \mathbb{Z}\}$  by

$$\mathbf{\Gamma}(k) = E[(\mathbf{X}_t - E(\mathbf{X}_t))(\mathbf{X}_{t-k} - E(\mathbf{X}_{t-k}))^T], \quad k \in \mathbb{Z}.$$

Then for  $k \geq 0$

$$\begin{aligned} \mathbf{\Gamma}(k) &= E(\mathbf{X}_t \mathbf{X}_{t-k}^T) - E(\mathbf{X}_t)E(\mathbf{X}_{t-k})^T \\ &= E((\mathbf{A} \odot \mathbf{X}_{t-1} + \boldsymbol{\epsilon}_t) \mathbf{X}_{t-k}^T) - E(\mathbf{X}_t)E(\mathbf{X}_{t-k})^T \\ &= \mathbf{A}E(\mathbf{X}_{t-1} \mathbf{X}_{t-k}^T) + E(\boldsymbol{\epsilon}_t)E(\mathbf{X}_{t-k})^T - E(\mathbf{X}_t)E(\mathbf{X}_{t-k})^T \\ &= \mathbf{A}E(\mathbf{X}_{t-1} \mathbf{X}_{t-k}^T) - \{E(\mathbf{X}_t) - E(\boldsymbol{\epsilon}_t)\}E(\mathbf{X}_{t-k})^T \\ &= \mathbf{A}\{E(\mathbf{X}_{t-1} \mathbf{X}_{t-k}^T) - E(\mathbf{X}_{t-1})E(\mathbf{X}_{t-k})^T\} \\ &= \mathbf{A}\mathbf{\Gamma}(k-1), \end{aligned}$$

where we are using the property that  $E(\mathbf{X}_t) = \mathbf{A}E(\mathbf{X}_{t-1}) + E(\boldsymbol{\epsilon}_t)$  and the shift invariant property of the autocovariance function. We can then get the univariate autocovariance at lag  $k$  ( $k \geq 0$ ) as

$$\gamma_X(k) = \alpha_1 \gamma_X(k-1) + \alpha_2 \gamma_X(k-2) + \dots + \alpha_p \gamma_X(k-p),$$

and consequently the autocorrelation at lag  $k$  is

$$\rho_X(k) = \alpha_1 \rho_X(k-1) + \alpha_2 \rho_X(k-2) + \dots + \alpha_p \rho_X(k-p).$$

For a proof of the spectral density function property refer to [Silva and Oliveira \[2005, pages 30-31\]](#). □

## 2.5 Transition Probabilities

In this section we discuss the calculations for obtaining transition probabilities of the GINAR( $p$ ) process. Following Definition 2.2.1, the GINAR( $p$ ) process is a Markov chain. The transition probabilities are given in the next theorem.

**Theorem 2.5.1** [*Hadri, 2009, Theorem 1*] For  $\{X_t : t \in \mathbb{Z}\}$ , a GINAR( $p$ ) process as defined in Definition 2.2.1, the transition probabilities are given by

$$\begin{aligned}
& P(X_t = x | X_{t-1} = x_{t-1}, \dots, X_{t-p} = x_{t-p}) \\
&= \sum_{i_1=0}^x P(\alpha_1 \odot x_{t-1} = i_1 | X_{t-1} = x_{t-1}) \times \\
&\quad \sum_{i_2=0}^{x-i_1} P(\alpha_2 \odot x_{t-2} = i_2 | X_{t-2} = x_{t-2}) \times \dots \times \\
&\quad \sum_{i_p=0}^{x-(i_1+i_2+\dots+i_{p-1})} P(\alpha_p \odot x_{t-p} = i_p | X_{t-p} = x_{t-p}) \times \\
&\quad P(\epsilon_t = x - (i_1 + i_2 + \dots + i_p)).
\end{aligned}$$

The proof of Theorem 2.5.1 follows directly by considering the convolutions that define the process. Below we show an example calculation for obtaining the transition probabilities.

**GINAR(1) Example:** Here we look at the example of a GINAR(1) process with binomial thinning and Poisson innovations expressed as

$$X_t = \alpha \odot X_{t-1} + \epsilon_t$$

where,  $\epsilon_t \sim \text{Poisson}(\lambda)$  and  $\alpha \odot X_{t-1} = \sum_{k=1}^{X_{t-1}} Y_k$ , where  $Y_k$  are IID Bernoulli( $\alpha$ ) RVs. Hence,  $\alpha \odot X_{t-1}$  given  $X_{t-1} = x_{t-1}$  follows a Binomial distribution with size  $x_{t-1}$  and success probability  $\alpha$ , Binomial( $x_{t-1}, \alpha$ ). As mentioned earlier, this is the most widely studied class of GINAR processes, called the Poisson INAR(1) process, and was introduced first by [Al-Osh and Alzaid \[1987\]](#) and [McKenzie \[1985\]](#) independently. This process can be interpreted as a birth and death process, where the process at time  $t$  depends on the survivors of the process at time  $t-1$  (with survival probability  $\alpha$ ), and the elements that entered the system during  $(t-1, t]$  are denoted by the innovation sequence.

To calculate the transition probabilities for the Poisson INAR(1) process, we first note the following lemma.

**Lemma 2.5.2** *Let  $X$  and  $Y$  be two independent RVs from some discrete distribution. Then the distribution of their sum,  $Z = X+Y$  is given by*

$$P(Z = z) = \sum_{s=-\infty}^{\infty} P(X = s)P(Y = z - s)$$

Then, using Lemma 2.5.2 we can calculate the transition probabilities of the Poisson INAR(1) process as follows:

$$\begin{aligned} P(X_t = x | X_{t-1} = x_{t-1}) &= \sum_{s=-\infty}^{s=\infty} P\left(\sum_{j=1}^{x_{t-1}} Y_j = s | X_{t-1} = x_{t-1}\right) P(\epsilon_t = x - s | X_{t-1} = x_{t-1}) \\ &= \sum_{s=0}^{\min(x, x_{t-1})} \binom{x_{t-1}}{s} \alpha^s (1 - \alpha)^{x_{t-1} - s} P(\epsilon_t = x - s) \\ &= \sum_{s=0}^{\min(x, x_{t-1})} \binom{x_{t-1}}{s} \alpha^s (1 - \alpha)^{x_{t-1} - s} \frac{\lambda^{x-s} e^{-\lambda}}{(x-s)!} \end{aligned}$$

Note that the bounds follow due to the constraint that  $s \geq 0$  and  $s \leq x_{t-1}$ . Also,  $x - s \geq 0$  since  $\epsilon_t$  follows a Poisson distribution. Combining all the inequalities we have the bound,  $0 \leq s \leq \min(x, x_{t-1})$ .

Furthermore, in this Poisson INAR(1) process, when the starting value  $X_0$  is Poisson with mean  $\lambda/(1 - \alpha)$ , the marginal distribution of the process is also Poisson( $\lambda/(1 - \alpha)$ ). For a formal proof refer to [Freeland \[1998\]](#).

For larger values of  $p$  the calculation of these convolutions become computationally burdensome (on average  $O(\mu_X^p)$ , where  $\mu_X$  is the mean of the process). Instead, noticing that convolutions can be rewritten as products in the Fourier domain, [Joe \[2019\]](#) proposes calculating the transition probabilities by integrating the characteristic function (chf) (e.g., [Davies \[1973\]](#)). This leads to a more computationally efficient algorithm for calculating transition probabilities, as demonstrated in the next proposition.

**Proposition 2.5.3** For  $\{X_t : t \in \mathbb{Z}\}$  a  $GINAR(p)$  process as defined in Definition 2.2.1, suppose that  $\phi_{X_t|X_{t-1}, \dots, X_{t-p}}(u)$  is the chf for the transition probability defined in Theorem 2.5.1. Then the cumulative distribution function is

$$a(x) = P(X_t < x | X_{t-1}, \dots, X_{t-p}) = \frac{1}{2} - \frac{1}{2\pi} \int_{-\pi}^{\pi} \operatorname{Re} \left( \frac{\phi_{X_t|X_{t-1}, \dots, X_{t-p}}(u) e^{-iux}}{1 - e^{-iu}} \right) du, \quad (2.4)$$

where

$$\phi_{X_t|X_{t-1}, \dots, X_{t-p}}(u) = \phi_{\epsilon_t}(u) \prod_{j=1}^p [\phi_{Y^{(j)}}(u)]^{X_{t-j}}.$$

Then the transition probabilities can be calculated

$$b(x) = P(X_t = x | X_{t-1}, \dots, X_{t-p}) = \begin{cases} a(1), & x = 0; \\ a(x+1) - a(x), & x = 1, 2, \dots, \end{cases}$$

where

$$b(x) = \frac{1}{\pi} \int_0^{\pi} \operatorname{Re}(\phi_{X_t|X_{t-1}, \dots, X_{t-p}}(u) e^{-iux}) du, \quad x = 1, 2, \dots$$

Through Proposition 2.5.3 one can obtain transition probabilities for a non-negative integer valued random variable based on the chf. For this method, we only need the process to have a closed form pgf. Hence this proposition is particularly useful when the innovation distribution or thinning operator distribution has a simple form of the pgf but not the pmf.

## Chapter 3: Estimation of GINAR( $p$ ) Processes

In Section 3.1 we discuss various estimation methods for parameter estimation of GINAR( $p$ ) processes like conditional maximum likelihood (CML), pseudo maximum likelihood, Whittle likelihood, Yule-Walker, conditional least squares, and saddlepoint approximation. We also present asymptotic theory for these estimation methods when available, in particular we prove asymptotic normality for the CML estimates for the GINAR( $p$ ) process. In Section 3.2 we discuss statistical inference procedures, in particular we investigate the construction of confidence regions for parameters and highlight methods for model selection. Lastly in Section 3.3 we discuss forecasting methods for GINAR( $p$ ) processes.

### 3.1 Estimation Methods

Let  $\mathbf{X} = (X_1, \dots, X_n)^T$  be observations from a GINAR( $p$ ) process where the thinning operator and innovation process is known. We are interested in estimation of the parameter vector  $\boldsymbol{\theta} = (\alpha_1, \alpha_2, \dots, \alpha_p, \mu_\epsilon, \sigma_\epsilon^2)^T$ , where  $\sum_{j=1}^p \alpha_j < 1$  and the constraints on  $\mu_\epsilon$  and  $\sigma_\epsilon^2$  depend on the form of the innovation distribution. Let  $\Theta$  be the resulting parameter space for  $\boldsymbol{\theta}$ , which we assume to be compact. We further assume that our GINAR( $p$ ) process is identifiable; we can tell apart different values of the parameter vector on the basis of the transition probabilities.

In this section we discuss the following methods for parametric estimation: conditional maximum likelihood (CML), Yule-Walker (Y-W), pseudo maximum likelihood, conditional least squares (CLS), Whittle likelihood, and saddlepoint methods. We introduce each estimator and when possible provide asymptotic theory for the distribution of each estimator.

### 3.1.1 Conditional maximum likelihood

Conditioning on the first  $p$  observations, conditional maximum likelihood (CML) calculates the conditional log likelihood using

$$\ell(\boldsymbol{\theta}) = \sum_{t=p+1}^n \log P(X_t = x_t | X_{t-1}, \dots, X_{t-p}) = \sum_{t=p+1}^n \log b(x_t), \quad (3.1)$$

where  $b(\cdot)$  is defined in Proposition 2.5.3. The CML-based parameter estimates of  $\boldsymbol{\theta}$  can then be computed as

$$\hat{\boldsymbol{\theta}}_{CML} = \arg \max_{\boldsymbol{\theta} \in \Theta} \ell(\boldsymbol{\theta}). \quad (3.2)$$

Numerical optimization techniques are typically used to maximize (3.2). As discussed in Section 2.4, the form of the transition probabilities will depend on the thinning operation and marginal distribution of innovation. We know the computational complexity of CML increases with  $p$ , and a brute force method of coding the transition probabilities is not efficient. Thus we use Proposition 2.5.3 to efficiently evaluate transition probabilities using numerical quadrature techniques.

We provide a proof for the consistency of  $\hat{\boldsymbol{\theta}}_{CML}$  in the following theorem.

**Theorem 3.1.1** *Let  $\{X_t : t \in \mathbb{Z}\}$  be a GINAR( $p$ ) process. Then,*

$$\hat{\boldsymbol{\theta}}_{CML} \rightarrow \boldsymbol{\theta}_0$$

*in probability as  $n \rightarrow \infty$ .*

Proof: We will show that the conditions of Theorem 4.1.2 in Amemiya [1985] are satisfied, giving us consistency of the estimator.

(i) By assumption and properties of the GINAR( $p$ ) process, the parameter space  $\Theta = \{\boldsymbol{\theta} : \boldsymbol{\theta} = (\alpha_1, \alpha_2, \dots, \alpha_p, \lambda_\epsilon, \sigma_\epsilon^2)^T\}$  is compact. All elements of  $\boldsymbol{\theta}$  are bounded by positive constants and  $\boldsymbol{\theta}_0$  is an interior point of  $\Theta$ . Therefore, the parameter space  $\Theta$  is an open subset of  $\mathbb{R}^k$ .

(ii) Let  $P_{\boldsymbol{\theta}}$  be the probability measure under the parameter  $\boldsymbol{\theta}$  and assume the following notation:

$$U_t(\boldsymbol{\theta}) = \log P_{\boldsymbol{\theta}}(X_t | X_{t-1}, X_{t-2}, \dots, X_{t-p});$$

$$Q_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{t=p+1}^n \log P_{\boldsymbol{\theta}}(X_t | X_{t-1}, \dots, X_{t-p}) = \frac{1}{n} \sum_{t=p+1}^n U_t(\boldsymbol{\theta}).$$

Note that  $U_t(\boldsymbol{\theta})$  is continuous for all  $\boldsymbol{\theta} \in \Theta$  and hence is a measurable function of the data for all  $\boldsymbol{\theta} \in \Theta$ . Also note that

$$\frac{\partial U_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{t=p+1}^n \frac{\partial P_{\boldsymbol{\theta}}(X_t | X_{t-1}, X_{t-2}, \dots, X_{t-p}) / \partial \boldsymbol{\theta}}{P_{\boldsymbol{\theta}}(X_t | X_{t-1}, X_{t-2}, \dots, X_{t-p})}$$

is clearly continuous in an open neighborhood of  $\boldsymbol{\theta}_0$ .

(iii) First, note that  $U_t(\boldsymbol{\theta})$  is continuous in an open and convex neighborhood of  $\boldsymbol{\theta}_0$ , denoted by  $N(\boldsymbol{\theta}_0) = (\boldsymbol{\theta}_0 - \epsilon, \boldsymbol{\theta}_0 + \epsilon)$ , for all  $\epsilon > 0$ . This implies the existence of at least one point such that

$$U_t(\tilde{\boldsymbol{\theta}}) = \arg \max_{\boldsymbol{\theta} \in \Theta} U_t(\boldsymbol{\theta}).$$

Then using Jensen's inequality we can get the following result:

$$\begin{aligned} E \left( \sup_{\boldsymbol{\theta} \in N(\boldsymbol{\theta}_0)} U_t(\boldsymbol{\theta}) \right) &= E \left( \log U_t(\tilde{\boldsymbol{\theta}}) \right) \\ &= E \left( \log P_{\tilde{\boldsymbol{\theta}}}(X_t | X_{t-1}, \dots, X_{t-p}) \right) \\ &\leq \log E \left( P_{\tilde{\boldsymbol{\theta}}}(X_t | X_{t-1}, \dots, X_{t-p}) \right) < \infty. \end{aligned}$$

Thus, the conditions for the Uniform Law of Large Numbers (ULLN) [Andrews, 1987] is satisfied and we get

$$\frac{1}{n} \sum_{t=p+1}^n U_t(\boldsymbol{\theta}) \xrightarrow{i.p.} E(U_t(\boldsymbol{\theta})) \equiv Q_0(t),$$

as  $n \rightarrow \infty$ .

Next, we want to show that  $Q_0(t)$  attains a local maximum at  $\boldsymbol{\theta}_0$ . We can show that  $\boldsymbol{\theta}_0$  uniquely maximizes  $E(U_t(\boldsymbol{\theta}))$  by

$$\begin{aligned} E(U_t(\boldsymbol{\theta})) &= E(\log P_{\boldsymbol{\theta}}(X_t|X_{t-1}, \dots, X_{t-p})) \\ &= E(\log P_{\boldsymbol{\theta}_0}(X_t|X_{t-1}, \dots, X_{t-p})) - E\left(\log \frac{P_{\boldsymbol{\theta}_0}(X_t|X_{t-1}, \dots, X_{t-p})}{P_{\boldsymbol{\theta}}(X_t|X_{t-1}, \dots, X_{t-p})}\right) \\ &\leq E(\log P_{\boldsymbol{\theta}_0}(X_t|X_{t-1}, \dots, X_{t-p})) = E(U_t(\boldsymbol{\theta}_0)). \end{aligned}$$

Hence, because we have assumed identifiability and given the argument above, we have that  $\boldsymbol{\theta}_0$  uniquely maximizes  $E(U_t(\boldsymbol{\theta}))$ .

The assumptions of Theorem 4.1.2 of Amemiya [1985] are satisfied, implying that the roots of  $\frac{\partial Q_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$  are consistent for  $\boldsymbol{\theta}_0$ .  $\square$

Next, we prove asymptotic normality of  $\widehat{\boldsymbol{\theta}}_{CML}$ .

**Theorem 3.1.2** *Let  $\{X_t : t \in \mathbb{Z}\}$  be a GINAR( $p$ ) process. Then the conditional maximum likelihood estimator,  $\widehat{\boldsymbol{\theta}}_{CML}$ , has the following asymptotic distribution*

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_{CML} - \boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}_{CML}), \quad (3.3)$$

as  $n \rightarrow \infty$ , where

$$\begin{aligned} \boldsymbol{\Sigma}_{CML} &= \mathbf{J}(\boldsymbol{\theta})^{-1} \mathbf{K}(\boldsymbol{\theta}) \mathbf{J}(\boldsymbol{\theta})^{-1}, \\ \mathbf{K}(\boldsymbol{\theta}) &= \lim_{n \rightarrow \infty} \frac{1}{n} E \left( \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \left( \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T \right), \end{aligned}$$

and

$$\mathbf{J}(\boldsymbol{\theta}) = \lim_{n \rightarrow \infty} \frac{1}{n} E \left( \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right).$$

Proof: We show that the conditions of Theorem 4.1.3 of Amemiya [1985] are satisfied, giving asymptotic normality of the CML estimator,  $\boldsymbol{\theta}_{CML}$ . We have already shown that assumptions of Theorem 4.1.2 of Amemiya [1985] are fulfilled which is also a requirement for Theorem 4.1.3.

(i) Note that  $Q_n(\boldsymbol{\theta})$  is a measurable function of  $X_t$  for all  $\boldsymbol{\theta} \in \Theta$  and is open and continuous in an open and convex neighborhood,  $N(\boldsymbol{\theta}_0)$ , of  $\boldsymbol{\theta}_0$ .

(ii) Let

$$\ell(\boldsymbol{\theta}) = \sum_{t=p+1}^n U_t(\boldsymbol{\theta}).$$

We want to show that

$$n^{-1} \frac{\partial^2 \ell(\boldsymbol{\theta}_n^*)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}$$

converges to a finite non-singular matrix

$$\mathbf{J}(\boldsymbol{\theta}_0) = \lim_{n \rightarrow \infty} n^{-1} E \left( \frac{\partial^2 \ell(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right)$$

in probability for any sequence  $\boldsymbol{\theta}_n^*$  such that  $\boldsymbol{\theta}_n^* \rightarrow \boldsymbol{\theta}_0$  as  $n \rightarrow \infty$ . Note that  $\boldsymbol{\theta}_n^*$  lies in between  $\boldsymbol{\theta}_0$  and  $\boldsymbol{\theta}_{CML}$ .

In Proposition 2.5.3 we represent the transition probability using characteristics functions, which are three times differentiable. Hence, we have that all partial derivatives of

$$\frac{\partial U_t(\boldsymbol{\theta})}{\partial \theta_i}$$

exist and are three times continuously differentiable in a neighborhood of  $\boldsymbol{\theta}_0$ ,  $N(\boldsymbol{\theta}_0)$ . Then, similar to step (iii) in the proof of Theorem 3.1.1 we have

$$E \left( \sup_{\boldsymbol{\theta} \in N(\boldsymbol{\theta}_0)} \frac{\partial^2 U_t(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right) = E \left( \frac{\partial^2 U_t(\tilde{\boldsymbol{\theta}})}{\partial \theta_i \partial \theta_j} \right) < \infty,$$

where  $\tilde{\boldsymbol{\theta}} \in N(\boldsymbol{\theta}_0)$ . This is because

$$\frac{\partial U_t(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j}$$

exists and is continuous in  $N(\boldsymbol{\theta}_0)$  for all  $i = 1, 2, \dots$  and  $j = 1, 2, \dots$

The conditions for ULLN is satisfied and we get

$$\frac{1}{n} \sum_{t=p+1}^n \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \xrightarrow{i.p.} E \left( \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right),$$

as  $n \rightarrow \infty$  uniformly in  $\boldsymbol{\theta} \in N(\boldsymbol{\theta}_0)$ . And by ULLN we have

$$\frac{1}{n} \sum_{t=p+1}^n \frac{\partial \ell(\boldsymbol{\theta}_n^*)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \xrightarrow{i.p.} E \left( \frac{\partial \ell(\boldsymbol{\theta}_n^*)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right),$$

when  $\boldsymbol{\theta}_n^* \rightarrow \boldsymbol{\theta}_0$  as  $n \rightarrow \infty$ .

(iii) We have the covariance matrix given by

$$\text{cov} \left( \frac{\partial U_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \right) = E \left[ \left( \frac{\partial U_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \right) \left( \frac{\partial U_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \right)^T \right],$$

since

$$E \left( \frac{\partial U_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \right) = 0.$$

Also, because of the ergodicity we have

$$n^{-1} \frac{\partial \ell(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \xrightarrow{i.p.} E \left( \frac{\partial U_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \right)$$

as  $n \rightarrow \infty$ . Then, using the martingale CLT and Cramer-Wold device we get

$$n^{-1} \frac{\partial \ell(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \xrightarrow{d} N(\mathbf{0}, \mathbf{K}(\boldsymbol{\theta}_0)),$$

as  $n \rightarrow \infty$ , where

$$\mathbf{K}(\boldsymbol{\theta}_0) = \lim_{n \rightarrow \infty} E \left( \frac{\partial \ell(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \left( \frac{\partial \ell(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \right)^T \right).$$

(iv) We note from [Latour \[1998\]](#) that the  $k^{\text{th}}$  order moment of  $X_t$  is finite for all  $t$ . Thus there exists  $W(X_t)$  such that  $E(W(X_t)) < \infty$  and

$$\left| \frac{\partial^3 \ln l(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j \partial \theta_k} \right| < W(X_t).$$

Then, taking a Taylor expansion we get

$$\frac{\partial \ell(\boldsymbol{\theta}_{CML})}{\partial \boldsymbol{\theta}} = \frac{\partial \ell(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} + \frac{\partial \ell(\boldsymbol{\theta}_n^*)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} (\widehat{\boldsymbol{\theta}}_{CML} - \boldsymbol{\theta}_0).$$

Then

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_{CML} - \boldsymbol{\theta}_0) = - \left[ n^{-1} \frac{\partial^2 \ell(\boldsymbol{\theta}_n^*)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right] \left[ n^{-1/2} \frac{\partial \ell(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \right].$$

Hence, the conditions of Theorem 4.1.3 of [Amemiya \[1985\]](#) are fulfilled. This proof is adapted from [Liu et al. \[2021\]](#). □

[Hadri \[2009, Theorem 2 and Theorem 3\]](#) provide expressions for first and second derivatives of the log likelihood function for a GINAR( $p$ ) process and show calculations for the expectations of derivatives for the GINAR(2) process with binomial thinning and Poisson innovations. Alternatively, we can calculate these quantities using the chf, as follows.

**Proposition 3.1.3** *In Proposition 2.5.3 we defined the transition probability  $b(x) = P(X_t | X_{t-1}, \dots, X_{t-p})$  for  $x = 0, 1, \dots$ . Fixing  $x$ , let  $b^{(j)}(x)$  denote the partial derivative of  $b(x)$  with respect to  $\theta_j$  and  $b^{(j,k)}(x)$  denote the second partial derivative of  $b(x)$  with respect to  $\theta_j$  and  $\theta_k$ . Then*

$$\frac{\partial}{\partial \theta_j} \ell(\boldsymbol{\theta}) = \sum_{t=p+1}^n \frac{b^{(j)}(x_t)}{b(x_t)},$$

with

$$\frac{\partial^2}{\partial \theta_j \partial \theta_k} \ell(\boldsymbol{\theta}) = \sum_{t=p+1}^n \frac{b(x_t) b^{(j,k)}(x_t) - b^{(j)}(x_t) b^{(k)}(x_t)}{b^2(x_t)}.$$

Proof: Note that for  $x = 1, 2, \dots$

$$\begin{aligned} b(x) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \operatorname{Re} \left( \frac{\phi_{X_t|X_{t-1}, \dots, X_{t-p}}(u) [e^{-iux_t} - e^{-iu(x_t+1)}]}{1 - e^{-iu}} \right) du \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \operatorname{Re} \left( \frac{\phi_{X_t|X_{t-1}, \dots, X_{t-p}}(u) e^{-iux_t} (1 - e^{-iu})}{1 - e^{-iu}} \right) du \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \operatorname{Re} (\phi_{X_t|X_{t-1}, \dots, X_{t-p}}(u) e^{-iux_t}) du \\ &= \frac{1}{\pi} \int_0^{\pi} \operatorname{Re} (\phi_{X_t|X_{t-1}, \dots, X_{t-p}}(u) e^{-iux_t}) du. \end{aligned}$$

Then

$$\ell(\boldsymbol{\theta}) = \sum_{t=p+1}^n \log b(x_t).$$

The result of the proposition follows directly by taking partial derivatives of  $\ell(\boldsymbol{\theta})$  with respect to  $\theta_j$ . □

### 3.1.2 Yule-Walker

Yule-Walker (Y-W) estimation [e.g., [Brockwell and Davis, 2016](#), Section 5.1.1] is a method of moments approach. From Proposition 2.4, we see that the GINAR( $p$ ) and AR( $p$ ) processes have the same autocorrelation structure. Consequently the Y-W equations hold for the GINAR( $p$ ) process:

$$\mathbf{\Gamma} \boldsymbol{\alpha} = \boldsymbol{\gamma}, \tag{3.4}$$

where  $\mathbf{\Gamma} = [\gamma_X(|i - j|)]_{p \times p}$ ,  $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_p]^T$ , and  $\boldsymbol{\gamma} = [\gamma_X(1), \dots, \gamma_X(p)]^T$ . Replacing the quantities in (3.4) with the corresponding sample estimates, i.e.  $\widehat{\mathbf{\Gamma}} \widehat{\boldsymbol{\alpha}}_{YW} = \widehat{\boldsymbol{\rho}}$ , provides the

Y-W estimate for  $\alpha$ . Here the sample autocovariance,  $\hat{\gamma}_X(k)$ , is

$$\hat{\gamma}_X(k) = \frac{1}{n} \sum_{t=1}^{n-|k|} (X_t - \bar{X})(X_{t+|k|} - \bar{X}), \quad k \in \mathbb{Z}.$$

The Y-W estimates of  $\mu_\epsilon$  and  $\sigma_\epsilon^2$  are [Silva and Silva, 2006]

$$\hat{\mu}_\epsilon = \left(1 - \sum_{j=1}^p \hat{\alpha}_j\right) \bar{X}$$

and

$$\sigma_\epsilon^2 = \hat{V}_p - \bar{X} \sum_{j=1}^p \hat{\beta}_j,$$

where  $\hat{V}_p = \hat{\gamma}_X(0) - \sum_{j=1}^p \hat{\alpha}_j \hat{\gamma}_X(j)$ . The asymptotic distribution of  $\hat{\alpha}_{YW}$  is as follows.

**Theorem 3.1.4** [Silva and Silva, 2006, Theorem 2] *Let  $\{X_t : t \in \mathbb{Z}\}$  be a GINAR( $p$ ) process. Assume that for  $\{Y_k : k \in \mathbb{Z}\}$  defined in Definition 2.1.1, that  $E(Y_k^3)$  and  $E(Y_k^4)$  are finite for all  $k$ . Let  $\hat{\alpha}_{YW}$  be the Y-W estimator of  $\alpha$ . Then,*

$$\sqrt{n}(\hat{\alpha}_{YW} - \alpha) \xrightarrow{d} N(\mathbf{0}, \Sigma_{YW})$$

as  $n \rightarrow \infty$ , where  $\Sigma_{YW} = \mathbf{D}^T \mathbf{Q} \mathbf{D}$ . In the definition of  $\Sigma_{YW}$ ,

$$\mathbf{D}^T = -[\gamma_X(1)\mathbf{I}_p \dots \gamma_X(p)\mathbf{I}_p](\mathbf{\Gamma}^T \otimes \mathbf{\Gamma}^{-1})[\mathbf{I}_{p^2} \ \mathbf{0}_{p^2 \times p}] + [\mathbf{0}_{p^2 \times p} \ \mathbf{\Gamma}^{-1}]$$

and  $\mathbf{Q}$  is the  $p(p+1) \times p(p+1)$  covariance matrix defined by

$$\mathbf{Q} = \text{cov}(\hat{V}(j), \hat{V}(k)), \text{ where}$$

$$\hat{V}(j) = \begin{cases} \hat{\gamma}_X(|(j-1) \bmod p - [(j-1)/p]|) & \text{if } j \leq p^2; \\ \hat{\gamma}_X(|j \bmod p|) & \text{if } j > p^2. \end{cases}$$

for  $p > 1$ , and  $\hat{V}(j) = [\hat{\gamma}_X(0) \ \hat{\gamma}_X(1)]^T$  for  $p = 1$ . Note that  $[r]$  represents the integer part of  $r \in \mathbb{R}$  and  $\otimes$  is the Kronecker product [Graham, 1981].

The asymptotic distribution for  $\hat{\sigma}_\epsilon^2$  and  $\hat{\mu}_\epsilon$  follows from Silva and Silva [2006, Theorem 1, Theorem 2] and the Delta method.

### 3.1.3 Conditional least squares

The conditional least squares (CLS) method for estimation of parameters of GINAR( $p$ ) processes was proposed by Klimko and Nelson [1978]. We first define the modified parameter space which leaves out  $\sigma_\epsilon^2$  from  $\boldsymbol{\theta}$ . Let  $\boldsymbol{\eta} = (\alpha_1, \alpha_2, \dots, \mu_\epsilon)^T$ . Then define

$$U_n(\boldsymbol{\eta}) = \sum_{t=p+1}^n \{X_t - \mu_{X_t|X_{t-1}, \dots, X_{t-p}}\}^2. \quad (3.5)$$

The CLS estimator,  $\hat{\boldsymbol{\eta}}_{CLS}$ , satisfies

$$\hat{\boldsymbol{\eta}}_{CLS} = \arg \min_{\boldsymbol{\eta} \in \Theta} U_n(\boldsymbol{\eta}). \quad (3.6)$$

The following is due to Latour [1998, Proposition 6.1].

**Theorem 3.1.5** *Let  $\{X_t : t \in \mathbb{Z}\}$  be a GINAR( $p$ ) process such that  $E(\epsilon_t^3) < \infty$  for all  $t \in \mathbb{Z}$ . Let  $\{Y_k^{(j)}\}$  be as Definition 2.1.1 where  $j$  indicates that it is associated with the  $j^{\text{th}}$  thinning operator  $\alpha_j \odot$ , and  $E\{(Y_k^{(j)})^3\} < \infty$ , for all  $j = 1, \dots, p$  and  $k \geq 1$ . Then the CLS estimator,  $\hat{\boldsymbol{\eta}}_{CLS}$ , has the following asymptotic distribution*

$$\sqrt{n}(\hat{\boldsymbol{\eta}}_{CLS} - \boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}_{CLS})$$

as  $n \rightarrow \infty$ , where  $\boldsymbol{\Sigma}_{CLS} = \mathbf{V}^{-1} \mathbf{W} \mathbf{V}$ . In the definition of  $\boldsymbol{\Sigma}_{CLS}$ ,

$$\mathbf{V} = \begin{bmatrix} \boldsymbol{\Gamma} + \mu_X^2 \mathbf{1}_p \mathbf{1}_p^T & \mu_X \mathbf{1}_p \\ \mu_X \mathbf{1}_p^T & 1 \end{bmatrix},$$

and

$$\mathbf{V}^{-1} = \begin{bmatrix} \boldsymbol{\Gamma}^{-1} & -\mu_X \boldsymbol{\Gamma}^{-1} \mathbf{1}_p \\ -\mu_X \mathbf{1}_p^T \boldsymbol{\Gamma}^{-1} & 1 + \mu_X^2 \mathbf{1}_p^T \boldsymbol{\Gamma}^{-1} \mathbf{1}_p \end{bmatrix},$$

where  $\mathbf{1}_p$  is a  $p \times 1$  matrix of ones. The elements of the  $\mathbf{W}$  matrix are

$$W_{lj} = \sum_{k=1}^p \beta_k E(X_{p-l+1} X_{p-j+1} X_{p-k+1}) + \sigma_\epsilon^2 V_{lj}, \quad 1 \leq l, j \leq p.$$

This method is computationally more tractable than CML, however it does not take into account other conditional moment restrictions, like the conditional heteroscedasticity, leading to a possible loss of efficiency in comparison to CML estimation [Hadri, 2009].

We can estimate the innovation variance  $\sigma_\epsilon^2$  using a two-step CLS method [Karlsen and Tjøstheim, 1988, Ristić et al., 2012, Bourguignon et al., 2019] – we minimize

$$S_n(\boldsymbol{\eta}) = \sum_{t=p+1}^n \left[ \{X_t - \mu_{X_t|X_{t-1}, \dots, X_{t-p}}\}^2 - \sigma_{X_t|X_{t-1}, \dots, X_{t-p}}^2 \right]^2,$$

with respect to  $\sigma_\epsilon^2$ , while replacing  $\boldsymbol{\alpha}$  and  $\mu_\epsilon$  with the CLS estimates from (3.6). (Also see Freeland [2010], Bourguignon and Vasconcellos [2015b], and Huang and Zhu [2021].)

### 3.1.4 Pseudo maximum likelihood

For the pseudo maximum likelihood method we approximate the transition probability using normal distribution with mean and variance equal to the conditional mean and variance presented in Lemma 2.4. Let

$$\ell_P(\boldsymbol{\theta}) = -\frac{1}{2} \sum_{t=p+1}^n \log \left( 2\pi\sigma_{X_t|X_{t-1}, \dots, X_{t-p}}^2(\boldsymbol{\theta}) \right) + \sum_{t=p+1}^n \frac{(x_t - \mu_{X_t|X_{t-1}, \dots, X_{t-p}}(\boldsymbol{\theta}))^2}{2\sigma_{X_t|X_{t-1}, \dots, X_{t-p}}^2(\boldsymbol{\theta})}.$$

Then the pseudo maximum likelihood estimator  $\widehat{\boldsymbol{\theta}}_P$  is

$$\widehat{\boldsymbol{\theta}}_P = \arg \max_{\boldsymbol{\theta} \in \Theta} \ell_P(\boldsymbol{\theta}).$$

**Theorem 3.1.6** *Let  $\{X_t : t \in \mathbb{Z}\}$  be a GINAR( $p$ ) process. The pseudo maximum likelihood estimator,  $\widehat{\boldsymbol{\theta}}_P$ , has the asymptotic distribution*

$$\sqrt{n} \left( \widehat{\boldsymbol{\theta}}_P - \boldsymbol{\theta} \right) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}_P), \quad (3.7)$$

as  $n \rightarrow \infty$ , where

$$\begin{aligned} \boldsymbol{\Sigma}_P &= \mathbf{U}(\boldsymbol{\theta})^{-1} \mathbf{Z}(\boldsymbol{\theta}) \mathbf{U}(\boldsymbol{\theta})^{-1}, \\ \mathbf{Z}(\boldsymbol{\theta}) &= \lim_{n \rightarrow \infty} \frac{1}{n} E \left( \frac{\partial \ell_P(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \left( \frac{\partial \ell_P(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T \right), \end{aligned}$$

and

$$\mathbf{U}(\theta) = \lim_{n \rightarrow \infty} \frac{1}{n} E \left( \frac{\partial^2 \ell_P(\theta)}{\partial \theta \partial \theta^T} \right).$$

This method is computationally much more tractable compared to CML, and also has a relatively simple implementation.

### 3.1.5 Saddlepoint methods

Saddlepoint approximation techniques were introduced by Daniels [1954], and are used in a wide range of applications. An advantage of these methods is that they tend to provide accurate approximations even for small sample sizes, and is often believed to be more accurate than using a normal approximation [e.g., Reid, 1988, Davison, 2003]. Pedeli et al. [2015] propose the saddlepoint approximation method for the GINAR( $p$ ) case with binomial thinning. Their main idea is that the convolutions in the likelihood expression can be removed by considering the corresponding moment-generating functions. First, they note that the conditional cumulant generating function for the GINAR( $p$ ) process with binomial thinning and a general innovation distribution is

$$\begin{aligned} K_t(u) &= \log E[\exp(uX_t) | X_{t-1}, \dots, X_{t-p}] \\ &= \sum_{j=1}^p x_{t-j} \log(1 - \alpha_j + \alpha_j \exp(u)) + K_{\epsilon_t}(u). \end{aligned}$$

Generalizing to other thinning operators we get

$$K_t(u) = \sum_{j=1}^p K_{\alpha \odot X_{t-j}}(u) + K_{\epsilon_t}(u).$$

Then, an approximation to the true conditional log likelihood is provided by the saddlepoint approximation is

$$\ell_S(\boldsymbol{\theta}) = \sum_{t=p+1}^n \left( \frac{1}{2\pi K_t''(\tilde{u}_t)} \right)^{1/2} \exp(K_t(\tilde{u}_t) - \tilde{u}_t x_t),$$

where  $\tilde{u}_t$  is found by solving  $K'_t(u) = x_t$ , and the derivatives are taken with respect to  $u$ . [Pedeli et al. \[2015\]](#) propose using a quasi-likelihood based approach to show that the resulting estimator is asymptotically normal for the binomial thinning case, however they remark that the estimator is not consistent. They posit that the saddlepoint method is more computationally efficient than the CML method with transition probabilities calculated using the inefficient equation in [Theorem 2.5.1](#), especially for cases where observed counts are large.

### 3.1.6 Spectral-based Whittle estimation

We can also consider estimating the parameters on the basis of a spectral analysis of the process. The Whittle criterion is

$$\ell_W(\boldsymbol{\theta}) = \frac{1}{N} \sum_{j=1}^{\lfloor N/2 \rfloor} \left( \log f(\nu_j, \boldsymbol{\theta}) + \frac{I_N(\nu_j)}{f(\nu_j, \boldsymbol{\theta})} \right), \quad (3.8)$$

where  $f(\nu_j, \boldsymbol{\theta})$  is the spectral density function and  $I_N(\nu_j)$  is the periodogram at frequency  $\nu_j = 2\pi j/N$  [[Silva and Oliveira, 2005](#)]. The periodogram is defined as follows:

$$I_N(\nu_j) = (2\pi N)^{-1} \left| \sum_{t=1}^N X_t \exp(i\omega t) \right|^2, \quad j = 1, \dots, \lfloor N/2 \rfloor.$$

The Whittle estimate of  $\boldsymbol{\theta}$ ,  $\hat{\boldsymbol{\theta}}_W$ , is

$$\hat{\boldsymbol{\theta}}_W = \arg \min_{\boldsymbol{\eta} \in \Theta} \ell_W(\boldsymbol{\theta}).$$

[Rice \[1979\]](#) showed that the Whittle estimator for a non-Gaussian series is also asymptotically normally distributed. However, the asymptotic variance depends on the fourth order spectra of the process, which is difficult to obtain for the case of a GINAR process.

For a large class of processes the Whittle likelihood is known to provide biased estimates. Hence, [Sykulski et al. \[2019\]](#) propose the debiased Whittle likelihood which replaces the spectral density with the expected value of the periodogram while still being computationally

efficient with  $O(n \log n)$  efficiency. The authors show that employing this method leads to less biased estimates compared to using the Whittle likelihood. However this has not been investigated for the analysis of count time series.

### 3.1.7 Other methods

Other estimation methods have also been proposed. For instance, the squared difference (SD) estimator is proposed for the GINAR(1) process with binomial thinning and Poisson innovations have, which provides bias reduced estimates using a method of moments approach [Bourguignon and Vasconcellos, 2015a]. Bootstrap based techniques have also been studied. For instance, Jentsch and Weiß [2019] and Weiß and Jentsch [2019] propose some flexible bootstrap techniques for the GINAR( $p$ ) process with binomial thinning which could be extended to a generalized thinning case. Lu and Wang [2022] propose a new parameter estimator based on empirical likelihood for the GINAR(1) process with binomial thinning and Poisson innovations which has a small bias and is efficient as shown by simulation studies. Generalized method of moments have also been used for estimation for different models under the GINAR( $p$ ) class; some examples are Brannas [1993] and Zhang et al. [2010].

## 3.2 Inference

In this section we discuss methods of inference for GINAR( $p$ ) processes, in particular building confidence intervals, and provide a discussion of model selection methodologies for these processes.

### 3.2.1 Confidence regions

We can calculate confidence intervals for the model parameters based on the asymptotic theory given in Section 3.1. An estimate of the asymptotic variance for methods such as CML, CLS, Yule-Walker and pseudo maximum likelihood can be obtained by substituting the appropriate estimate into the expression for the asymptotic covariance matrix. We can also estimate the covariance matrix using the Hessian, calculated numerically from the optimization algorithm that generates the estimator.

For example, with an estimated covariance matrix for the parameters of  $\Sigma$ , an approximate  $(1 - \nu)100\%$  confidence interval for  $\theta_j$  is

$$\hat{\theta}_j \pm z_{1-\nu/2} [\Sigma]_{jj},$$

and a  $100(1 - \nu)\%$  simultaneous confidence region for  $\theta$  is

$$\left\{ \theta : (\hat{\theta} - \theta) \Sigma^{-1} (\hat{\theta} - \theta)^T \leq \frac{\chi_{p,1-\nu}^2}{n} \right\}.$$

Weiß [2011] use simulations to compare the confidence interval coverage for the CLS and CML methods for a GINAR(1) process with binomial thinning and Poisson innovations. They indicate that the CML method has better coverage properties. The CLS method has lower coverage when the dependence parameter  $\alpha_1$  is large for shorter time series.

### 3.2.2 Model selection

Model selection for  $\text{GINAR}(p)$  processes includes selection of the model order  $p$ , thinning operator, and innovation distribution. There are many techniques discussed in the literature to make these choices which are discussed in this section.

[Weiß and Feld \[2020\]](#) carry out a comprehensive simulation study on the performance of AIC and BIC for model selection in the context of count time series models. The authors choose candidate models to provide insight into tasks such as identifying serial correlation, overdispersion or zero-inflation, order selection, and non-nested model families. Their findings show that both AIC and BIC are generally successful model selection criteria – for shorter series, AIC is recommended, while BIC outperforms AIC for longer series.

[Alzahrani et al. \[2018\]](#) discuss other methods for model selection which are useful when the competing models are non-nested, which is usually the case. They develop an effective algorithm in the Bayesian framework, implementing direct computation of the marginal likelihood and using the deviance information criteria (DIC). Other than information criteria methods, the forecasting potential of different models can also be used as a method of model selection. For example, [Cardinal et al. \[1999\]](#) propose the relative forecast error (RFE) to measure the forecasting performance of a model which can then be used to select the optimal model. [da Silva \[2005\]](#) also propose to use the final prediction error to choose model order, among other metrics.

[Bu and McCabe \[2008\]](#) propose that model order can be selected sample autocorrelation/partial autocorrelation functions. They also present a model selection approach using a new residual process which is defined for each of the  $p+1$  model components (the  $p$  thinning operations of past values and the innovation sequence). Here they look at the residuals of each model component, in addition to residuals of the fitted model. They posit that this

allows for a more robust analysis into the goodness of fit of the model. [Freeland and McCabe \[2004a\]](#) also suggest defining two sets of residuals for the GINAR(1) process with binomial thinning, one for each random component part of the model. However, the challenge here is interpreting these residuals when the lags are higher. [Park and Kim \[2012\]](#) present two new expected residual calculations for the GINAR( $p$ ) process with binomial thinning, and demonstrate its usage in assessing model adequacy and predictive performance. They show that these residuals can be useful in model selection, identifying overdispersion and correlation arising from incorrect parameter estimates or choice of model order  $p$ . These residuals require that the conditional probabilities,  $P(X_t|X_{t-1}, \dots, X_{t-p})$  and  $P(\epsilon_t|X_{t-1}, \dots, X_{t-p})$  are specified, hence this method can be extended to the GINAR( $p$ ) case whenever the conditional distributions are available, or alternatively we can use the Davies approach to calculate the transition probabilities. [Forughi et al. \[2022\]](#) develops some portmanteau test statistics to check goodness of fit for GINAR( $p$ ) processes and derive asymptotic distributions of the test statistics.

### 3.3 Forecasting for GINAR( $p$ ) processes

We next discuss forecasting strategies for the GINAR( $p$ ) process. We first start with a summary of different methodologies used in the literature and then present the methodology used in this dissertation.

#### 3.3.1 Literature Review

When it comes to forecasting GINAR( $p$ ) processes, we ideally want a method that produces integer-valued forecasts. [Jin-Guan and Yuan \[1991\]](#) propose the minimum variance predictor of  $X_{n+1}$ , denoted  $\widehat{X}_n(1)$ , to be the conditional expectation,  $E(X_{n+1}|X_n, \dots, X_{n+1-p})$ .

Similarly then, the  $k$ -step ahead prediction is given by,

$$\begin{aligned}
\widehat{X}_n(k) &= E(X_{n+k}|X_1, \dots, X_n) = \sum_{j=1}^p E(\alpha_j \odot X_{n+k-j}) + \mu_\epsilon \\
&= \sum_{j=1}^p \alpha_j E(X_{n+k-j}|X_1, \dots, X_n) + \mu_\epsilon \\
&= \sum_{j=1}^p \alpha_j \widehat{X}_n(k-j) + \mu_\epsilon.
\end{aligned}$$

Although, this method produces forecasts with minimum mean square errors, it this has the disadvantage of producing non-integer forecasts due to the multiplications involved. Hence, various other methods have been proposed in the literature. [Freeland and McCabe \[2004b\]](#), present the  $k$ -step ahead predictive distribution for the GINAR(1) process with binomial thinning and Poisson innovations, but analytical solutions are not easily derived for higher order lags and other innovation distributions. They provide the asymptotic distribution of the  $k$ -step ahead pmf which can be used to construct confidence intervals for the probability associated with any value of  $x$  in the forecast distribution. [Bu and McCabe \[2008\]](#) propose a forecasting methodology for the GINAR( $p$ ) process with binomial thinning and Poisson innovations, using the transition probability function of the process. They treat the model as a Markov chain, where they assume a maximum count of  $M$ , and obtain forecasts using the transition matrix. They also posit a methodology for calculating confidence intervals for these forecast probabilities.

Other commonly used forecasting approach uses bootstrap based prediction methods, which allow for a distribution free approach, making them more flexible to incorporate higher order lags, and different thinning and innovation distributions. [Bisaglia and Gerolimetto \[2019\]](#), [Kim and Park \[2008\]](#), [Cardinal et al. \[1999\]](#) implement bootstrap based prediction methods. [Bisaglia and Gerolimetto \[2019\]](#) propose a model-based bootstrap approach to

estimate the probability mass function conditional on the data available at the time the forecast is made. They propose a new approach based on the autoregressive nature of INAR model that allows for the integer nature of the data. They propose a model-based bootstrap approach to estimate the pmf, conditional on the data available at the time the forecast is made. They also carry out an extensive simulation study comparing their proposed method to that of [Cardinal et al. \[1999\]](#) and [Kim and Park \[2008\]](#) show that their method is either better or comparable.

[Jung and Tremayne \[2006\]](#) propose a simulation based approach which is easily extendable to higher order lags and other innovation distributions and thinning operations.

Bayesian methods to forecast GINAR processes have also been studied in the literature. [McCabe and Martin \[2005\]](#) introduced a Bayesian method to produce coherent forecasts for low-count time series data, and specifically look at GINAR(1) processes with binomial thinning and any innovation distribution. They produce forecasts using a  $p$ -step ahead predictive probability mass function (pmf) defined as follows:

$$P(X_{n+p} = x_{n+p}|\mathbf{x}) = \sum_{k=1}^K P(X_{n+p} = x_{n+p}|\mathbf{x}, M_k)P(M_k|\mathbf{x}),$$

where  $\{X_t : t \in \mathbb{Z}\}$  is as defined in Definition 2.2.1. The process generating  $X_t$  is assumed to be within a set of pre-defined  $K$  processes denoted  $M_k$  for  $k = 1, 2, \dots, K$ , and  $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ , the vector of observed data. Note that  $P(X_{n+p} = x_{n+p}|\mathbf{x}, M_k)$  is the  $p$ -step ahead predictive pmf for the  $k^{th}$  process considered and  $P(M_k|\mathbf{x})$  is the posterior probability of process  $M_k$ . [McCabe and Martin \[2005\]](#) use Bayesian methods to simplify this equation and use numerical approaches for the evaluation. This approach relies on parametric assumptions for the innovation distribution, and also, if the true distribution is not included in the set of processes considered it could lead to misspecification errors and poor performance. To circumvent the need for parametric assumptions on the innovation

distribution, [Bisaglia and Canale \[2016\]](#) propose Bayesian nonparametric forecasting methods for INAR models. For the innovation distribution they use a Dirichlet process mixture of rounded Gaussians which allows for a large support on the space and pmf, and allows to model various count distributions. Another non-parametric approach is proposed by [McCabe et al. \[2011\]](#); they estimate the forecast distribution of the discrete random variable non-parametrically. For the broad INAR class they develop an asymptotically efficient estimator of the forecast distribution using non-parametric maximum likelihood. They show the merits of their methodology by comparing performance to misspecified parametric models.

### 3.3.2 A simulation-based forecasting method

In the previous section we provided a discussion on the various forecasting methods studied in the literature. In this section we describe the forecasting methodology used for the simulations and applications in this dissertation.

We use a simulation based approach for forecasting so that it can easily be extended to higher-order lags and other innovation distributions and thinning operators. The prediction distribution is calculated by means of a Monte Carlo procedure. Then, the median of this distribution can be used as a point estimate and a confidence interval can be obtained from the quantiles. This ensures that we always obtain an integer-valued forecast. Then for a positive integer  $n$  the steps for calculating the  $h$ -step ahead forecast distribution, denoted  $X_{n+h}$ , for a  $\text{GINAR}(p)$  process is as follows:

1. Generate one realization of  $\hat{\alpha}_j \odot X_{n+h-j}^b = \sum_{k=1}^{X_{n+h-j}^b} Y_k^j$  where  $\{Y_k^j\}$  are IID RVs as described in [Definition 2.1.1](#), for  $j = 1, 2, \dots, p$ ;
2. Generate one realization of  $\{\epsilon_{n+h}^b\}$  which are IID RVs from a pre-specified non-negative distribution with parameters  $\hat{\mu}_\epsilon$  and  $\hat{\sigma}_\epsilon^2$ ;

3. Set  $X_{N+h}^b = \sum_{j=1}^p \hat{\alpha}_j \odot X_{n+h-j}^b + \epsilon_{n+h}^b$ ;

4. Repeat steps 1-3,  $b$  times, for large  $b$ .

Then we can estimate a percentile-based  $100(1 - \nu)\%$  forecast interval for the prediction.

## Chapter 4: Simulation Studies for GINAR( $p$ ) Processes

Up to this point we have reviewed and discussed a number of different estimation, inference, and forecasting methods that can be applied to the statistical analysis of GINAR( $p$ ) processes. Now we compare the performance of different estimation methods for finite samples, taking time to compare the performance as we vary the order  $p$ , as well as the thinning operator and innovation process used to define the GINAR process.

Throughout, Po-INAR refers to a GINAR process with binomial thinning and Poisson innovations, NB-INAR refers to binomial thinning and negative binomial innovations, and Geom-INAR refers to negative binomial thinning defined with geometric random variables along with Poisson innovations. In all simulations we estimate each quantity of interest (the bias, standard deviation (SD), root mean squared error (RMSE), or coverage) using 10,000 replicates, and estimate standard errors for each quantity using 10,000 bootstrap samples. We present GINAR(1) simulations in Section 4.1, GINAR(2) simulations in Section 4.2 and GINAR(4) simulations in Section 4.3. In Section 4.4 we present coverage simulation for some estimation methods for the Po-INAR(1) and Po-INAR(2) processes.

## 4.1 Estimating GINAR(1) process parameters

In this section we look at simulation studies for the estimation of GINAR(1) process parameters. In particular we consider the estimation of parameters of the Po-INAR(1), Geom-INAR(1) and NB-INAR(1) processes.

### Varying the Sample Size

In our first simulation we estimate the parameters of the most commonly used Po-INAR(1) process when  $\alpha_1 = 0.5$  and  $\mu_\epsilon = 1$ . Table 4.1 shows estimates of the bias, SD, and RMSE for both parameters of the process using different estimation methods as we vary the sample length  $n$ , with  $n = 100, 500$  and  $1,000$ . For most estimation methods, we learn that the bias tends to get closer to zero as we increase  $n$ , and that naturally the SD and RMSE gets smaller as  $n$  is increased. However, due to the lack of consistency discussed in Section 3, the bias of the saddlepoint estimator does not go to zero. For longer sample lengths the CML method tends to do best in terms of bias, SD, and RMSE, although the pseudo-likelihood method is comparable in performance.

There is a loss of efficiency for other estimators. For the CML method we found no difference in the results when using the slow transition probability calculation given by Theorem 2.5.1 and the much faster chf-based calculation in Proposition 2.5.3, evaluating the integrals using Gauss quadrature with 300 weights.

In Table 4.2 we move to using a NB-INAR(1) process instead of the Po-INAR(1) process. Both processes are defined with binomial thinning but we now use negative binomial innovations with mean  $\mu_\epsilon$  and variance  $\sigma_\epsilon^2 = \mu_\epsilon + r\mu_\epsilon^2$  with  $r = 1$  for the NB-INAR(1) process, generating more overdispersion as compared to Po-INAR(1) process. In general the

Table 4.1: Estimated bias, SD, and RMSE when estimating the parameters of a Po-INAR(1) process with  $\alpha_1 = 0.5$  and  $\mu_\epsilon = 1$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.002 for estimating  $\alpha_1$  and 0.004 for estimating  $\mu_\epsilon$ .

Method		$n = 100$		$n = 500$		$n = 1,000$	
		$\hat{\alpha}_1$	$\hat{\mu}_\epsilon$	$\hat{\alpha}_1$	$\hat{\mu}_\epsilon$	$\hat{\alpha}_1$	$\hat{\mu}_\epsilon$
CML	Bias	-0.009	0.013	-0.002	0.003	-0.001	0.001
	SD	0.076	0.166	0.032	0.073	0.023	0.051
	RMSE	0.076	0.166	0.033	0.073	0.023	0.051
CLS	Bias	-0.026	0.049	-0.006	0.011	-0.002	0.004
	SD	0.093	0.203	0.042	0.089	0.030	0.063
	RMSE	0.097	0.209	0.043	0.090	0.030	0.063
YW	Bias	-0.032	0.059	-0.006	0.012	-0.003	0.005
	SD	0.093	0.206	0.042	0.089	0.029	0.063
	RMSE	0.099	0.215	0.042	0.090	0.029	0.063
Pseudo	Bias	-0.011	0.008	-0.002	0.002	-0.001	0.001
	SD	0.083	0.176	0.035	0.077	0.025	0.055
	RMSE	0.083	0.176	0.035	0.077	0.025	0.055
Whittle	Bias	-0.017	0.012	-0.003	0.003	-0.002	0.000
	SD	0.094	0.187	0.042	0.083	0.030	0.058
	RMSE	0.096	0.187	0.042	0.083	0.030	0.058
Saddle	Bias	-0.045	0.086	-0.037	0.073	-0.036	0.071
	SD	0.075	0.169	0.033	0.073	0.023	0.052
	RMSE	0.087	0.189	0.049	0.103	0.043	0.088

Table 4.2: Estimated bias, SD, and RMSE when estimating the parameters of a NB-INAR(1) process with  $\alpha_1 = 0.5$ ,  $r = 1$  and  $\mu_\epsilon = 1$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.006 for estimating  $\alpha_1$ , 0.05 for estimating  $r$ , and 0.05 for estimating  $\mu_\epsilon$ .

Method		$n = 100$			$n = 500$			$n = 1,000$		
		$\hat{\alpha}_1$	$\hat{\mu}_\epsilon$	$\hat{r}$	$\hat{\alpha}_1$	$\hat{\mu}_\epsilon$	$\hat{r}$	$\hat{\alpha}_1$	$\hat{\mu}_\epsilon$	$\hat{r}$
CML	Bias	-0.013	0.021	-0.014	-0.002	0.003	-0.005	-0.002	0.002	-0.001
	SD	0.068	0.184	0.505	0.028	0.079	0.216	0.020	0.055	0.151
	RMSE	0.069	0.186	0.505	0.029	0.079	0.216	0.020	0.055	0.151
CLS	Bias	-0.030	0.054	-0.071	-0.006	0.011	-0.013	-0.003	0.004	-0.006
	SD	0.095	0.233	0.713	0.043	0.101	0.315	0.030	0.072	0.225
	RMSE	0.099	0.239	0.716	0.043	0.102	0.315	0.031	0.072	0.225
YW	Bias	-0.031	0.057	-0.058	-0.006	0.011	-0.010	-0.003	0.008	-0.004
	SD	0.094	0.233	0.705	0.042	0.103	0.316	0.030	0.073	0.224
	RMSE	0.098	0.240	0.708	0.043	0.103	0.316	0.031	0.073	0.224
Pseudo	Bias	-0.029	0.053	-0.084	-0.006	0.011	-0.024	-0.003	0.005	-0.009
	SD	0.098	0.240	0.578	0.043	0.102	0.267	0.030	0.071	0.190
	RMSE	0.103	0.245	0.584	0.043	0.102	0.268	0.030	0.071	0.190
Whittle	Bias	-0.013	0.053	-0.040	-0.001	0.008	0.002	0.000	0.000	0.016
	SD	0.094	0.227	0.690	0.043	0.101	0.318	0.030	0.070	0.217
	RMSE	0.095	0.233	0.691	0.043	0.102	0.318	0.030	0.069	0.217
Saddle	Bias	-0.075	0.147	-0.490	-0.064	0.127	-0.468	-0.063	0.123	-0.466
	SD	0.074	0.208	0.318	0.034	0.146	0.142	0.028	0.066	0.099
	RMSE	0.105	0.254	0.584	0.072	0.193	0.489	0.069	0.140	0.477

Table 4.3: Estimated bias, SD, and RMSE when estimating the parameters of a Geom-INAR(1) process with  $\alpha_1 = 0.5$  and  $\mu_\epsilon = 1$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.002 for estimating  $\alpha_1$  and 0.007 for estimating  $\mu_\epsilon$ .

Method		$n = 100$		$n = 500$		$n = 1,000$	
		$\hat{\alpha}_1$	$\hat{\mu}_\epsilon$	$\hat{\alpha}_1$	$\hat{\mu}_\epsilon$	$\hat{\alpha}_1$	$\hat{\mu}_\epsilon$
CML	Bias	-0.027	0.041	-0.006	0.010	-0.003	0.004
	SD	0.096	0.182	0.042	0.079	0.029	0.055
	RMSE	0.100	0.186	0.042	0.080	0.029	0.055
CLS	Bias	-0.040	0.066	-0.009	0.014	-0.004	0.007
	SD	0.109	0.212	0.051	0.098	0.036	0.069
	RMSE	0.116	0.222	0.052	0.099	0.037	0.069
YW	Bias	-0.042	0.073	-0.009	0.015	-0.004	0.008
	SD	0.106	0.209	0.051	0.097	0.036	0.069
	RMSE	0.114	0.221	0.051	0.098	0.036	0.070
Pseudo	Bias	-0.022	0.022	-0.005	0.006	-0.002	0.001
	SD	0.106	0.210	0.045	0.091	0.032	0.063
	RMSE	0.108	0.211	0.046	0.091	0.032	0.063
Whittle	Bias	-0.029	0.060	-0.006	0.014	-0.003	0.006
	SD	0.109	0.338	0.050	0.151	0.036	0.107
	RMSE	0.113	0.344	0.051	0.152	0.036	0.107

results stay consistent with the results to the Po-INAR(1) simulation. However, the pseudo-likelihood estimator does worse in comparison to CML and the saddlepoint estimator also performs much worse, in terms of bias, SD and RMSE. Also, it is much harder to estimate  $r$  compared to the other parameters of the model. In particular, for the saddlepoint estimation method we had to restrict  $r$  values to be between zero and five in the optimization step to obtain reasonable results. One can clearly see the merits of using CML here as it tends to do best in terms of bias, SD and RMSE.

Table 4.3 compares the performance of different estimators when we simulate Geom-INAR(1) processes with  $\alpha_1 = 0.5$  and  $\mu_\epsilon = 1$ , at the same sample lengths  $n$ . For this process

we have negative binomial thinning and Poisson innovations, leading to overdispersion that varies over time. Also for the Geom-INAR(1) case the saddlepoint method is not available, and thus is not shown. The simulation results in Table 4.3 show that CML and pseudo-likelihood are comparable in terms of Bias, SD and RMSE. We see that Whittle likelihood estimates perform much worse as compared to the Po-INAR(1) simulation, especially for smaller sample lengths with the RMSE for  $\mu_\epsilon$  being almost two times larger than that of CML.

### Varying the Innovation Mean

We now estimate the parameter for the Po-INAR(1), NB-INAR(1) and Geom-INAR(1) processes when varying the innovation mean. We consider values  $\mu_\epsilon = 5, 10,$  and  $20$  for the innovation mean.

When varying the innovation mean, we see that the saddlepoint estimate of all parameters except  $\mu_\epsilon$  improves in terms of RMSE for larger values of  $\mu_\epsilon$  for all three processes. However, it has much worse bias properties than all other estimation methods for smaller values of  $\mu_\epsilon$ , for both NB-INAR(1) and Po-INAR(1) processes. A large bias and small RMSE suggests that saddlepoint estimator is consistently producing estimates that are far off from the true values, which is not desirable.

The simulation results in Table 4.4 show that the CML estimation outperforms most methods, however for the Po-INAR(1) process the pseudo maximum likelihood estimator is comparable to CML as we increase the innovation mean. This is not the case for NB-INAR(1) as pseudo performs worse, with larger RMSE and bias, as the innovation mean increases. We know the marginal distribution of the Po-INAR(1) process to be  $\text{Poisson}(\mu_\epsilon/(1 - \alpha))$ , and theoretically a  $\text{Poisson}(\lambda)$  random variable approaches a normal distribution as  $\lambda$  goes to infinity. This is the mathematical intuition for the pseudo estimator performing well when we

increase the innovation mean for the Po-INAR(1) case, but not NB-INAR(1). The results of the Geom-INAR(1) process as seen in Table 4.6 are also consistent with Po-INAR(1) results.

### Varying the Dependence Parameter

We consider varying the dependence parameter of the Po-INAR(1), Geom-INAR(1), and NB-INAR(1) processes. We consider the values  $\alpha_1 = 0.2, 0.5,$  and  $0.8$ . Varying the values of the dependence parameter shows that the CML method has the lowest RMSE for all parameters estimated for all three processes and most methods have large bias, SD and RMSE as  $\alpha_1$  gets close to the boundary point of  $\alpha_1 = 1$ , as seen in Tables 4.7, 4.8, and 4.9.

When the dependence parameter is closer to the boundary point; i.e. the simulations where  $\alpha_1 = 0.8$ , CML outperforms other methods by a much larger factor. For instance, in the NB-INAR(1) simulation, the RMSE for  $\hat{\mu}_\epsilon$  is almost 40% smaller than the next best estimator - pseudo maximum likelihood. Note that for this simulation the saddlepoint estimator has a RMSE almost 57 times larger than CML, this is clearly not a desirable estimator to use when the dependence parameter is believed to be large.

Table 4.4: Estimated bias, SD, and RMSE when estimating the parameters of a Po-INAR(1) process with  $\alpha_1 = 0.5$  and  $n = 500$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.001 for estimating  $\alpha_1$  and 0.031 for estimating  $\mu_\epsilon$ .

Method		$\mu_\epsilon = 5$		$\mu_\epsilon = 10$		$\mu_\epsilon = 20$	
		$\hat{\alpha}_1$	$\hat{\mu}_\epsilon$	$\hat{\alpha}_1$	$\hat{\mu}_\epsilon$	$\hat{\alpha}_1$	$\hat{\mu}_\epsilon$
CML	Bias	-0.001	0.011	-0.002	0.029	-0.001	0.029
	SD	0.031	0.314	0.031	0.625	0.030	1.224
	RMSE	0.031	0.314	0.031	0.626	0.030	1.224
CLS	Bias	-0.005	0.051	-0.005	0.106	-0.005	0.189
	SD	0.039	0.399	0.040	0.800	0.039	1.554
	RMSE	0.040	0.403	0.040	0.807	0.039	1.565
YW	Bias	-0.006	0.060	-0.006	0.118	-0.007	0.268
	SD	0.040	0.405	0.039	0.796	0.039	1.555
	RMSE	0.040	0.409	0.040	0.805	0.039	1.578
Pseudo	Bias	-0.001	0.012	-0.001	0.025	-0.001	0.052
	SD	0.031	0.320	0.031	0.621	0.030	1.224
	RMSE	0.031	0.320	0.031	0.622	0.030	1.225
Whittle	Bias	-0.003	0.005	-0.004	0.018	-0.004	0.037
	SD	0.039	0.363	0.040	0.709	0.039	1.390
	RMSE	0.040	0.363	0.040	0.709	0.039	1.391
Saddle	Bias	-0.011	0.111	-0.007	0.129	-0.004	0.161
	SD	0.031	0.313	0.030	0.613	0.031	1.226
	RMSE	0.033	0.332	0.031	0.626	0.031	1.237

Table 4.5: Estimated bias, SD, and RMSE when estimating the parameters of a NB-INAR(1) process with  $\alpha_1 = 0.5$ ,  $r = 1$  and  $n = 500$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.001 for estimating  $\alpha_1$ , 0.004 for estimating  $r$  and 0.036 for estimating  $\mu_\epsilon$ .

Method		$\mu_\epsilon = 5$			$\mu_\epsilon = 10$			$\mu_\epsilon = 20$		
		$\hat{\alpha}_1$	$\hat{\mu}_\epsilon$	$\hat{r}$	$\hat{\alpha}_1$	$\hat{\mu}_\epsilon$	$\hat{r}$	$\hat{\alpha}_1$	$\hat{\mu}_\epsilon$	$\hat{r}$
CML	Bias	-0.002	0.020	-0.002	-0.001	0.009	0.000	0.000	0.025	0.000
	SD	0.021	0.314	0.136	0.017	0.574	0.118	0.014	1.062	0.105
	RMSE	0.021	0.314	0.136	0.017	0.574	0.118	0.014	1.062	0.105
CLS	Bias	-0.006	0.055	-0.010	-0.005	0.104	-0.009	-0.005	0.227	-0.011
	SD	0.040	0.465	0.210	0.039	0.922	0.192	0.039	1.796	0.186
	RMSE	0.040	0.468	0.210	0.040	0.928	0.192	0.039	1.811	0.186
YW	Bias	-0.007	0.067	-0.014	-0.007	0.130	-0.013	-0.006	0.242	-0.009
	SD	0.040	0.475	0.205	0.039	0.914	0.192	0.039	1.797	0.187
	RMSE	0.041	0.480	0.206	0.040	0.923	0.193	0.039	1.813	0.187
Pseudo	Bias	-0.005	0.049	-0.009	-0.006	0.115	-0.010	-0.005	0.200	-0.008
	SD	0.040	0.462	0.196	0.039	0.919	0.188	0.039	1.815	0.184
	RMSE	0.040	0.464	0.196	0.040	0.926	0.188	0.039	1.826	0.184
Whittle	Bias	-0.003	-0.012	0.023	-0.003	-0.060	0.031	-0.003	-0.269	0.044
	SD	0.040	0.462	0.212	0.040	0.893	0.200	0.040	1.798	0.200
	RMSE	0.040	0.462	0.213	0.040	0.895	0.202	0.040	1.818	0.205
Saddle	Bias	-0.024	0.242	-0.101	-0.009	0.182	0.013	0.000	-0.023	0.093
	SD	0.022	0.322	0.124	0.018	0.572	0.123	0.014	1.045	0.119
	RMSE	0.033	0.403	0.160	0.020	0.600	0.123	0.014	1.046	0.151

Table 4.6: Estimated bias, SD, and RMSE when estimating the parameters of a Geom-INAR(1) process with  $\alpha_1 = 0.5$  and  $n = 500$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.001 for estimating  $\alpha_1$  and 0.050 for estimating  $\mu_\epsilon$ .

Method		$\mu_\epsilon = 5$		$\mu_\epsilon = 10$		$\mu_\epsilon = 20$	
		$\hat{\alpha}_1$	$\hat{\mu}_\epsilon$	$\hat{\alpha}_1$	$\hat{\mu}_\epsilon$	$\hat{\alpha}_1$	$\hat{\mu}_\epsilon$
CML	Bias	-0.005	0.048	-0.006	0.118	-0.006	0.238
	SD	0.036	0.353	0.036	0.712	0.036	1.411
	RMSE	0.037	0.356	0.036	0.721	0.036	1.431
CLS	Bias	-0.006	0.057	-0.005	0.099	-0.006	0.222
	SD	0.042	0.415	0.040	0.798	0.040	1.588
	RMSE	0.042	0.418	0.041	0.804	0.040	1.604
YW	Bias	-0.006	0.061	-0.007	0.129	-0.006	0.229
	SD	0.041	0.409	0.040	0.793	0.040	1.587
	RMSE	0.042	0.413	0.041	0.804	0.040	1.603
Pseudo	Bias	-0.006	0.053	-0.006	0.109	-0.006	0.236
	SD	0.038	0.379	0.037	0.729	0.036	1.415
	RMSE	0.039	0.383	0.037	0.737	0.036	1.435
Whittle	Bias	-0.004	0.058	-0.004	0.097	-0.003	0.187
	SD	0.042	0.663	0.041	1.287	0.040	2.567
	RMSE	0.042	0.665	0.041	1.290	0.040	2.574

Table 4.7: Estimated bias, SD, and RMSE when estimating the parameters of a Po-INAR(1) process with  $\mu_\epsilon = 1$  and  $n = 500$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.002 for estimating  $\alpha_1$  and 0.011 for estimating  $\mu_\epsilon$ .

Method		$\alpha_1 = 0.2$		$\alpha_1 = 0.5$		$\alpha_1 = 0.8$	
		$\hat{\alpha}_1$	$\hat{\mu}_\epsilon$	$\hat{\alpha}_1$	$\hat{\mu}_\epsilon$	$\hat{\alpha}_1$	$\hat{\mu}_\epsilon$
CML	Bias	-0.002	0.002	-0.001	0.002	-0.001	0.004
	SD	0.045	0.069	0.032	0.073	0.014	0.074
	RMSE	0.045	0.069	0.032	0.073	0.014	0.075
CLS	Bias	-0.003	0.003	-0.006	0.011	0.064	-0.330
	SD	0.047	0.071	0.041	0.089	0.111	0.552
	RMSE	0.047	0.071	0.042	0.090	0.128	0.643
YW	Bias	-0.004	0.004	-0.006	0.010	-0.009	0.042
	SD	0.047	0.071	0.042	0.089	0.028	0.145
	RMSE	0.047	0.071	0.043	0.090	0.030	0.151
Pseudo	Bias	-0.002	0.002	-0.002	0.003	-0.001	0.003
	SD	0.050	0.074	0.035	0.078	0.015	0.078
	RMSE	0.050	0.074	0.035	0.078	0.015	0.078
Whittle	Bias	-0.001	0.000	-0.003	0.002	-0.005	0.007
	SD	0.047	0.085	0.042	0.082	0.029	0.081
	RMSE	0.047	0.085	0.042	0.082	0.029	0.081
Saddle	Bias	-0.022	0.027	-0.037	0.072	-0.012	0.059
	SD	0.039	0.064	0.033	0.072	0.016	0.079
	RMSE	0.045	0.070	0.049	0.102	0.020	0.099

## 4.2 Estimating GINAR(2) process parameters

In this section we consider estimation of GINAR processes when the order of the process,  $p$ , is higher, with  $p = 2$ . We estimate the parameters of Po-INAR(2), Geom-INAR(2), and NB-INAR(2) processes.

### Varying the Sample Size

We estimate the parameters of Po-INAR(2), Geom-INAR(2), and NB-INAR(2) processes when  $\boldsymbol{\alpha} = (0.2, 0.4)^T$  with  $\mu_\epsilon = 1$  when the sample length is  $n = 100$  and  $n = 500$ . We

Table 4.8: Estimated bias, SD, and RMSE when estimating the parameters of a NB-INAR(1) process with  $\mu_\epsilon = 1$ ,  $n = 500$  and  $r = 1$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.005 for estimating  $\alpha_1$ , and 0.012 for estimating  $r$ , and 2.665 for estimating  $\mu_\epsilon$ .

Method		$\alpha = 0.2$			$\alpha = 0.5$			$\alpha = 0.8$		
		$\hat{\alpha}_1$	$\hat{\mu}_\epsilon$	$\hat{r}$	$\hat{\alpha}_1$	$\hat{\mu}_\epsilon$	$\hat{r}$	$\hat{\alpha}_1$	$\hat{\mu}_\epsilon$	$\hat{r}$
CML	Bias	-0.002	0.002	-0.003	-0.002	0.003	-0.007	-0.001	0.004	-0.008
	SD	0.036	0.075	0.192	0.029	0.080	0.214	0.014	0.083	0.238
	RMSE	0.036	0.075	0.192	0.029	0.080	0.214	0.014	0.084	0.238
CLS	Bias	-0.003	0.005	-0.011	-0.005	0.010	-0.004	-0.007	0.035	-0.026
	SD	0.048	0.084	0.231	0.043	0.103	0.322	0.029	0.155	0.542
	RMSE	0.048	0.084	0.231	0.043	0.103	0.322	0.030	0.159	0.543
YW	Bias	-0.004	0.004	-0.008	-0.007	0.014	-0.018	-0.009	0.044	-0.033
	SD	0.048	0.085	0.231	0.042	0.102	0.308	0.029	0.155	0.541
	RMSE	0.048	0.085	0.231	0.043	0.103	0.308	0.030	0.161	0.542
Pseudo	Bias	-0.005	0.006	-0.018	-0.006	0.011	-0.017	-0.005	0.025	-0.032
	SD	0.059	0.094	0.230	0.043	0.102	0.275	0.024	0.130	0.410
	RMSE	0.059	0.094	0.231	0.044	0.103	0.276	0.024	0.132	0.411
Whittle	Bias	-0.001	-0.004	0.021	-0.004	0.011	-0.006	-0.005	0.035	0.002
	SD	0.048	0.083	0.234	0.043	0.102	0.321	0.029	0.154	0.552
	RMSE	0.048	0.083	0.235	0.043	0.102	0.321	0.029	0.158	0.552
Saddle	Bias	-0.075	0.095	-0.145	-0.064	0.129	-0.468	-0.041	0.317	-0.726
	SD	0.039	0.082	0.157	0.036	0.191	0.142	0.030	4.850	0.132
	RMSE	0.085	0.125	0.214	0.073	0.230	0.489	0.051	4.859	0.738

see that for the Po-INAR(2) and Geom-INAR(2) processes the CML, pseudo, and Whittle estimation methods have comparable performance with low bias, SD and RMSE (Tables 4.10 and 4.12). CML has a lower RMSE when estimating  $\alpha_1$  and  $\alpha_2$  for both  $n = 100$  and  $n = 500$ . It also seems easier for almost all estimation methods in the Po-INAR(1) case to estimate  $\alpha_2$  as compared to  $\alpha_1$ .

For the NB-INAR(2) process, the Whittle and pseudo estimators are no longer comparable to CML. Here CML has the best performance in terms of bias, SD and RMSE (Table 4.11). Also note that in order for the optimization to run for the Whittle and saddlepoint

Table 4.9: Estimated bias, SD, and RMSE when estimating the parameters of a Geom-INAR(1) process with  $\mu_\epsilon = 1$  and  $n = 500$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.001 for estimating  $\alpha_1$  and 0.004 for estimating  $\mu_\epsilon$ .

Method		$\alpha_1 = 0.2$		$\alpha_1 = 0.5$		$\alpha_1 = 0.8$	
		$\hat{\alpha}_1$	$\hat{\mu}_\epsilon$	$\hat{\alpha}_1$	$\hat{\mu}_\epsilon$	$\hat{\alpha}_1$	$\hat{\mu}_\epsilon$
CML	Bias	-0.005	0.005	-0.006	0.008	-0.007	0.017
	SD	0.045	0.068	0.042	0.078	0.031	0.105
	RMSE	0.045	0.069	0.042	0.078	0.032	0.106
CLS	Bias	-0.003	0.003	-0.008	0.014	-0.015	0.058
	SD	0.048	0.071	0.050	0.095	0.040	0.170
	RMSE	0.048	0.071	0.051	0.096	0.043	0.179
YW	Bias	-0.005	0.006	-0.010	0.016	-0.016	0.062
	SD	0.048	0.071	0.051	0.096	0.041	0.170
	RMSE	0.049	0.072	0.052	0.098	0.044	0.181
Pseudo	Bias	-0.004	0.002	-0.004	0.004	-0.005	0.012
	SD	0.049	0.076	0.045	0.090	0.034	0.126
	RMSE	0.049	0.076	0.046	0.090	0.034	0.127
Whittle	Bias	-0.002	-0.001	-0.007	0.017	-0.011	0.062
	SD	0.049	0.106	0.050	0.152	0.041	0.228
	RMSE	0.049	0.106	0.051	0.153	0.042	0.237

Table 4.10: Estimated bias, SD, and RMSE when estimating the parameters of a Po-INAR(2) process with  $\alpha_1 = 0.2$ ,  $\alpha_2 = 0.4$  and  $\mu_\epsilon = 1$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.002 for estimating  $\alpha_1$ , 0.002 for estimating  $\alpha_2$  and 0.006 for estimating  $\mu_\epsilon$ .

Method		$n = 100$			$n = 500$		
		$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\mu}_\epsilon$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\mu}_\epsilon$
CML	Bias	-0.011	-0.022	0.071	-0.002	-0.004	0.014
	SD	0.099	0.093	0.266	0.042	0.039	0.112
	RMSE	0.100	0.095	0.275	0.042	0.039	0.113
CLS	Bias	0.130	-0.005	0.137	0.129	0.000	0.027
	SD	0.029	0.030	0.312	0.013	0.013	0.130
	RMSE	0.134	0.030	0.341	0.130	0.013	0.133
Y-W	Bias	-0.017	-0.049	0.155	-0.004	-0.010	0.033
	SD	0.100	0.097	0.307	0.044	0.044	0.131
	RMSE	0.101	0.109	0.344	0.044	0.045	0.135
Pseudo	Bias	-0.009	-0.024	0.061	-0.002	-0.005	0.012
	SD	0.105	0.100	0.275	0.044	0.041	0.115
	RMSE	0.105	0.102	0.282	0.044	0.041	0.115
Whittle	Bias	0.000	-0.036	0.044	0.000	-0.007	0.009
	SD	0.098	0.101	0.266	0.043	0.044	0.116
	RMSE	0.098	0.107	0.269	0.043	0.045	0.117
Saddle	Bias	-0.002	-0.040	0.091	0.009	-0.022	0.027
	SD	0.094	0.083	0.287	0.040	0.035	0.121
	RMSE	0.094	0.092	0.301	0.041	0.041	0.124

Table 4.11: Estimated bias, SD, and RMSE when estimating the parameters of a NB-INAR(2) process with  $\alpha_1 = 0.2$ ,  $\alpha_2 = 0.4$ ,  $r = 1$  and  $\mu_\epsilon = 1$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.002 for estimating  $\alpha_1$ , 0.002 for estimating  $\alpha_2$ , 0.007 for estimating  $\mu_\epsilon$  and 0.039 for estimating  $r$ .

Method		$n = 100$				$n = 500$			
		$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\mu}_\epsilon$	$\hat{r}$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\mu}_\epsilon$	$\hat{r}$
CML	Bias	-0.011	-0.027	0.087	-0.069	-0.003	-0.004	0.015	-0.011
	SD	0.090	0.085	0.282	0.699	0.038	0.036	0.112	0.290
	RMSE	0.090	0.089	0.295	0.703	0.038	0.036	0.113	0.290
CLS	Bias	-0.018	-0.039	0.130	-0.155	-0.003	-0.008	0.029	-0.033
	SD	0.099	0.100	0.329	0.805	0.044	0.044	0.139	0.382
	RMSE	0.101	0.107	0.354	0.820	0.044	0.045	0.142	0.383
Y-W	Bias	-0.017	-0.048	0.153	-0.146	-0.003	-0.010	0.031	-0.024
	SD	0.100	0.098	0.328	0.816	0.044	0.045	0.140	0.395
	RMSE	0.102	0.109	0.362	0.829	0.044	0.046	0.143	0.396
Pseudo	Bias	-0.017	-0.041	0.134	-0.144	-0.004	-0.008	0.028	-0.033
	SD	0.109	0.110	0.353	0.756	0.051	0.048	0.150	0.352
	RMSE	0.111	0.117	0.378	0.769	0.051	0.048	0.152	0.353
Whittle	Bias	0.000	-0.033	0.055	-0.026	-0.001	-0.007	-0.067	0.792
	SD	0.100	0.102	0.301	0.585	0.044	0.045	0.264	1.136
	RMSE	0.100	0.107	0.306	0.585	0.044	0.046	0.272	1.385
Saddle	Bias	-0.036	-0.075	0.269	-0.639	-0.028	-0.056	0.209	-0.604
	SD	0.085	0.078	0.297	0.280	0.038	0.034	0.128	0.127
	RMSE	0.092	0.108	0.401	0.698	0.047	0.065	0.245	0.617

likelihood methods we had to constrain the  $r$  parameter to be in between one and five. This is already a drawback of these two methods as we need to decide on a range for  $r$ , which may not be appropriate for real-world applications. Note that [Pedeli et al. \[2015\]](#) fixed the  $r$  parameter in their simulations for the saddlepoint method and thus we cannot compare our results to theirs.

Table 4.12: Estimated bias, SD, and RMSE when estimating the parameters of a Geom-INAR(2) process with  $\alpha_1 = 0.2$ ,  $\alpha_2 = 0.4$  and  $\mu_\epsilon = 1$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.002 for estimating  $\alpha_1$ , 0.003 for estimating  $\alpha_2$  and 0.008 for estimating  $\mu_\epsilon$ .

Method		$n = 100$			$n = 500$		
		$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\mu}_\epsilon$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\mu}_\epsilon$
CML	Bias	-0.014	-0.042	0.119	-0.002	-0.009	0.024
	SD	0.095	0.102	0.285	0.042	0.044	0.115
	RMSE	0.096	0.110	0.309	0.042	0.045	0.118
CLS	Bias	-0.017	-0.044	0.135	-0.004	-0.010	0.030
	SD	0.107	0.120	0.335	0.047	0.050	0.138
	RMSE	0.109	0.128	0.361	0.047	0.051	0.141
Y-W	Bias	-0.018	-0.056	0.169	-0.003	-0.012	0.034
	SD	0.104	0.104	0.305	0.046	0.049	0.133
	RMSE	0.105	0.118	0.349	0.046	0.050	0.137
Pseudo	Bias	-0.012	-0.037	0.089	-0.002	-0.007	0.014
	SD	0.104	0.110	0.326	0.047	0.047	0.131
	RMSE	0.105	0.116	0.338	0.047	0.047	0.132
Whittle	Bias	-0.004	-0.043	0.102	-0.001	-0.009	0.022
	SD	0.103	0.109	0.409	0.047	0.049	0.180
	RMSE	0.103	0.117	0.422	0.047	0.050	0.181

Table 4.13: Estimated bias, SD, and RMSE when estimating the parameters of a Po-INAR(2) process with  $\alpha_1 = 0.2$ ,  $\alpha_2 = 0.4$  and  $n = 500$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.001 for estimating  $\alpha_1$ , 0.001 for estimating  $\alpha_2$  and 0.024 for estimating  $\mu_\epsilon$ .

Method		$\mu_\epsilon = 5$			$\mu_\epsilon = 10$		
		$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\mu}_\epsilon$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\mu}_\epsilon$
CML	Bias	-0.002	-0.003	0.070	-0.001	-0.004	0.137
	SD	0.042	0.038	0.538	0.042	0.037	1.053
	RMSE	0.042	0.038	0.542	0.042	0.038	1.063
CLS	Bias	0.129	0.000	0.139	0.129	0.000	0.301
	SD	0.012	0.013	0.615	0.012	0.013	1.221
	RMSE	0.130	0.013	0.631	0.130	0.013	1.258
Y-W	Bias	-0.003	-0.010	0.152	-0.003	-0.009	0.278
	SD	0.042	0.042	0.619	0.042	0.041	1.218
	RMSE	0.042	0.043	0.637	0.042	0.042	1.249
Pseudo	Bias	-0.003	-0.003	0.077	-0.002	-0.003	0.138
	SD	0.042	0.038	0.550	0.042	0.038	1.070
	RMSE	0.043	0.038	0.555	0.042	0.038	1.079
Whittle	Bias	0.000	-0.007	0.049	0.000	-0.007	0.099
	SD	0.042	0.042	0.545	0.042	0.042	1.080
	RMSE	0.042	0.042	0.547	0.042	0.042	1.084
Saddle	Bias	0.000	-0.010	0.116	-0.001	-0.007	0.197
	SD	0.042	0.037	0.545	0.041	0.037	1.070
	RMSE	0.042	0.038	0.558	0.041	0.037	1.088

### Varying the Innovation Mean

We estimate the parameters of Po-INAR(2), Geom-INAR(2), and NB-INAR(2) processes when  $\boldsymbol{\alpha} = (0.2, 0.4)^T$  with  $n = 500$  when the innovation mean is  $\mu_\epsilon = 5$  and  $\mu_\epsilon = 10$ . For the Po-INAR(2) process we see that CLS has a low bias, SD, and RMSE when estimating  $\alpha_2$  but has the worst performance in terms of bias, SD, and RMSE for  $\alpha_1$  when compared to all the other estimation methods (Table 4.13). This indicates that the CLS estimator is not consistent in its performance for all the parameters. When  $\mu_\epsilon$  is larger we clearly see

Table 4.14: Estimated bias, SD, and RMSE when estimating the parameters of a NB-INAR(2) process with  $\alpha_1 = 0.2$ ,  $\alpha_2 = 0.4$ ,  $r = 1$  and  $n = 500$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.001 for estimating  $\alpha_1$ , 0.001 for estimating  $\alpha_2$ , 0.135 for estimating  $\mu_\epsilon$  and 0.012 for estimating  $r$ .

Method		$\mu_\epsilon = 5$				$\mu_\epsilon = 10$			
		$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\mu}_\epsilon$	$\hat{r}$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\mu}_\epsilon$	$\hat{r}$
CML	Bias	-0.001	-0.002	0.043	-0.004	0.000	-0.001	0.041	-0.001
	SD	0.030	0.028	0.414	0.182	0.025	0.024	0.709	0.154
	RMSE	0.030	0.028	0.416	0.182	0.025	0.024	0.710	0.154
CLS	Bias	-0.002	-0.008	0.133	-0.024	-0.003	-0.007	0.249	-0.021
	SD	0.042	0.042	0.660	0.275	0.042	0.042	1.288	0.265
	RMSE	0.042	0.043	0.674	0.276	0.042	0.043	1.312	0.265
Y-W	Bias	-0.003	-0.010	0.161	-0.030	-0.003	-0.009	0.303	-0.026
	SD	0.042	0.042	0.661	0.274	0.042	0.041	1.306	0.262
	RMSE	0.042	0.043	0.681	0.276	0.042	0.043	1.340	0.264
Pseudo	Bias	-0.003	-0.008	0.135	-0.017	-0.003	-0.008	0.253	-0.016
	SD	0.045	0.044	0.693	0.282	0.043	0.043	1.323	0.266
	RMSE	0.046	0.044	0.706	0.282	0.043	0.043	1.347	0.266
Whittle	Bias	0.000	-0.006	1.606	-0.264	0.000	-0.007	2.950	-0.247
	SD	0.042	0.043	2.109	0.586	0.042	0.042	4.438	0.453
	RMSE	0.042	0.043	2.651	0.643	0.042	0.043	5.328	0.516
Saddle	Bias	-0.034	-0.026	0.744	-0.297	-0.022	-0.010	0.819	-0.156
	SD	0.031	0.029	0.439	0.119	0.026	0.025	0.727	0.125
	RMSE	0.046	0.039	0.864	0.320	0.035	0.027	1.095	0.200

the merits of CML as it overall has the lowest RMSE. The Whittle and pseudo estimators are comparable for  $\mu_\epsilon$  estimates but not for  $\alpha_1$  and  $\alpha_2$ . For the Geom-INAR(2) process we too see that CML, Whittle and pseudo are comparable and have a low bias, SD, and RMSE (Table 4.15). For the NB-INAR(2) process CML outperforms all other estimation methods in terms of bias, SD, and RMSE (Table 4.14).

Table 4.15: Estimated bias, SD, and RMSE when estimating the parameters of a Geom-INAR(2) process with  $\alpha_1 = 0.2$ ,  $\alpha_2 = 0.4$  and  $n = 500$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.002 for estimating  $\alpha_1$ , 0.006 for estimating  $\alpha_2$ , and 0.104 for estimating  $\mu_\epsilon$ .

Method		$\mu_\epsilon = 5$			$\mu_\epsilon = 10$		
		$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\mu}_\epsilon$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\mu}_\epsilon$
CML	Bias	-0.004	-0.005	0.109	-0.003	-0.008	0.295
	SD	0.044	0.062	0.702	0.041	0.040	1.132
	RMSE	0.045	0.063	0.710	0.041	0.041	1.169
CLS	Bias	-0.013	0.016	-0.037	-0.061	0.190	-3.256
	SD	0.060	0.130	1.230	0.116	0.288	5.319
	RMSE	0.061	0.131	1.231	0.131	0.345	6.236
Y-W	Bias	-0.003	-0.009	0.156	-0.002	-0.010	0.293
	SD	0.042	0.043	0.620	0.042	0.042	1.204
	RMSE	0.042	0.044	0.639	0.042	0.043	1.239
Pseudo	Bias	-0.003	-0.009	0.138	-0.003	-0.008	0.274
	SD	0.043	0.042	0.588	0.042	0.040	1.151
	RMSE	0.043	0.043	0.604	0.042	0.040	1.183
Whittle	Bias	-0.001	-0.007	0.105	0.000	-0.007	0.214
	SD	0.043	0.043	0.812	0.042	0.042	1.605
	RMSE	0.043	0.044	0.819	0.042	0.043	1.619

### 4.3 Estimating GINAR(4) process parameters

In this section we consider estimation of GINAR processes when the order of the process,  $p$ , is higher, with  $p = 4$ . We estimate the parameters of Po-INAR(4), Geom-INAR(4), and NB-INAR(4) processes. Note that we no longer consider the saddlepoint estimator due to its poor performance in previous simulations. We only include methods for which we can estimate all parameters without imposing additional constraints. Thus, we do not include results for the Whittle likelihood estimator for NB-INAR(4) simulations as we had to constrain the bounds of the  $r$  parameter (like NB-INAR(2) simulations) which is not ideal.

#### Varying the Sample Size

We estimate the parameters of Po-INAR(4), Geom-INAR(4), and NB-INAR(4) processes when  $\boldsymbol{\alpha} = (0.1, 0.2, 0.1, 0.3)^T$  with  $\mu_\epsilon = 1$  when the sample length is  $n = 100$  and  $n = 500$ . The saddlepoint method is not available in this case.

Table 4.16 and 4.20 show the results for the Geom-INAR(4) simulations, Table 4.17 and 4.19 show the results for the Po-INAR(4) simulations, and Tables 4.18 and 4.21 show the results for the NB-INAR(4) simulations. As expected, increasing sample size leads to lower RMSE, SD, and a smaller bias. Across both processes, the CML method tends to perform the best with respect to bias, SD, and RMSE. For the Po-INAR(4) process the pseudo likelihood and Whittle methods do comparably well as compared to the CML method, whereas for the Geom-INAR(4) process the Yule-Walker and pseudo methods are comparable to the CML method.

The CLS method for NB-INAR(4) process had some issues in terms of estimating the  $r$  parameter. In particular for the  $n = 100$  simulation, there were 11 simulations with estimated values of  $r$  over 100. These were discarded for the results presented however it is

Table 4.16: Estimated bias, SD, and RMSE when estimating the parameters of a Geom-INAR(4) process with  $\alpha_1 = 0.1$ ,  $\alpha_2 = 0.2$ ,  $\alpha_3 = 0.3$ ,  $\alpha_4 = 0.1$ , and  $\mu_\epsilon = 1$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.002 for estimating  $\alpha_1$ , 0.002 for estimating  $\alpha_2$ , 0.002 for estimating  $\alpha_3$ , 0.003 for estimating  $\alpha_4$  and 0.014 for estimating  $\mu_\epsilon$ .

Method		$n = 100$				
		$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$	$\hat{\mu}_\epsilon$
CML	Bias	-0.008	-0.032	-0.010	-0.049	0.310
	SD	0.082	0.096	0.078	0.099	0.482
	RMSE	0.082	0.102	0.079	0.110	0.573
CLS	Bias	-0.009	-0.032	-0.009	-0.043	0.286
	SD	0.087	0.110	0.083	0.115	0.500
	RMSE	0.087	0.115	0.084	0.123	0.576
Y-W	Bias	-0.016	-0.035	-0.023	-0.059	0.416
	SD	0.105	0.106	0.098	0.101	0.546
	RMSE	0.106	0.112	0.101	0.117	0.686
Pseudo	Bias	-0.008	-0.030	-0.010	-0.044	0.261
	SD	0.085	0.102	0.083	0.106	0.522
	RMSE	0.086	0.106	0.084	0.115	0.583
Whittle	Bias	0.011	-0.018	0.003	-0.044	0.128
	SD	0.092	0.104	0.087	0.105	0.584
	RMSE	0.092	0.106	0.087	0.114	0.598

safe to assume that this method should be avoided when we want to estimate the  $r$  parameter as well. We also note that most methods have issues estimating  $r$ , with a larger absolute bias, larger SD, and larger RMSE.

Table 4.17: Estimated bias, SD, and RMSE when estimating the parameters of a Po-INAR(4) process with  $\alpha_1 = 0.1$ ,  $\alpha_2 = 0.2$ ,  $\alpha_3 = 0.3$ ,  $\alpha_4 = 0.1$ , and  $\mu_\epsilon = 1$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.002 for estimating  $\alpha_1$ , 0.003 for estimating  $\alpha_2$ , 0.002 for estimating  $\alpha_3$ , 0.003 for estimating  $\alpha_4$  and 0.014 for estimating  $\mu_\epsilon$ .

Method		$n = 100$				
		$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$	$\hat{\mu}_\epsilon$
CML	Bias	0.001	-0.017	0.003	-0.024	0.118
	SD	0.089	0.104	0.089	0.104	0.493
	RMSE	0.089	0.105	0.089	0.106	0.507
CLS	Bias	-0.009	-0.032	-0.009	-0.038	0.277
	SD	0.085	0.107	0.084	0.114	0.506
	RMSE	0.085	0.111	0.084	0.120	0.576
Y-W	Bias	-0.012	-0.033	-0.024	-0.058	0.405
	SD	0.104	0.104	0.097	0.097	0.545
	RMSE	0.105	0.109	0.100	0.114	0.679
Pseudo	Bias	-0.002	-0.025	-0.003	-0.031	0.175
	SD	0.091	0.107	0.088	0.106	0.448
	RMSE	0.091	0.109	0.088	0.111	0.481
Whittle	Bias	0.010	-0.016	0.003	-0.043	0.099
	SD	0.090	0.102	0.088	0.102	0.459
	RMSE	0.091	0.103	0.088	0.111	0.469

Table 4.18: Estimated bias, SD, and RMSE when estimating the parameters of a NB-INAR(4) process with  $\alpha_1 = 0.1$ ,  $\alpha_2 = 0.2$ ,  $\alpha_3 = 0.3$ ,  $\alpha_4 = 0.1$ , and  $\mu_\epsilon = 1$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.002 for estimating  $\alpha_1$ , 0.002 for estimating  $\alpha_2$ , 0.002 for estimating  $\alpha_3$ , 0.003 for estimating  $\alpha_4$  and 0.014 for estimating  $\mu_\epsilon$ , and 0.681 for estimating  $r$ .

Method		$n = 100$					
		$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$	$\hat{\mu}_\epsilon$	$\hat{r}$
CML	Bias	-0.005	-0.026	-0.007	-0.034	0.230	-0.017
	SD	0.084	0.098	0.080	0.096	0.473	1.710
	RMSE	0.084	0.101	0.080	0.102	0.526	1.710
CLS	Bias	-0.009	-0.030	-0.010	-0.038	0.272	-0.142
	SD	0.085	0.106	0.083	0.114	0.521	1.964
	RMSE	0.085	0.110	0.084	0.120	0.588	1.964
Y-W	Bias	-0.013	-0.034	-0.022	-0.057	0.401	-0.292
	SD	0.102	0.104	0.098	0.098	0.557	0.992
	RMSE	0.103	0.109	0.100	0.113	0.687	0.992
Pseudo	Bias	-0.004	-0.031	-0.008	-0.045	0.280	-0.175
	SD	0.091	0.109	0.086	0.109	0.534	1.690
	RMSE	0.091	0.113	0.086	0.118	0.603	1.690

Table 4.19: Estimated bias, SD, and RMSE when estimating the parameters of a Po-INAR(4) process with  $\alpha_1 = 0.1$ ,  $\alpha_2 = 0.2$ ,  $\alpha_3 = 0.3$ ,  $\alpha_4 = 0.1$ , and  $\mu_\epsilon = 1$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.001 for estimating  $\alpha_1$ , 0.001 for estimating  $\alpha_2$ , 0.001 for estimating  $\alpha_3$ , 0.001 for estimating  $\alpha_4$  and 0.004 for estimating  $\mu_\epsilon$ .

Method		$n = 500$				
		$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$	$\hat{\mu}_\epsilon$
CML	Bias	-0.003	-0.004	-0.003	-0.005	0.047
	SD	0.043	0.044	0.043	0.042	0.181
	RMSE	0.043	0.044	0.043	0.043	0.187
CLS	Bias	-0.003	-0.006	-0.004	-0.008	0.066
	SD	0.045	0.047	0.044	0.046	0.216
	RMSE	0.045	0.048	0.044	0.046	0.225
Y-W	Bias	-0.003	-0.006	-0.004	-0.011	0.078
	SD	0.045	0.045	0.044	0.045	0.205
	RMSE	0.045	0.046	0.044	0.046	0.219
Pseudo	Bias	-0.002	-0.005	-0.003	-0.005	0.042
	SD	0.045	0.045	0.045	0.044	0.188
	RMSE	0.045	0.046	0.045	0.044	0.192
Whittle	Bias	0.000	-0.004	-0.001	-0.009	0.034
	SD	0.044	0.045	0.044	0.045	0.188
	RMSE	0.044	0.045	0.044	0.046	0.191

Table 4.20: Estimated bias, SD, and RMSE when estimating the parameters of a Geom-INAR(4) process with  $\alpha_1 = 0.1$ ,  $\alpha_2 = 0.2$ ,  $\alpha_3 = 0.3$ ,  $\alpha_4 = 0.1$ , and  $\mu_\epsilon = 1$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.001 for estimating  $\alpha_1$ , 0.001 for estimating  $\alpha_2$ , 0.001 for estimating  $\alpha_3$ , 0.001 for estimating  $\alpha_4$  and 0.004 for estimating  $\mu_\epsilon$ .

Method		$n = 500$				
		$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$	$\hat{\mu}_\epsilon$
CML	Bias	-0.003	-0.007	-0.002	-0.008	0.066
	SD	0.043	0.044	0.042	0.044	0.189
	RMSE	0.043	0.045	0.042	0.045	0.200
CLS	Bias	-0.003	-0.007	-0.003	-0.009	0.067
	SD	0.045	0.048	0.045	0.048	0.213
	RMSE	0.045	0.049	0.045	0.049	0.223
Y-W	Bias	-0.001	-0.007	-0.005	-0.014	0.081
	SD	0.046	0.047	0.045	0.047	0.207
	RMSE	0.046	0.047	0.045	0.049	0.222
Pseudo	Bias	-0.002	-0.006	-0.004	-0.008	0.054
	SD	0.046	0.048	0.046	0.047	0.209
	RMSE	0.046	0.048	0.046	0.048	0.216
Whittle	Bias	0.001	-0.005	-0.001	-0.010	0.039
	SD	0.045	0.048	0.045	0.048	0.236
	RMSE	0.045	0.048	0.045	0.049	0.239

Table 4.21: Estimated bias, SD, and RMSE when estimating the parameters of a NB-INAR(4) process with  $\alpha_1 = 0.1$ ,  $\alpha_2 = 0.2$ ,  $\alpha_3 = 0.3$ ,  $\alpha_4 = 0.1$ , and  $\mu_\epsilon = 1$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.001 for estimating  $\alpha_1$ , 0.001 for estimating  $\alpha_2$ , 0.001 for estimating  $\alpha_3$ , 0.001 for estimating  $\alpha_4$ , 0.005 for estimating  $\mu_\epsilon$  and 0.015 for estimating  $r$ .

Method		$n = 500$					
		$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$	$\hat{\mu}_\epsilon$	$\hat{r}$
CML	Bias	-0.002	-0.004	-0.002	-0.006	0.047	-0.038
	SD	0.041	0.042	0.041	0.040	0.181	0.416
	RMSE	0.041	0.042	0.041	0.041	0.187	0.416
CLS	Bias	-0.003	-0.006	-0.003	-0.009	0.066	-0.050
	SD	0.044	0.047	0.044	0.047	0.215	0.535
	RMSE	0.044	0.047	0.044	0.048	0.225	0.535
Y-W	Bias	-0.002	-0.006	-0.005	-0.012	0.079	-0.073
	SD	0.045	0.046	0.045	0.046	0.208	0.503
	RMSE	0.045	0.046	0.045	0.047	0.222	0.503
Pseudo	Bias	-0.004	-0.007	-0.003	-0.009	0.071	-0.065
	SD	0.049	0.050	0.049	0.048	0.229	0.501
	RMSE	0.049	0.050	0.049	0.049	0.239	0.501

Table 4.22: Estimated bias, SD, and RMSE when estimating the parameters of a Po-INAR(4) process with  $\alpha_1 = 0.1$ ,  $\alpha_2 = 0.2$ ,  $\alpha_3 = 0.3$  and  $\alpha_4 = 0.1$  and  $\mu_\epsilon = 10$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.002 for estimating  $\alpha_1$ , 0.003 for estimating  $\alpha_2$ , 0.001 for estimating  $\alpha_3$ , 0.006 for estimating  $\alpha_4$ , and 0.102 for estimating  $\mu_\epsilon$ .

Method		$n = 500$				
		$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$	$\hat{\mu}_\epsilon$
CML	Bias	0.002	-0.004	-0.001	-0.008	0.287
	SD	0.043	0.043	0.042	0.044	1.830
	RMSE	-0.004	0.043	0.042	0.044	1.853
CLS	Bias	-0.022	0.006	-0.023	0.082	-1.440
	SD	0.062	0.121	0.061	0.217	5.136
	RMSE	0.006	0.122	0.066	0.232	5.334
YW	Bias	-0.001	-0.008	-0.005	-0.010	0.774
	SD	0.043	0.043	0.044	0.042	1.963
	RMSE	-0.008	0.044	0.044	0.044	2.110
Pseduo	Bias	-0.003	-0.006	-0.003	-0.005	0.526
	SD	0.044	0.043	0.043	0.042	1.849
	RMSE	-0.006	0.044	0.043	0.042	1.923
Whittle	Bias	0.002	-0.004	-0.001	-0.008	0.287
	SD	0.043	0.043	0.042	0.044	1.830
	RMSE	-0.004	0.043	0.042	0.044	1.853

### Varying the Innovation Mean

We estimate the parameters of Po-INAR(4), NB-INAR(4), and Geom-INAR(4) processes when

$\boldsymbol{\alpha} = (0.1, 0.2, 0.1, 0.3)^T$  with  $\mu_\epsilon = 1$  and  $\mu_\epsilon = 10$  when the sample length is  $n = 500$ .

We see that for all methods, estimating the thinning probability parameters  $\boldsymbol{\alpha}$  tends to be easier than estimating the innovation mean  $\mu_\epsilon$  and this is especially true as the innovation mean increases (Tables 4.22, 4.23 and 4.24). Overall the results are consistent with the

Table 4.23: Estimated bias, SD, and RMSE when estimating the parameters of a Geom-INAR(4) process with  $\alpha_1 = 0.1$ ,  $\alpha_2 = 0.2$ ,  $\alpha_3 = 0.3$  and  $\alpha_4 = 0.1$  and  $\mu_\epsilon = 10$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.001 for estimating  $\alpha_1$ , 0.003 for estimating  $\alpha_2$ , 0.001 for estimating  $\alpha_3$ , 0.006 for estimating  $\alpha_4$ , and 0.102 for estimating  $\mu_\epsilon$ .

Method		$n = 500$				
		$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$	$\hat{\mu}_\epsilon$
CML	Bias	-0.002	-0.007	-0.003	-0.009	0.700
	SD	0.043	0.043	0.042	0.042	1.928
	RMSE	0.043	0.044	0.042	0.043	2.051
CLS	Bias	-0.022	0.006	-0.023	0.082	-1.440
	SD	0.062	0.121	0.061	0.217	5.136
	RMSE	0.066	0.122	0.066	0.232	5.334
YW	Bias	-0.002	-0.006	-0.004	-0.011	0.727
	SD	0.044	0.044	0.043	0.043	1.978
	RMSE	0.044	0.044	0.043	0.045	2.108
Pseudo	Bias	-0.002	-0.007	-0.003	-0.010	0.687
	SD	0.043	0.044	0.043	0.042	1.944
	RMSE	0.043	0.044	0.043	0.043	2.062
Whittle	Bias	0.001	-0.004	-0.001	-0.008	0.365
	SD	0.044	0.043	0.043	0.044	2.247
	RMSE	0.044	0.044	0.043	0.044	2.277

Table 4.24: Estimated bias, SD, and RMSE when estimating the parameters of a NB-INAR(4) process with  $\alpha_1 = 0.1$ ,  $\alpha_2 = 0.2$ ,  $\alpha_3 = 0.3$  and  $\alpha_4 = 0.1$  and  $\mu_\epsilon = 10$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.001 for estimating  $\alpha_1$ , 0.001 for estimating  $\alpha_2$ , 0.001 for estimating  $\alpha_3$ , 0.001 for estimating  $\alpha_4$ , 0.042 for estimating  $\mu_\epsilon$  and 0.009 for estimation  $r$ .

Method		$n = 500$					
		$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$	$\hat{\mu}_\epsilon$	$\hat{r}$
CML	Bias	0.001	-0.001	0.000	-0.002	0.067	0.005
	SD	0.030	0.030	0.030	0.030	0.944	0.209
	RMSE	0.030	0.030	0.030	0.030	0.947	0.209
Y-W	Bias	-0.001	-0.006	-0.004	-0.011	0.767	-0.054
	SD	0.044	0.043	0.043	0.043	2.022	0.390
	RMSE	0.044	0.044	0.043	0.044	2.163	0.390
Pseudo	Bias	-0.003	-0.007	-0.003	-0.008	0.682	-0.040
	SD	0.044	0.045	0.045	0.044	2.092	0.396
	RMSE	0.045	0.045	0.045	0.045	2.201	0.396

previous section, where the CML method tends to perform the best with respect to bias, SD, and RMSE.

Note that for the NB-INAR(4) process, the CLS method starts failing in almost 200 cases out of 10,000 when  $\mu_\epsilon = 10$ , and hence the results are not shown. The results are not comparable to the CML, Yule-Walker or pseudo estimation methods.

## 4.4 Coverage Simulations

In this section we provide a simple coverage simulation for the GINAR(1) processes studied in Section 4.1 for varying sample sizes. We calculate confidence intervals for the parameters in R using the Hessian calculated as part of the optimization used to calculate each parameter estimate.

Table 4.25: A comparison of the estimated coverages for a 95% confidence interval for  $\mu_\epsilon$  and  $\alpha_1$  in the Po-INAR(1) model, as  $n$  is varied, for select estimation methods. The number of replicates is 10,000. The maximum standard error for  $\mu_\epsilon$  and  $\alpha_1$  is 0.43 and 0.45 respectively. The bold values indicate coverage probabilities no different than 0.95, on the basis of a significance test with level 0.05.

Method	$n = 500$		$n = 1,000$	
	$\mu_\epsilon$	$\alpha_1$	$\mu_\epsilon$	$\alpha_1$
CML	<b>94.6%</b>	<b>95.0%</b>	<b>94.8%</b>	<b>94.6%</b>
Pseudo	87.7%	93.7%	88.5%	94.2%
Saddlepoint	88.1%	84.5%	77.7%	71.0%
Whittle	75.0%	80.2%	75.8%	79.6%

Table 4.26: A comparison of the estimated coverages for a 95% confidence interval for  $\mu_\epsilon$  and  $\alpha_1$  in the NB-INAR(1) model, as  $n$  is varied, for select estimation methods. The number of replicates is 10,000. The maximum standard error for  $\mu_\epsilon$  and  $\alpha_1$  is 0.43 and 0.45 respectively. The bold values indicate coverage probabilities no different than 0.95, on the basis of a significance test with level 0.05.

Method	$n = 500$			$n = 1,000$		
	$\mu_\epsilon$	$\alpha_1$	$r$	$\mu_\epsilon$	$\alpha_1$	$r$
CML	<b>94.9%</b>	<b>95.2%</b>	<b>95.4%</b>	<b>95.0%</b>	<b>95.0%</b>	<b>95.1%</b>
Pseudo	94.2%	<b>95.1%</b>	98.5%	<b>95.0%</b>	<b>95.1%</b>	98.5%
Saddlepoint	64.0%	52.2%	15.5%	41.5%	22.1%	1.8%
Whittle	100.0%	79.7%	100.0%	100.0%	78.9%	100.0%

Table 4.27: A comparison of the estimated coverages for a 95% confidence interval for  $\mu_\epsilon$  and  $\alpha_1$  in the Geom-INAR(1) model, as  $n$  is varied, for select estimation methods. The number of replicates is 10,000. The maximum standard error for  $\mu_\epsilon$  and  $\alpha_1$  is 0.43 and 0.45 respectively. The bold values indicate coverage probabilities no different than 0.95, on the basis of a significance test with level 0.05.

Method	$n = 500$		$n = 1,000$	
	$\mu_\epsilon$	$\alpha_1$	$\mu_\epsilon$	$\alpha_1$
CML	<b>94.9%</b>	<b>95.4%</b>	<b>94.6%</b>	<b>94.8%</b>
Pseudo	85.9%	85.8%	86.2%	85.8%
Whittle	74.8%	71.5%	75.1%	71.3%

We use the Hessian since calculating the asymptotic variance for each estimator from the asymptotic theory provided in Section 3 is not straightforward for larger model orders. Tables 4.25, 4.26, and 4.27 give the coverage of Po-INAR(1), NB-INAR(1), and Geom-INAR(1) process parameters for the given estimation methods when  $n = 500$  and  $n = 1,000$  respectively. First, 95% confidence intervals were built for the parameter estimates, and the coverage was calculated as the proportion of these confidence intervals that had the true parameter value. The true values for all simulations are  $\mu_\epsilon = 1$ ,  $\alpha_1 = 0.5$ , and  $r = 1$ . Furthermore, in Tables 4.25, 4.26, and 4.27 we test whether or not the population coverages are equal to 95% for each method and sample size. In particular, we carry out a two-sided test for a proportion that each true coverage is equal to 95% or not. For an 0.05 level test, we fail to reject the test when the observed proportions lie between 94.6% and 95.4%, and denote such cases in Tables 4.25, 4.26, and 4.27 in bold.

The CML method has the best coverage performance for all parameters for all processes considered since the coverage probabilities are no different than 95%. In Table 4.25 we see that the Whittle estimator has poor coverage for both parameters, even though it had good bias, SD, and RMSE performance in Section 4.1. For the NB-INAR(1) process coverage probabilities presented in Table 4.26, we see again that only the CML method has appropriate performance with coverage no different than 95%, although pseudo does depict similar performance for the  $\mu_\epsilon$  and  $\alpha_1$  parameters when  $n = 1,000$ .

Similarly, for the Geom-INAR(1) process in Table 4.27 we see that the CML estimation method has the best performance compared to the other methods, with coverage values no different than 95%.

## 4.5 Conclusions

In this chapter we presented various simulations demonstrating the small sample performance of different estimation methods for  $\text{GINAR}(p)$  processes. We looked at metrics like bias, SD, RMSE, and coverage probabilities. We studied the performance of these methods for two different thinning operators and innovation distributions, three model orders, and varying parameter and sample size values.

The main takeaway from these simulations is that CML performs well over all other methods, even when changing the model order  $p$ , the values of the parameters, thinning operation, and innovation distribution. Our simulations over different  $\text{GINAR}$  processes show that some methods are competitive to conditional maximum likelihood (CML) for some processes (e.g. pseudo likelihood estimator for  $\text{Geom-INAR}$  processes), however the overall best choice is CML, especially for longer sample sizes. CML also tends to do better compared to other methods for overdispersed processes and when the processes is close to being nonstationary (i.e., when the sum of the  $\alpha_j$  dependence parameters is close to one). Therefore, for applications where we have an overdispersed or highly dependent series we should be using CML to estimate the parameters. This is further supported by the coverage simulations in Section 4.4 where the CML method consistently outperformed other methods for all three processes as it had coverage probabilities no different than 95%. The only other method with comparable performance was the pseudo likelihood estimation method, but this was only true for the  $\text{NB-INAR}(1)$  process with  $n = 1,000$ .

When calculating the transition probabilities the Davies method using Gauss quadrature yields the same results as the convolution calculation and is computationally more efficient. Hence, we recommend using the Davies method for CML estimation of  $\text{GINAR}(p)$  process parameters. It is important to note that even with this faster implementation, CML can be

computationally costly as compared to other estimation methods especially when the model order  $p$  is large. In cases where computational complexity is an issue, methods like pseudo maximum likelihood can be used for certain processes like GINAR(1) processes with binomial thinning and Poisson innovations, as their performance is comparable to CML. However, caution should be taken when varying the thinning operator and innovation process. Varying the type of GINAR process, the Yule-Walker method used extensively in the literature may not be an ideal choice for estimation of model parameters. Similarly the performance of other (non-CML) estimators are more mixed. In particular, saddlepoint methods have issues, and thus are less effective for estimating GINAR process parameters.

## Chapter 5: Time Varying GINAR( $p$ ) Processes

### 5.1 Introduction

The stationarity assumption has been an important part of time series analysis as there exists a large variety of models and methods for stationary processes, and important mathematical theorems like ergodic theorem and various central limit theorems rely on the assumption of stationarity. Up until now in this dissertation we have assumed stationarity of GINAR( $p$ ) processes. Although stationarity offers us convenient and well-defined mathematical properties, it is not always applicable or possible in the real-world setting. Hence, a lot of attention has been focused on non-stationary time series. The biggest challenge is deriving meaningful asymptotic properties when we no longer have the stationarity assumption. In the continuous time series domain, there exists well defined theory surrounding locally stationary processes (e.g. [Dahlhaus \[2012\]](#)), but this is not available for count-time series processes like the GINAR( $p$ ) process and the theory does not extend naturally to this class of processes.

In an attempt to capture non-stationarity in count-time series researchers in the field have studied variations of GINAR( $p$ ) processes. For instance, [Kim and Park \[2008\]](#) introduce a INAR( $p$ ) model with signed binomial thinning to account for non-stationarity in time series with large dispersion. Their process allows for negative autocorrelations, whereas

the GINAR( $p$ ) process we have defined thus far can only work with non-negative counts. They show the application and merit of their model to an AIDS disease dataset which depicts large variability. [Nastić et al. \[2016\]](#) introduce a random environment integer-valued autoregressive process where the marginal distribution of the parameters can take different values that correspond to finite possible states of the environment.

Introducing structural breaks in a process is another way to capture non-stationary. For example, [Kashikar et al. \[2013\]](#) introduce a non-stationary INAR process with structural breaks to capture situations where the parameters of the process are not constant over time. They define the process with  $m$  break points and a different INAR process is used to define these  $m + 1$  sections; their process is essentially a state-space model. Their methodology for the selection of the number of breakpoints involves comparing multiple models with different values of  $m$  based on model selection criteria like AIC, BIC, and forecast error metrics. They apply their new process to a rare disease application of H1N1 cases in India and show that their non-stationary process outperforms stationary processes.

Researchers like [Joe \[1996\]](#) and [Freeland and McCabe \[2004a\]](#) have briefly discussed how to allow for time-varying parameters for the GINAR processes with binomial thinning and Poisson innovations. [Brannas \[1995\]](#) and [Enciso-Mora et al. \[2009\]](#) incorporated explanatory variables into the dependence and innovation parameters of the GINAR(1) process with binomial thinning and Poisson innovations. The majority of the research in this area has focused on modifying the GINAR( $p$ ) process to allow for a time-varying innovation parameter, for example, [Böckenholt \[1998\]](#) introduces a mixture version of integer-valued autoregressive Poisson regression models where the rate parameter is expressed as a function of non-negative covariates. [Roy and Karmakar \[2021\]](#) propose a semiparametric time-varying autoregressive process and propose a Bayesian framework to study it. They introduce the

time-varying Bayesian integer valued generalized autoregressive conditional heteroscedastic (TVBINGARCH) model, where the mean of the process varies with time. They also consider a simplified version of this model called the time-varying Bayesian autoregressive model for counts (TVBARC). They study the spread of COVID-19 in NYC for a specific time period with the aim to determine which lags are significant so that it can be used to determine the incubation period for symptoms to show up. [Pedeli et al. \[2015\]](#) use a time-varying innovation mean that to account for the seasonality in meningococcal disease cases in Germany. This dataset is explored further in Chapter 7.

There has not been much research carried out in allowing for a time-varying dependence parameter which is what we want to focus our attention on. We introduce a new GINAR( $p$ ) process, called the time-varying GINAR( $p$ ) process, for non-stationary count time series. In particular, we allow both the dependence and parameters characterizing the innovation process, such as innovation mean and innovation variance, to vary with time. The time-varying innovation and dependence parameters are parameterized by a transformation of basis functions. These smooth functions of the process parameters are able to capture changes in the time-varying parameters. We discuss parameter estimation for this process using conditional maximum likelihood, pseudo maximum likelihood, and conditional least squares, and investigate the construction of confidence regions for parameters. We further present simulations and show the application of our model to a disease dataset and a patient scores dataset.

## 5.2 Defining TV-GINAR( $p$ ) Processes

We define the time-varying GINAR( $p$ ) (TV-GINAR( $p$ )) process using the thinning operator defined in Definition 2.1 and following the ideas from Dahlhaus [2012] on locally stationary processes. We define these processes using triangular arrays  $\{(t, T) : t = 1, 2, \dots, T\}$  for positive integers  $T = 1, 2, \dots$ . Our parameter curves are then modeled in local time  $u = t/T$ ,  $t = 1, 2, \dots, T$  for a given  $T$ . As  $T$  increases we are able to learn more about the parameter curves over  $[0, 1]$ .

**Definition 5.2.1** *The TV-GINAR( $p$ ) process  $\{X_{t,T} : t = 1, 2, \dots, T\}$  for a positive integer  $T$ , is the non-negative integer-valued process defined by*

$$X_{t,T} = \sum_{j=1}^p \alpha_{j,t,T} \odot X_{t-j,T} + \xi_{t,T}. \quad (5.1)$$

Let  $u = t/T$  be the rescaled time unit. In the definition above, the dependence parameter  $\alpha_{j,t,T}$  can be represented using the parameter curve  $\{\alpha_j(u) : u \in [0, 1]\}$ , where  $j = 1, \dots, p$  and  $\sum_{j=1}^p \alpha_j(u) < 1$  for all  $u \in [0, 1]$ . Also  $0 < \alpha_j(u) < 1$  for all  $u \in [0, 1]$  and  $j = 1, 2, \dots, p$ .

The innovation process  $\{\xi_{t,T}\}$  is a set of IID non-negative integer-valued RVs with mean  $\mu_\xi(u) > 0$  and variance  $\sigma_\xi^2(u) > 0$ , which again depends on the local time unit  $u \in [0, 1]$ . Also, the  $\{Y_{k,t}\}$  associated with each thinning operation (see (2.1)) are mutually independent and independent of  $\{\xi_{t,T}\}$ , and  $\{\xi_{t,T}\}$  is independent of  $X_{t-j}$  for all  $t \in T$  and  $j \geq 1$ .

For local stationarity we assume that there exists a positive constant  $K$  and dependence parameter  $\alpha_j(u)$  such that,

$$\sup_t |\alpha_{j,t,T} - \alpha_j(u)| \leq K/T, \quad (5.2)$$

for all  $T$ . The functions  $\alpha_j(u)$  ( $j = 1, 2, \dots, p$ ),  $\mu_\epsilon(u)$ , and  $\sigma_\epsilon^2(u)$  are assumed to be continuous over  $u \in [0, 1]$ .

Note that in the above definition, (5.2) gives us a locally stationary type representation (e.g. Dahlhaus [2012]) for an integer time series. However, rather than using linear filtering this process is defined through thinning operations.

Each dependence parameter curve  $\{\alpha_j(u) : u \in [0, 1]\}$  is defined via generalized linear functions of basis vectors as follows:

$$\alpha_j(u) = \frac{\exp(\mathbf{v}_j^T(u)\boldsymbol{\eta}_j)}{1 + \sum_{k=1}^p \exp(\mathbf{v}_k^T(u)\boldsymbol{\eta}_k)}, \quad u \in [0, 1],$$

where  $\{\mathbf{v}_j(u)\}$  is a vector of  $m_j$  smooth basis functions and  $\boldsymbol{\eta}_j$  is a  $m_j$  dimensional vector of model coefficients for  $j = 1, 2, \dots, p$ . Note that this transformation was chosen to ensure that the constraints on parameter curves (i.e.  $\sum_{j=1}^p \alpha_j(u) < 1$  for all  $u \in [0, 1]$  with  $0 < \alpha_j(u) < 1$  for all  $u$ ), as described in Definition 5.2.1, are satisfied.

Similarly, the innovation mean parameter curves  $\{\mu_\xi(u) : u \in [0, 1]\}$  and the innovation variance parameter curve  $\{\sigma_\xi^2(u) : u \in [0, 1]\}$  are defined via generalized linear functions of basis vectors as follows:

$$\mu_\xi(u) = \exp(\mathbf{v}_{p+1}^T(u)\boldsymbol{\eta}_{p+1}), \quad u \in [0, 1],$$

and

$$\sigma_\xi^2(u) = \exp(\mathbf{v}_{p+2}^T(u)\boldsymbol{\eta}_{p+2}), \quad u \in [0, 1],$$

where  $\{\mathbf{v}_{p+1}(u)\}$  is a vector of  $m_{p+1}$  smooth basis functions and  $\boldsymbol{\eta}_{p+1}$  is a  $m_{p+1}$  dimensional vector of model coefficients. Lastly,  $\{\mathbf{v}_{p+2}(u)\}$  is a vector of  $m_{p+2}$  smooth basis functions and  $\boldsymbol{\eta}_{p+2}$  is a  $m_{p+2}$  dimensional vector of model coefficients.

Note that we choose basis functions to model the time-varying parameter curves due to the flexibility in choice of basis functions, for example b-splines, polynomials, and Fourier

basis functions. This allows flexibility in modeling a variety of different time-varying processes. We mainly focus on b-spline basis functions in this dissertation (e.g. [Piegl and Tiller \[1995\]](#)).

### 5.3 Statistical properties of TV-GINAR( $p$ ) processes

In this section we present useful statistical properties of both the generalized thinning operator and the TV-GINAR( $p$ ) process.

**Lemma 5.3.1** *Let  $\{X_{t,T} : t = 1, 2, \dots, T\}$  for a positive integer  $T$  be a TV-GINAR( $p$ ) process of Definition 5.2.1. Then for each  $t = 1, 2, \dots, T$ , the conditional mean is*

$$\mu_{X_{t,T}|X_{t-1,T}, \dots, X_{t-p,T}} = E(X_{t,T}|X_{t-1,T}, \dots, X_{t-p,T}) = \sum_{j=1}^p \alpha_{j,t,T} X_{t-j} + \mu_{\xi}(t/T),$$

and the conditional variance is

$$\sigma_{X_{t,T}|X_{t-1,T}, \dots, X_{t-p,T}}^2 = \text{var}(X_{t,T}|X_{t-1,T}, \dots, X_{t-p,T}) = \sum_{j=1}^p \beta_{j,t,T} + \sigma_{\xi}^2(t/T).$$

Further, the marginal mean is

$$E(X_{t,T}) = \sum_{j=1}^p \alpha_{j,t,T} E(X_{t-j,T}) + \mu_{\xi}(t/T),$$

and the marginal variance is

$$\text{var}(X_{t,T}) = \sum_{j=1}^p \alpha_{j,t,T}^2 \text{var}(X_{t-j,T}) + \sum_{j=1}^p \beta_{j,t,T} E(X_{t-j,T}) + \sigma_{\xi}^2(t/T).$$

The autocovariance sequence  $\{\gamma_X(t, k) : k \in \mathbb{Z}, t = 1, 2, \dots, T\}$  satisfies the following for all lags  $k \neq 0$ :

$$\gamma_X(t, k) = \sum_{j=1}^p \alpha_{j,t,T} \gamma_X(t-j, k-j).$$

Similarly, the autocorrelation sequence satisfies for all lags  $k \neq 0$ :

$$\rho_X(t, k) = \frac{\gamma_X(t, k)}{\sqrt{\text{var}(X_{t,T})\text{var}(X_{t-k,T})}}.$$

Proof: The conditional expectation and variance property can be derived as follows:

$$\begin{aligned}\mu_{X_{t,T}|X_{t-1,T},\dots,X_{t-p,T}} &= \sum_{j=1}^p E(\alpha_{j,t,T} \odot X_{t-j,T} | X_{t-j,T}) + E(\xi_{t,T}) \\ &= \sum_{j=1}^p \alpha_{j,t,T} X_{t-j,T} + \mu_{\xi}(t/T),\end{aligned}$$

with

$$\begin{aligned}\sigma_{X_{t,T}|X_{t-1,T},\dots,X_{t-p,T}}^2 &= \sum_{j=1}^p \text{var}(\alpha_{j,t,T} \odot X_{t-j,T} | X_{t-j,T}) + \text{var}(\xi_{t,T}) \\ &= \sum_{j=1}^p \beta_{j,t,T} + \sigma_{\xi}^2(t/T).\end{aligned}$$

The marginal mean and variance is obtained using law of iterated expectations as follows:

$$\begin{aligned}E(X_{t,T}) &= \sum_{j=1}^p \alpha_{j,t,T} E(X_{t-j,T}) + \mu_{\xi}(t/T), \text{ and} \\ \text{var}(X_{t,T}) &= \sum_{j=1}^p \alpha_{j,t,T}^2 \text{var}(X_{t-j,T}) + \sum_{j=1}^p \beta_{j,t,T} E(X_{t-j,T}) + \sigma_{\xi}^2(t/T).\end{aligned}$$

Next, we present a proof for the autocorrelation and autocovariance properties of the process. To derive the autocovariance structure we present the TV-GINAR( $p$ ) model in matrix form. Let  $\mathbf{X}_{t,T} = (X_{t,T}, X_{t-1,T}, \dots, X_{t-p+1,T})^T$ , and let  $\mathbf{A}_{t,T}$  and  $\boldsymbol{\xi}_{t,T}$  be defined as

$$\mathbf{A}_{t,T} = \begin{bmatrix} \alpha_{1,t,T} & \alpha_{2,t,T} & \dots & \alpha_{p-1,t,T} & \alpha_{p,t,T} \\ 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix} \quad \boldsymbol{\xi}_{t,T} = \begin{bmatrix} \xi_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Then the TV-GINAR( $p$ ) process can be expressed as

$$\mathbf{X}_{t,T} = \mathbf{A}_{t,T} \odot \mathbf{X}_{t-1,T} + \boldsymbol{\xi}_{t,T}.$$

We define the multivariate autocovariance sequence  $\{\boldsymbol{\Gamma}(k, t) : k \in \mathbb{Z}, t = 1, 2, \dots, T\}$  by

$$\boldsymbol{\Gamma}(k, t) = E[(\mathbf{X}_{t,T} - E(\mathbf{X}_{t,T}))(\mathbf{X}_{t-k,T} - E(\mathbf{X}_{t-k,T}))^T], \quad k \in \mathbb{Z}.$$

Then

$$\begin{aligned}
\mathbf{\Gamma}(k, t) &= E(\mathbf{X}_{t,T} \mathbf{X}_{t-k,T}^T) - E(\mathbf{X}_{t,T}) E(\mathbf{X}_{t-k,T})^T \\
&= E((\mathbf{A}_{t,T} \odot \mathbf{X}_{t-1,T} + \boldsymbol{\xi}_{t,T}) \mathbf{X}_{t-k,T}^T) - E(\mathbf{X}_{t,T}) E(\mathbf{X}_{t-k,T})^T \\
&= \mathbf{A}_{t,T} E(\mathbf{X}_{t-1,T} \mathbf{X}_{t-k,T}^T) + E(\boldsymbol{\xi}_{t,T}) E(\mathbf{X}_{t-k,T})^T - E(\mathbf{X}_{t,T}) E(\mathbf{X}_{t-k,T})^T \\
&= \mathbf{A}_{t,T} E(\mathbf{X}_{t-1,T} \mathbf{X}_{t-k,T}^T) - \{E(\mathbf{X}_{t,T}) - E(\boldsymbol{\xi}_{t,T})\} E(\mathbf{X}_{t-k,T})^T \\
&= \mathbf{A}_{t,T} \{E(\mathbf{X}_{t-1,T} \mathbf{X}_{t-k,T}^T) - E(\mathbf{X}_{t-1,T}) E(\mathbf{X}_{t-k,T})^T\} \\
&= \mathbf{A}_{t,T} \mathbf{\Gamma}(k-1, t-1),
\end{aligned}$$

where we are using the property that  $E(\mathbf{X}_{t,T}) = \mathbf{A}_{t,T} E(\mathbf{X}_{t-1,T}) + E(\boldsymbol{\xi}_{t,T})$ . Then we can get the univariate autocovariance at lag  $k$  and time  $t$  as

$$\gamma_X(t, k) = \sum_{j=1}^p \alpha_{j,t,T} \gamma(t-j, k-j),$$

and consequently the autocorrelation at lag  $k$  and time  $t$  is

$$\rho_X(t, k) = \frac{\gamma(t, k)}{\sqrt{\text{var}(X_{t,T}) \text{var}(X_{t-k,T})}}.$$

As an example, the ACVF for the TV-GINAR(1) process with binomial thinning and Poisson innovations can be obtained recursively as follows.

$$\begin{aligned}
\gamma(t, k) &= \alpha_{1,t,T} \gamma_X(t-1, k-1) \\
&= \alpha_{1,t,T} (\alpha_{1,t-1,T} \gamma_X(t-2, k-2)) \\
&= \left[ \prod_{l=0}^{k-1} \alpha_{1,t-l,T} \right] \text{var}(X_{t-k,T}),
\end{aligned}$$

which is the same expression obtained by [Brannas \[1995\]](#). Note that the last equality is obtained recursively. □

The time-varying transition probabilities follows similarly as that of the GINAR( $p$ ) process and is shown in [Theorem 5.3.2](#).

**Theorem 5.3.2** Let  $\{X_{t,T} : t = 1, 2, \dots, T\}$  for a positive integer  $T$  be a TV-GINAR( $p$ ) process of Definition 5.2.1. The transition probabilities of this process are given by

$$\begin{aligned}
& P(X_{t,T} = x | X_{t-1,T} = x_{t-1}, \dots, X_{t-p,T} = x_{t-p}) \\
&= \sum_{i_1=0}^x P(\alpha_{1,t,T} \odot x_{t-1} = i_1 | X_{t-1,T} = x_{t-1}) \times \\
&\quad \sum_{i_2=0}^{x-i_1} P(\alpha_{2,t,T} \odot x_{t-2} = i_2 | X_{t-2,T} = x_{t-2}) \times \dots \times \\
&\quad \sum_{i_p=0}^{x-(i_1+i_2+\dots+i_{p-1})} P(\alpha_{p,t,T} \odot x_{t-p} = i_p | X_{t-p,T} = x_{t-p}) \times \\
&\quad P(\xi_{t,T} = x - (i_1 + i_2 + \dots + i_p)).
\end{aligned}$$

We also have a Fourier domain representation, similar to the GINAR( $p$ ) case, which leads to a more computationally efficient algorithm as shown in Proposition 5.3.3.

**Proposition 5.3.3** Let  $\{X_{t,T} : t = 1, 2, \dots, T\}$  for a positive integer  $T$  be a TV-GINAR( $p$ ) process of Definition 5.2.1. Suppose that  $\phi_{X_{t,T}|X_{t-1,T}, \dots, X_{t-p,T}}(w)$  is the characteristic function for the transition probability defined in Theorem 5.3.2. Then the cumulative distribution function is

$$\begin{aligned}
a_t(x) &= P(X_{t,T} < x | X_{t-1,T}, \dots, X_{t-p,T}) \\
&= \frac{1}{2} - \frac{1}{2\pi} \int_{-\pi}^{\pi} \operatorname{Re} \left( \frac{\phi_{X_{t,T}|X_{t-1,T}, \dots, X_{t-p,T}}(w) e^{-iux}}{1 - e^{-iw}} \right) dw,
\end{aligned}$$

where

$$\phi_{X_{t,T}|X_{t-1,T}, \dots, X_{t-p,T}}(w) = \phi_{\xi_{t,T}}(w) \prod_{j=1}^p \left[ \phi_{Y_{t,T}^{(j)}}(w) \right]^{X_{t-j,T}}.$$

Then the transition probabilities can be calculated as

$$b_t(x) = P(X_{t,T} = x | X_{t-1,T}, \dots, X_{t-p,T}) = \begin{cases} a_t(1), & x = 0; \\ a_t(x+1) - a_t(x), & x = 1, 2, \dots, \end{cases}$$

where

$$b_t(x) = \frac{1}{\pi} \int_0^\pi \operatorname{Re}(\phi_{X_{t,T}|X_{t-1,T},\dots,X_{t-p,T}}(u) e^{-iux}) du, \quad x = 1, 2, \dots$$

In this chapter we defined a new class of processes for nonstationary count time series called TV-GINAR( $p$ ) processes and presented their statistical properties. We next discussed two methods of calculating the transition probabilities for these processes, which were an extension of the methods presented in Chapter 2 for stationary GINAR( $p$ ) processes. We also introduced possible basis functions we can use for modeling time-varying parameter curves in TV-GINAR processes. We will show further examples of defining time-varying parameter curves in later chapters.

## Chapter 6: Estimation for TV-INAR( $p$ ) Processes

In this chapter we present estimation methods for TV-GINAR( $p$ ) processes. In particular we extend the CML, CLS and pseudo likelihood estimation methods from Section 3 to TV-GINAR( $p$ ) processes. We further present the construction of confidence intervals using the Delta method and discuss model selection. Lastly, we present simulations comparing the performance of the aforementioned estimation methods for finite samples.

### 6.1 Estimation Methods

In this section we discuss parameter estimation for the TV-GINAR( $p$ ) model using conditional maximum likelihood (CML), conditional least squares (CLS) and pseudo maximum likelihood estimator. We are interested in estimation of the parameter vector  $\boldsymbol{\eta} = (\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \dots, \boldsymbol{\eta}_{p+2})^T$ . Let  $\Theta \subseteq \mathbb{R}^{p+2}$  be the resulting parameter space for  $\boldsymbol{\eta}$ , which we assume to be compact. We further assume that our TV-GINAR( $p$ ) process is identifiable; we can tell apart different values of the parameter vector on the basis of the transition probabilities.

#### 6.1.1 Conditional Maximum Likelihood (CML)

Conditioning on the first  $p$  observations, CML calculates the conditional log likelihood using

$$\ell(\boldsymbol{\eta}) = \sum_{t=p+1}^T \log P(X_{t,T} = x_{t,T} | X_{t-1,T}, \dots, X_{t-p,T}), \quad (6.1)$$

where the transition probabilities are defined in Theorem 5.3.2. Then CML-based parameter estimates of  $\boldsymbol{\eta}$  can then be computed as

$$\widehat{\boldsymbol{\eta}}_{CML} = \arg \max_{\boldsymbol{\eta} \in \Theta} \ell(\boldsymbol{\eta}). \quad (6.2)$$

Numerical optimization techniques are typically used to maximize (6.2), and the form of the transition probabilities will depend on the thinning operator and innovation distribution. The complexity of the optimization will depend on  $p$ , like the stationary GINAR( $p$ ) process case, but also the complexity of the basis functions used to model the dependence and innovation parameters.

### 6.1.2 Conditional Least Squares (CLS)

The CLS method for estimation of parameters of TV-GINAR( $p$ ) processes is a natural extension of the CLS method for GINAR( $p$ ) processes. We first define the modified parameter space which leaves out the innovation variance parameter vector  $\boldsymbol{\eta}_{p+2}$  from  $\boldsymbol{\eta}$ . Let  $\tilde{\boldsymbol{\eta}} = (\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \dots, \boldsymbol{\eta}_p, \boldsymbol{\eta}_{p+1})^T$ . Define

$$U_T(\tilde{\boldsymbol{\eta}}) = \sum_{t=p+1}^T \{X_{t,T} - \mu_{X_{t,T}|X_{t-1,T}, \dots, X_{t-p,T}}\}^2, \quad (6.3)$$

where  $\mu_{X_{t,T}|X_{t-1,T}, \dots, X_{t-p,T}}$  is defined in Lemma 5.3.1. Then the CLS estimator,  $\widehat{\boldsymbol{\eta}}_{CLS}$ , satisfies

$$\widehat{\boldsymbol{\eta}}_{CLS} = \arg \min_{\tilde{\boldsymbol{\eta}} \in \Theta} U_T(\tilde{\boldsymbol{\eta}}). \quad (6.4)$$

Similar to the stationary GINAR( $p$ ) process case we can estimate the innovation variance parameter vector,  $\boldsymbol{\eta}_{p+2}$ , using a two-step CLS method – we minimize

$$S_T(\tilde{\boldsymbol{\eta}}) = \sum_{t=p+1}^T \left[ \{X_{t,T} - \mu_{X_{t,T}|X_{t-1,T}, \dots, X_{t-p,T}}\}^2 - \sigma_{X_{t,T}|X_{t-1,T}, \dots, X_{t-p,T}}^2 \right]^2,$$

with respect to  $\boldsymbol{\eta}_{p+2}$ , while replacing  $\{\alpha_j(u) : u \in [0, 1], j = 1, 2, \dots, p\}$  and  $\{\mu_\xi(u) : u \in [0, 1]\}$  with the CLS estimates obtained from (6.4). The equation for  $\sigma_{X_{t,T}|X_{t-1,T}, \dots, X_{t-p,T}}^2$  is also given in Lemma 5.3.1.

### 6.1.3 Pseudo Maximum Likelihood

For the pseudo maximum likelihood method we approximate the transition probability using a normal distribution with mean and variance equal to the conditional mean and variance presented in Lemma 5.3.1. Let

$$\ell_P(\boldsymbol{\eta}) = -\frac{1}{2} \sum_{t=p+1}^T \log \left( 2\pi\sigma_{X_{t,T}|X_{t-1,T},\dots,X_{t-p,T}}^2(\boldsymbol{\eta}) \right) + \sum_{t=p+1}^T \frac{\left( x_{t,T} - \mu_{X_{t,T}|X_{t-1,T},\dots,X_{t-p,T}}(\boldsymbol{\eta}) \right)^2}{2\sigma_{X_{t,T}|X_{t-1,T},\dots,X_{t-p,T}}^2(\boldsymbol{\eta})}.$$

Then the pseudo maximum likelihood estimator  $\widehat{\boldsymbol{\eta}}_P$  is

$$\widehat{\boldsymbol{\eta}}_P = \arg \max_{\boldsymbol{\eta} \in \Theta} \ell_P(\boldsymbol{\eta}).$$

## 6.2 Confidence Intervals

In this section, we derive confidence intervals for the TV-GINAR( $p$ ) process parameters. We make normality assumptions about the estimated process vectors and then use the Delta method to obtain the form of the pointwise confidence intervals. We assume the following asymptotic distribution for  $\widehat{\boldsymbol{\eta}}$ :

$$\sqrt{T}(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}) \xrightarrow{d} N_m(\mathbf{0}, V(\boldsymbol{\eta})) \quad (6.5)$$

as  $T \rightarrow \infty$  where  $V(\boldsymbol{\eta}) = [V_{jk}(\boldsymbol{\eta}) : j, k = 1, 2, \dots, p+2]$  and  $m = \sum_{j=1}^{p+2} m_j$  ( $m_j$  was defined in Section 5.2). Note that  $V_{jk}(\boldsymbol{\eta})$  are block covariance matrices. In practice, we estimate  $V_{jk}(\boldsymbol{\eta})$  using the Hessian, calculated numerically from the optimization algorithm that generates the estimator.

We first define the following notation:

$$\mathbf{v}_j^T(u) = \begin{bmatrix} \mathbf{v}_{j,1}^T(u) \\ \mathbf{v}_{j,2}^T(u) \\ \vdots \\ \mathbf{v}_{j,T}^T(u) \end{bmatrix}_{T \times m_j} \quad \text{and} \quad \boldsymbol{\eta}_j = \begin{bmatrix} \eta_{j,1} \\ \eta_{j,2} \\ \vdots \\ \eta_{j,m_j} \end{bmatrix}_{m_j \times 1},$$

where  $j = 1, 2, \dots, p+2$  and  $u \in [0, 1]$ . Hence,  $\mathbf{v}_j^T(u)$  represents the  $j^{\text{th}}$  matrix of  $m_j$  smooth basis functions,  $\boldsymbol{\eta}_j$  is the corresponding  $m_j$  dimensional vector of coefficients, and  $\mathbf{v}_{j,t}^T$  is the  $t^{\text{th}}$  row of  $\mathbf{v}_j^T(u)$  for  $t = 1, 2, \dots, T$ . We represent the elements of this row vector as  $\mathbf{v}_{j,t}^T = (v_{j,t,1}, v_{j,t,2}, \dots, v_{j,t,m_j})$ .

Using the above notation, we define the following:

$$h_{q,t}(\boldsymbol{\eta}; u) = \frac{\exp(\mathbf{v}_{q,t}^T(u)\boldsymbol{\eta}_q)}{1 + \sum_{k=1}^p \exp(\mathbf{v}_{k,t}^T(u)\boldsymbol{\eta}_k)}, \quad u \in [0, 1], \quad (6.6)$$

where  $t = 1, 2, \dots, T$  and  $q = 1, 2, \dots, p$ . This represents the  $t^{\text{th}}$  element of the autocorrelation parameter curve  $\{\alpha_q(u) : u \in [0, 1]\}$ . We now derive the asymptotic distribution of (6.6).

Assume that (6.5) is true. Then by the Delta method we have the following:

$$\sqrt{T}(h_{q,t}(\hat{\boldsymbol{\eta}}; u) - h_{q,t}(\boldsymbol{\eta}; u)) \xrightarrow{d} N(0, \tau_{q,t}^2(u)), \quad u \in [0, 1],$$

as  $T \rightarrow \infty$  where  $\tau_{q,t}^2(u) = S_{q,t}V(\boldsymbol{\eta})S_{q,t}^T$ . We next derive the components of the asymptotic variance  $\tau_{q,t}^2(u)$ .

First let  $d_{q,t}(u) = \exp(\mathbf{v}_{q,t}^T(u)\boldsymbol{\eta}_q)$ . Let  $r = 1, 2, \dots, p+2$ , and  $s = 1, 2, \dots, m_r$ , and note that  $S_{q,t}$  is a  $1 \times m$  vector whose elements can be obtained as follows:

$$\frac{\partial h_{q,t}(\boldsymbol{\eta}, u)}{\partial \eta_{rs}} = \frac{\partial h_{q,t}(\boldsymbol{\eta}, u)}{\partial d_{r,t}(u)} \times \frac{\partial d_{r,t}(u)}{\partial \eta_{r,s}}. \quad (6.7)$$

The components of (6.7) are derived as follows:

$$\frac{\partial h_{q,t}(\boldsymbol{\eta}, u)}{\partial d_{r,t}(u)} = \begin{cases} \frac{1 + \sum_{k=1}^p d_{k,t}(u) - d_{q,t}(u)}{(1 + \sum_{k=1}^p d_{k,t}(u))^2}, & \text{if } r = q; \\ \frac{-d_{q,t}(u)}{(1 + \sum_{k=1}^p d_{k,t}(u))^2}, & \text{if } r \neq q \text{ and } r \leq p; \\ 0, & \text{if } r > p. \end{cases}$$

and

$$\frac{\partial d_{r,t}(u)}{\partial \eta_{rs}} = \frac{\partial \exp(\sum_{w=1}^{m_r} v_{r,t,w} \eta_{r,w})}{\partial \eta_{rs}} = v_{r,t,s} d_{r,t}(u).$$

Next, we show the asymptotic distribution for the innovation mean and variance. First, define the following:

$$w_{a,t}(\boldsymbol{\eta}_a, u) = \exp(\mathbf{v}_{a,t}^T(u) \boldsymbol{\eta}_a),$$

where  $a = p+1, p+2$ . Then again assuming that (6.5) is true, by the Delta method, as  $T \rightarrow \infty$ ,

$$\sqrt{T}(w_{a,t}(\widehat{\boldsymbol{\eta}}_a, u) - w_{a,t}(\boldsymbol{\eta}_a, u)) \xrightarrow{d} N(0, \zeta_{a,t}^2(u)), \quad u \in [0, 1],$$

where  $\zeta_{a,t}^2(u) = R_{q,t} V(\boldsymbol{\eta}) R_{q,t}^T$ . Here  $R_{q,t}$  is a  $1 \times m$  vector with elements:

$$\frac{\partial w_{a,t}(\boldsymbol{\eta}_a, u)}{\partial \eta_{rs}} = \begin{cases} v_{p+1,t,s} w_{p+1,t}(\boldsymbol{\eta}_{p+1}, u), & \text{if } r = p+1; \\ v_{p+2,t,s} w_{p+2,t}(\boldsymbol{\eta}_{p+2}, u), & \text{if } r = p+2; \\ 0, & \text{otherwise.} \end{cases}$$

We can utilize the covariance matrices derived in this section to construct pointwise confidence intervals for the parameters of the TV-GINAR( $p$ ) process, as we will demonstrate in the applications presented in Chapter 7.

As an example, consider a TV-GINAR(1) process with binomial thinning and Poisson innovations. Assume that the innovation mean is constant and the dependence parameter curve is defined as follows:

$$\alpha_1(u) = \frac{\exp(\eta_{1,1} + \eta_{1,1}u)}{1 + \exp(\eta_{1,0} + \eta_{1,1}u)}, \quad u \in [0, 1].$$

To express this dependence curve in terms of the notation of (6.6) we define the following components:

$$\mathbf{v}_1^T(u) = \begin{bmatrix} \mathbf{v}_{1,1}^T(u) \\ \mathbf{v}_{1,2}^T(u) \\ \vdots \\ \mathbf{v}_{1,T}^T(u) \end{bmatrix}_{T \times 2} \quad \text{and} \quad \boldsymbol{\eta}_1 = \begin{bmatrix} \eta_{1,1} \\ \eta_{1,2} \end{bmatrix}_{2 \times 1}.$$

Note that  $\mathbf{v}_{1,t}^T = (1, t/T)$ , for  $t = 1, 2, \dots, T$ . Then

$$h_{1,t}(\boldsymbol{\eta}, u) = \frac{\exp(\mathbf{v}_1^T(u)\boldsymbol{\eta}_1)}{1 + \exp(\mathbf{v}_1^T(u)\boldsymbol{\eta}_1)}, \quad u \in [0, 1].$$

Also, since the innovation mean is constant we express it as follows:

$$\mathbf{v}_2^T(u) = \begin{bmatrix} \mathbf{v}_{2,1}^T(u) \\ \mathbf{v}_{2,2}^T(u) \\ \vdots \\ \mathbf{v}_{2,T}^T(u) \end{bmatrix}_{T \times 1} \quad \text{and} \quad \boldsymbol{\eta}_2 = [\eta_{2,1}]_{1 \times 1},$$

where  $\mathbf{v}_{2,t}^T = 1$ , for  $t = 1, 2, \dots, T$ . Then the full parameter vector is  $\boldsymbol{\eta} = (\boldsymbol{\eta}_1^T, \boldsymbol{\eta}_2^T) = (\eta_{1,1}, \eta_{1,2}, \eta_{2,1})^T$ .

Consider fixed  $t$ , we can build an approximate  $(1 - \nu)100\%$  confidence interval for  $h_{1,1}(\boldsymbol{\eta}, u)$  as follows:

$$h_{1,1}(\hat{\boldsymbol{\eta}}, u) \pm z_{1-\nu/2} \tau_{1,1}(u), \quad (6.8)$$

where  $u = t/T$ . To calculate  $\tau_{1,1}(u)$  we use (6.7). First note that for this example  $r = 1, 2$ , with  $m_1 = 2$  and  $m_2 = 1$ . Then

$$\begin{aligned} \frac{\partial h_{1,1}(\boldsymbol{\eta}, u)}{\partial \eta_{1,1}} &= \frac{\exp(\mathbf{v}_{1,1}^T(u)\boldsymbol{\eta}_1)}{(1 + \exp(\mathbf{v}_{1,1}^T(u)\boldsymbol{\eta}_1))^2}, \\ \frac{\partial h_{1,1}(\boldsymbol{\eta}, u)}{\partial \eta_{1,2}} &= \frac{\exp(\mathbf{v}_{1,1}^T(u)\boldsymbol{\eta}_1)}{T (1 + \exp(\mathbf{v}_{1,1}^T(u)\boldsymbol{\eta}_1))^2}, \end{aligned}$$

and

$$\frac{\partial h_{1,1}(\boldsymbol{\eta}, u)}{\partial \eta_{2,1}} = 0.$$

Thus  $S_{1,1} = \left( \frac{\partial h_{1,1}(\boldsymbol{\eta}, u)}{\partial \eta_{1,1}}, \frac{\partial h_{1,1}(\boldsymbol{\eta}, u)}{\partial \eta_{1,2}}, \frac{\partial h_{1,1}(\boldsymbol{\eta}, u)}{\partial \eta_{2,1}} \right)$  and  $V(\boldsymbol{\eta})$  is obtained using the estimated Hessian. Hence, we have all the components to calculate  $\tau_{1,1}^2(u)$  and consequently evaluate (6.8).

### 6.3 Model Selection

In this section, we discuss model selection for TV-GINAR( $p$ ) processes. In particular, we discuss how to determine if a process is time-varying or not. We can do this via the model selection techniques already discussed in Section 3.2 using the AIC. In the case of a stationary process, we can estimate the process parameters using CML and then calculate the AIC based on the conditional likelihood from Section 2.5. For the non-stationary TV-GINAR( $p$ ) process we can use the CML method described in Chapter 6 and calculate the AIC. By comparing the AIC values from these two CML methods we can determine which process is a better fit. We investigate the performance of these methods in the simulations presented in Section 6.4.

As an example, consider the following TV-GINAR(1) process with binomial thinning and Poisson innovations. We let the innovation mean parameter be constant in time and characterize the time-varying dependence parameter as follows:

$$\alpha_1(u) = \frac{\exp(\eta_{1,0} + \eta_{1,1}u)}{1 + \exp(\eta_{1,0} + \eta_{1,1}u)}, \quad u \in [0, 1],$$

where  $\eta_{1,0}$  and  $\eta_{1,1}$  are constant and do not vary with time. Note that the stationary GINAR( $p$ ) process is a special case of this process when  $\eta_{1,1} = 0$ . Then the AIC for the two

models is calculated as follows:

$$AIC_{TV-GINAR(1)} = 2k_{TV} - 2\ell(\hat{\boldsymbol{\eta}}_{CML}), \text{ and}$$

$$AIC_{GINAR(1)} = 2k_G - 2\ell(\hat{\boldsymbol{\theta}}_{CML}),$$

where  $k_{TV}$  and  $k_G$  are the number of parameters in the TV-GINAR(1) and GINAR(1) process respectively. For this example,  $k_{TV} = 3$  and  $k_G = 2$ . Also,  $\hat{\boldsymbol{\eta}}_{CML}$  and  $\hat{\boldsymbol{\theta}}_{CML}$  are the CML estimates described in Chapter 6 and Chapter 3 respectively. We can then decide which model is appropriate, i.e. stationary or non-stationary, based on which one has a lower AIC value.

## 6.4 Simulation Studies for TV-GINAR( $p$ ) Processes

In this section, we compare the performance of the CML, CLS, and pseudo likelihood estimation methods for finite samples for the TV-GINAR( $p$ ) process with binomial thinning and Poisson innovations. We consider TV-GINAR(1) and TV-GINAR(2) processes.

In all simulations, we estimate each quantity of interest (the bias, standard deviation (SD), root mean squared error (RMSE), or coverage) using 10,000 replicates, and estimate standard errors for each quantity using 10,000 bootstrap samples.

### 6.4.1 Estimating TV-GINAR(1) process parameters

We start by considering the estimation of a TV-GINAR(1) process. In particular, we consider two cases. In the first case, we allow for time-varying dependence parameters while the innovation sequence parameters are held constant over time. In the second case we allow both the dependence parameter and innovation sequence parameters to vary with time. The

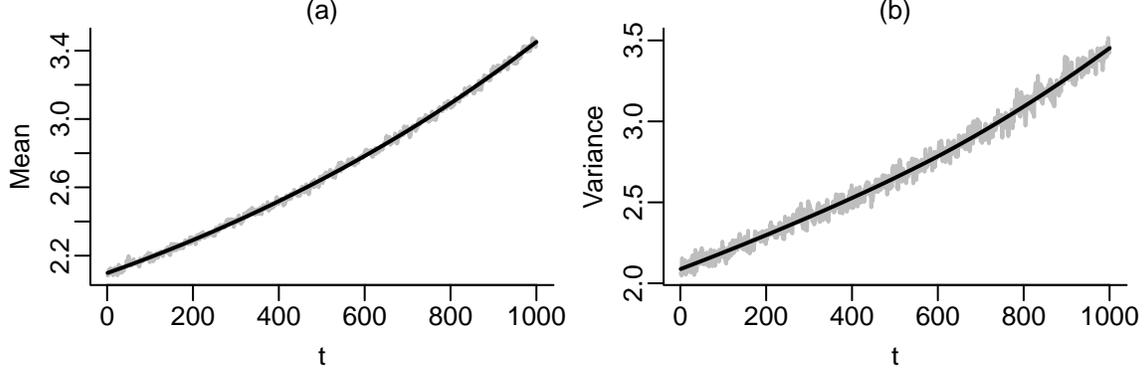


Figure 6.1: (a) Marginal mean as a function of time for the 10,000 realizations in gray and the smoothed result over time, using loess smoothing, shown in black. (b) Marginal variance as a function of time for the 10,000 realizations in gray and the smoothed result over time shown in black.

time-varying dependence parameter,  $\{\alpha_1(u) : u \in [0, 1]\}$ , curve is defined as follows:

$$\alpha_1(u) = \frac{\exp(\eta_{1,0} + \eta_{1,1}u)}{1 + \exp(\eta_{1,0} + \eta_{1,1}u)} \quad (6.9)$$

where  $\eta_{1,0} = 0.1$ ,  $\eta_{1,1} = 0.8$  and  $u \in [0, 1]$ .

The time-varying innovation mean parameter curve,  $\{\mu_\xi(u) : u \in [0, 1]\}$ , is defined as follows:

$$\mu_\xi(u) = \exp(\eta_{2,0} + \eta_{2,1}u), \quad (6.10)$$

where  $\eta_{2,0} = 0$ ,  $\eta_{2,1} = 0.5$  and  $u \in [0, 1]$ . In the case that the innovation mean is not time-varying we set  $\mu_\xi(u) = 1$  for all  $u$ .

Note that in the case while the innovation mean and variance parameter are constant over time, the marginal mean and variance of the process are not constant over time. To see this, consider 10,000 realizations of this TV-GINAR(1). We calculate the mean and variance of each of the 10,000 realizations for each time point  $t$ , where  $t = 1, 2, \dots, 1000$ . Figure 6.1 shows the marginal mean and variance with respect to time, and we can clearly see that

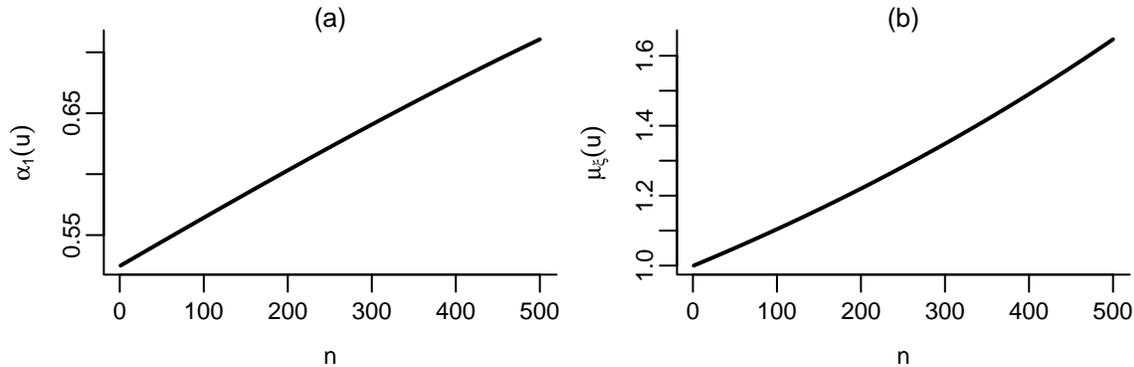


Figure 6.2: (a) Parametrization of  $\{\alpha_1(u) : u \in [0, 1]\}$  for TV-GINAR(1) process simulation. (b) Parametrization of  $\{\mu_\xi(u) : u \in [0, 1]\}$  for TV-GINAR(1) process simulation.

neither is constant over time. Note that the marginal mean and variance derived in Section 5.3.1 are not constant over time either.

For a process of length 500, the time-varying parameter curves are depicted in Figure 6.2. This representation allows for a series with increasing dependence and increasing innovation mean parameters over time.

Table 6.1 shows the simulation results for the estimation of parameters in the first case (constant innovation parameters over time and time-varying dependence parameter). We see that the SD and RMSE for estimating the constant over time innovation mean parameter,  $\mu_\xi$ , is smaller compared to estimating  $\eta_{1,0}$  and  $\eta_{1,1}$ . When the series length is  $T = 500$ , the RMSE of both CLS and pseudo estimation methods are much larger than the CML method for all parameters. When  $T = 1,000$ , the CML method still has the lowest bias, SD, and RMSE, with pseudo likelihood method being comparable to the CML method. In comparison to the  $T = 500$  simulation, the  $T = 1,000$  simulation results have biases closer to zero, as well as smaller SD and RMSE. This is to be expected as the sample size increases.

Table 6.2 shows the simulation results for the estimation of parameters in the second case (time-varying innovation and dependence parameters). The parameters associated with the

Table 6.1: Estimated bias, SD, and RMSE when estimating the parameters of a TV-PoINAR(1) process with  $\eta_{1,0} = 0.1$ ,  $\eta_{1,1} = 0.8$  and  $\mu_\xi = 1$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.002 for estimating  $\mu_\xi$ , 0.005 for estimating  $\eta_{1,0}$ , and 0.006 for estimating  $\eta_{1,1}$ .

Method		$T = 500$			$T = 1,000$		
		$\hat{\mu}_\xi$	$\hat{\eta}_{1,0}$	$\hat{\eta}_{1,1}$	$\hat{\mu}_\xi$	$\hat{\eta}_{1,0}$	$\hat{\eta}_{1,1}$
CML	Bias	-0.004	-0.006	0.002	-0.003	-0.001	-0.001
	SD	0.073	0.193	0.284	0.052	0.137	0.202
	RMSE	0.073	0.193	0.284	0.052	0.137	0.202
CLS	Bias	0.019	-0.040	-0.001	0.009	-0.021	0.003
	SD	0.101	0.238	0.316	0.072	0.169	0.221
	RMSE	0.103	0.241	0.316	0.073	0.171	0.221
Pseudo	Bias	0.001	-0.017	0.006	0.000	-0.009	0.002
	SD	0.078	0.207	0.303	0.055	0.142	0.208
	RMSE	0.078	0.207	0.303	0.055	0.142	0.208

Table 6.2: Estimated bias, SD, and RMSE when estimating the parameters of a TV-PoINAR(1) process with  $\eta_{1,0} = 0.1$ ,  $\eta_{1,1} = 0.8$ ,  $\eta_{2,0} = 0$  and  $\eta_{2,1} = 0.5$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.004 for estimating  $\hat{\eta}_{2,0}$ , 0.007 for estimating  $\hat{\eta}_{2,1}$ , 0.006 for estimating  $\eta_{1,0}$  and 0.011 for estimating  $\eta_{1,1}$ .

Method		$T = 500$				$T = 1,000$			
		$\hat{\eta}_{2,0}$	$\hat{\eta}_{2,1}$	$\hat{\eta}_{1,0}$	$\hat{\eta}_{1,1}$	$\hat{\eta}_{2,0}$	$\hat{\eta}_{2,1}$	$\hat{\eta}_{1,0}$	$\hat{\eta}_{1,1}$
CML	Bias	-0.017	0.025	0.009	-0.026	-0.010	0.015	0.005	-0.012
	SD	0.141	0.141	0.228	0.372	0.101	0.101	0.163	0.268
	RMSE	0.142	0.142	0.229	0.373	0.101	0.101	0.164	0.268
CLS	Bias	-0.004	0.051	-0.014	-0.066	-0.001	0.021	-0.008	-0.027
	SD	0.203	0.203	0.334	0.571	0.140	0.140	0.232	0.399
	RMSE	0.203	0.203	0.334	0.575	0.140	0.140	0.232	0.400
Pseudo	Bias	-0.005	0.006	-0.010	0.003	-0.001	0.002	-0.006	0.001
	SD	0.151	0.151	0.248	0.397	0.108	0.108	0.174	0.279
	RMSE	0.151	0.151	0.248	0.397	0.108	0.108	0.174	0.280

innovation mean parameter curve,  $\eta_{2,0}$  and  $\eta_{2,1}$ , have larger bias, SD, and RMSE as compared to the constant in time innovation mean simulation (Table 6.1) for all estimation methods. The RMSE for the dependence parameter curve parameters is also larger compared to those in Table 6.1. Overall, the CML method has the lowest bias, SD, and RMSE for both values of  $T$  considered. Again for larger sample length of  $T = 1,000$  we see that the bias is closer to zero, and the SD and RMSE are smaller, as compared to a sample length of  $T = 500$ . In terms of time complexity, the CML method performs well too. For the simulation in Table 6.2 ( $T = 500$ ), 100 simulations using the CML method take approximately 0.03 seconds, whereas CLS and pseudo likelihood estimators take an average of 1.35 seconds. Note that the CML method estimation code was run in C++ which explains the faster performance compared to CLS and pseudo likelihood estimation methods.

Lastly, we consider model selection for this TV-GINAR(1) process using the technique described in Section 6.3. The true process is a TV-GINAR(1) process of length  $T = 1,000$  defined with binomial thinning and Poisson innovations. Our goal is to determine if we can correctly identify if the process is time-varying or not. For this, we compare the AIC calculated using the true TV-GINAR(1) conditional maximum likelihood and the AIC calculated assuming the process is GINAR(1). The true process is the TV-GINAR(1) process described in this section with the dependence and innovation parameter curves as described in (6.9) and (6.10). We carried out 10,000 simulations comparing the two AICs, and then calculated the win rates. The win rate is defined as the percentage of times the TV-GINAR(1) model AIC value is less than the GINAR(1) AIC value. The true TV-GINAR(1) AIC had a win rate of 97.8% in the first case and a win rate of 100% in the second case. This shows that we are able to correctly identify a time-varying process using this method.

### 6.4.2 Estimating TV-GINAR(2) process parameters

In this section, we consider the estimation of the parameter curves of TV-GINAR(2) processes. Like the TV-GINAR(1) simulations, we consider two cases. In the first case, we allow for time-varying dependence parameters while the innovation sequence parameters are held constant over time. In the second case, we allow both the dependence parameter and innovation sequence parameters to vary with time. The time-varying dependence parameter curves,  $\{\alpha_1(u) : u \in [0, 1]\}$  and  $\{\alpha_2(u) : u \in [0, 1]\}$ , used for the simulations are defined as follows:

$$\alpha_1(u) = \frac{\exp(\eta_{1,0} + \eta_{1,1}u)}{1 + \sum_{k=1}^2 \exp(\eta_{k,0} + \eta_{k,1}u)} \quad (6.11)$$

and

$$\alpha_2(u) = \frac{\exp(\eta_{2,0} + \eta_{2,1}u)}{1 + \sum_{k=1}^2 \exp(\eta_{k,0} + \eta_{k,1}u)}, \quad u \in [0, 1], \quad (6.12)$$

where  $\eta_{1,0} = -0.1$ ,  $\eta_{1,1} = -0.8$ ,  $\eta_{2,0} = 0.1$ , and  $\eta_{2,1} = 0.6$ . The time-varying innovation mean parameter curve,  $\{\mu_\xi(u) : u \in [0, 1]\}$ , is defined as follows:

$$\mu_\xi(u) = \exp(\eta_{3,0} + \eta_{3,1}u), \quad u \in [0, 1], \quad (6.13)$$

where  $\eta_{3,0} = 0$  and  $\eta_{3,1} = 0.5$ . In the simulations where the innovation mean is constant over time we set  $\mu_\xi(u) = 1 = \mu_\xi$ , say, for all  $u \in [0, 1]$ . Similar to the TV-GINAR(1) simulations, even when the innovation sequence parameters are constant over time, the marginal mean and variance are not.

For a process of length  $T = 500$ , the time-varying dependence parameter curves and their spectral density for certain time points are depicted in Figures 6.3 and 6.4 respectively. With this choice of dependence parameters, we have a representation where  $\{\alpha_1(u) : u \in [0, 1]\}$  is gradually decreasing and  $\{\alpha_2(u) : u \in [0, 1]\}$  is increasing over time. Also, Figure 6.4 shows

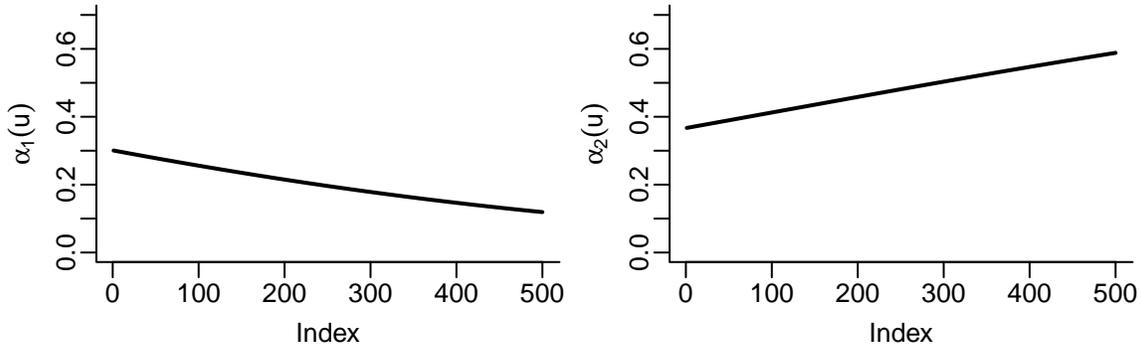


Figure 6.3: Parametrization of  $\{\alpha_1(u) : u \in [0, 1]\}$  and  $\{\alpha_2(u) : u \in [0, 1]\}$  for TV-GINAR(2) process simulation.

the spectral density for certain time points, showing how the spectral properties change over time. At larger time points we see more prominent peaks developing.

Tables 6.3 and 6.4 show the simulation results for the estimation of model parameters when the series length is 500 and 1,000 respectively and the innovation mean is constant over time. We see that when  $T = 500$ , all estimation methods have the largest absolute bias, SD, and RMSE when estimating the parameters as compared to  $T = 1,000$ . In this case, CLS and pseudo have much larger RMSE than CML. The performance for all estimation methods improves with the longer  $T = 1,000$  as compared to  $T = 500$ , as seen in Table 6.6, which is expected as sample size increases. Similar to the TV-GINAR(1) simulation results, in terms of bias closest to zero, smaller SD, and smaller RMSE, CML outperforms both CLS and pseudo estimation methods.

Tables 6.5 and 6.6 show the simulation results for the estimation of model parameters when the series length is 500 and 1,000 respectively, and both the innovation and dependence parameters are time-varying. We see that the SD and RMSE for all estimation methods has increased compared to the constant innovation mean simulations. Overall, compared to the CLS and pseudo likelihood estimation methods, CML has the lowest absolute bias,

Table 6.3: Estimated bias, SD, and RMSE when estimating the parameters of a TV-PoINAR(2) process with  $\eta_{1,0} = -0.1$ ,  $\eta_{1,1} = -0.8$ ,  $\eta_{2,0} = 0.1$ ,  $\eta_{2,1} = 0.6$  and  $\mu_\xi = 1$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.003 for estimating  $\mu_\xi$ , 0.152 for estimating  $\eta_{1,0}$  and 1.613 for estimating  $\eta_{1,1}$ , 0.006 for estimating  $\eta_{2,0}$  and 0.009 for estimating  $\eta_{2,1}$ .

Method		$T = 500$				
		$\hat{\mu}_\xi$	$\hat{\eta}_{1,0}$	$\hat{\eta}_{1,1}$	$\hat{\eta}_{2,0}$	$\hat{\eta}_{2,1}$
CML	Bias	0.021	-0.040	-0.126	-0.055	0.022
	SD	0.120	0.437	0.883	0.295	0.404
	RMSE	0.122	0.439	0.893	0.300	0.404
CLS	Bias	0.050	-0.093	-0.144	-0.110	0.028
	SD	0.143	0.449	2.678	0.334	0.437
	RMSE	0.152	0.458	2.678	0.352	0.438
Pseudo	Bias	0.018	-0.058	-0.103	-0.059	0.021
	SD	0.124	0.560	1.134	0.304	0.419
	RMSE	0.126	0.563	1.139	0.309	0.420

Table 6.4: Estimated bias, SD, and RMSE when estimating the parameters of a TV-PoINAR(2) process with  $\eta_{1,0} = -0.1$ ,  $\eta_{1,1} = -0.8$ ,  $\eta_{2,0} = 0.1$ ,  $\eta_{2,1} = 0.6$  and  $\mu_\xi = 1$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.002 for estimating  $\mu_\xi$ , 0.006 for estimating  $\eta_{1,0}$  and 0.011 for estimating  $\eta_{1,1}$ , 0.005 for estimating  $\eta_{2,0}$  and 0.005 for estimating  $\eta_{2,1}$ .

Method		$T = 1,000$				
		$\hat{\mu}_\xi$	$\hat{\eta}_{1,0}$	$\hat{\eta}_{1,1}$	$\hat{\eta}_{2,0}$	$\hat{\eta}_{2,1}$
CML	Bias	0.010	-0.024	-0.041	-0.026	0.009
	SD	0.085	0.284	0.526	0.200	0.276
	RMSE	0.086	0.285	0.528	0.201	0.276
CLS	Bias	0.025	-0.052	-0.031	-0.050	0.009
	SD	0.101	0.308	0.576	0.227	0.298
	RMSE	0.104	0.312	0.576	0.233	0.298
Pseudo	Bias	0.009	-0.029	-0.037	-0.032	0.016
	SD	0.088	0.297	0.554	0.210	0.286
	RMSE	0.089	0.299	0.556	0.212	0.286

Table 6.5: Estimated bias, SD, and RMSE when estimating the parameters of a TV-PoINAR(2) process with  $\eta_{1,0} = -0.1$ ,  $\eta_{1,1} = -0.8$ ,  $\eta_{2,0} = 0.1$ ,  $\eta_{2,1} = 0.6$ ,  $\eta_{3,0} = 0$  and  $\eta_{3,1} = 0.5$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.006 for estimating  $\eta_{3,0}$ , 0.010 for estimating  $\eta_{3,1}$ , 0.082 for estimating  $\eta_{1,0}$  and 1.615 for estimating  $\eta_{1,1}$ , 0.008 for estimating  $\eta_{2,0}$  and 0.015 for estimating  $\eta_{2,1}$ .

Method		$T = 500$					
		$\hat{\eta}_{3,0}$	$\hat{\eta}_{3,1}$	$\hat{\eta}_{1,0}$	$\hat{\eta}_{1,1}$	$\hat{\eta}_{2,0}$	$\hat{\eta}_{2,1}$
CML	Bias	0.023	-0.008	-0.060	-0.115	-0.058	0.027
	SD	0.241	0.423	0.596	1.398	0.371	0.602
	RMSE	0.242	0.423	0.598	1.402	0.376	0.603
CLS	Bias	0.037	0.022	-0.102	-0.165	-0.083	-0.027
	SD	0.291	0.513	0.705	3.118	0.450	0.756
	RMSE	0.293	0.514	0.712	3.120	0.457	0.756
Pseudo	Bias	0.022	-0.007	-0.065	-0.128	-0.060	0.026
	SD	0.244	0.431	0.611	1.443	0.381	0.619
	RMSE	0.245	0.431	0.614	1.450	0.386	0.620

Table 6.6: Estimated bias, SD, and RMSE when estimating the parameters of a TV-PoINAR(1) process with  $\eta_{1,0} = 0.1$ ,  $\eta_{1,1} = 0.8$  and  $\mu_\xi = 1$ . The number of replicates is 10,000. The maximum standard error for each quantity is 0.004 for estimating  $\eta_{3,0}$ , 0.007 for estimating  $\eta_{3,1}$ , 0.013 for estimating  $\eta_{1,0}$ , 0.017 for estimating  $\eta_{1,1}$ , 0.006 for estimating  $\eta_{2,0}$  and 0.010 for estimating  $\eta_{2,1}$ .

Method		$T = 1,000$					
		$\hat{\eta}_{3,0}$	$\hat{\eta}_{3,1}$	$\hat{\eta}_{1,0}$	$\hat{\eta}_{1,1}$	$\hat{\eta}_{2,0}$	$\hat{\eta}_{2,1}$
CML	Bias	0.010	0.000	-0.023	-0.062	-0.029	0.015
	SD	0.163	0.288	0.370	0.776	0.252	0.410
	RMSE	0.163	0.288	0.371	0.778	0.253	0.410
CLS	Bias	0.018	0.015	-0.044	-0.063	-0.041	-0.015
	SD	0.203	0.359	0.421	0.847	0.315	0.525
	RMSE	0.204	0.359	0.423	0.849	0.318	0.526
Pseudo	Bias	0.007	0.002	-0.021	-0.058	-0.027	0.009
	SD	0.172	0.302	0.397	0.818	0.265	0.428
	RMSE	0.172	0.302	0.397	0.821	0.266	0.428

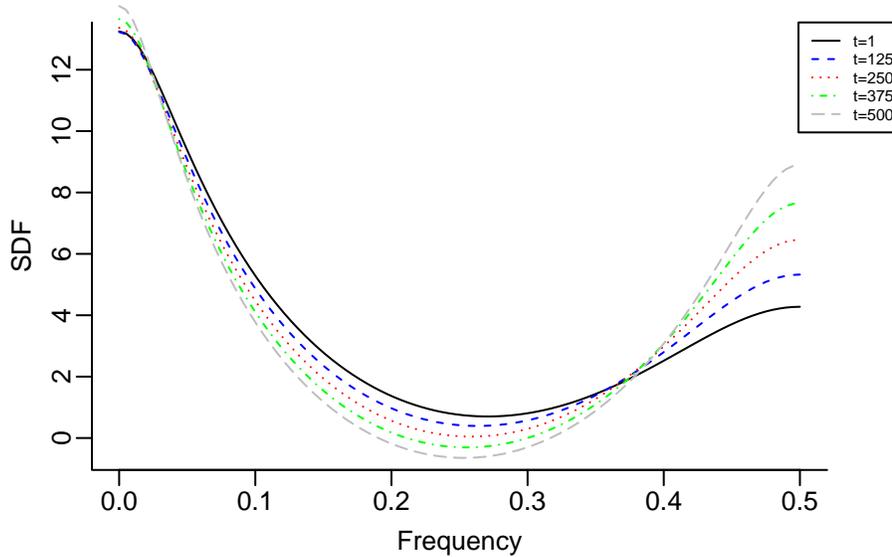


Figure 6.4: Spectral density for specified  $t$  values for TV-GINAR(2) process when innovation parameter is constant.

SD, and RMSE for these simulations as well. We look at the computational time for the simulation in Table 6.4. For 100 simulations, CML takes approximately 5 seconds, whereas CLS and pseudo estimators take an average of 3.5 seconds. Note that the CML method was implemented in C++ and CLS was implemented in R.

Lastly, we consider model selection for the TV-GINAR(2) process using the technique described in Section 6.3. The true process is a TV-GINAR(2) process with binomial thinning and Poisson innovations and  $T = 1,000$ . The dependence and innovation parameter curves are as described in (6.11), (6.12) and (6.13). Similar to the TV-GINAR(1) simulation we want to be able to determine if we can tell apart whether the model is time-varying or not. For this, we compare the AIC calculated using the true TV-GINAR(2) likelihood and the AIC calculated assuming the process is GINAR(2), both using CML. We carried out 10,000 simulations comparing the two AICs, and then calculated the win rates. The win rate is defined as the percentage of times the TV-GINAR(2) model AIC value is less than

the GINAR(2) AIC value. The true TV-GINAR(2) AIC had a win rate of 100% in both cases: when only the dependence parameter is time-varying and when both the dependence and innovation parameters are time-varying. Hence, we are correctly able to identify the time-varying process.

### 6.4.3 Conclusions

In conclusion, from the simulation presented in Sections 6.4.1 and 6.4.2 we see that the CML estimation method outperforms the CLS method and pseudo likelihood method in almost all simulations, especially in cases when all parameters of the process are time-varying. The merits of CML are even more pronounced when we increase model order from  $p = 1$  to  $p = 2$ , as the difference in absolute bias, SD, and RMSE of the CML estimation method compared to CLS and pseudo likelihood estimation methods gets larger. Also, it is harder to estimate the process parameters when we vary both the dependence and innovation parameters over time as compared to when only the dependence parameter is time-varying. Lastly, we also discussed model selection techniques for these processes, and showed that AIC can be used as a tool to determine if a process is time-varying or not since the win rates for the true time-varying model were approximately 100% for all cases studied.

## Chapter 7: Applications

In this chapter, we use  $\text{GINAR}(p)$  and  $\text{TV-GINAR}(p)$  processes to model, infer, and forecast a disease surveillance series and a time series of perceptual speed scores for a schizophrenic patient. A count time series model, and in particular  $\text{GINAR}(p)$  processes, are appropriate for these applications because they are low-count data and show evidence of a mean-variance relationship, and thus the Gaussianity assumption is not appropriate. Furthermore,  $\text{GINAR}$  models have been considered for applications in the medical sciences and shown successful performance (e.g. [Franke and Seligmann \[1993\]](#), [Cardinal et al. \[1999\]](#)). A recent paper by [Khan et al. \[2022\]](#) even applied  $\text{INAR}$  models to a COVID-19 death series. We consider various models for each application, with varying orders, innovation distributions, thinning operators, and model assumptions. We compare these models with each other based on AIC, root mean squared error, forecast coverage, and goodness of fit and interpret the time-varying parameter curves for the two applications.

### 7.1 Disease Surveillance

In this section we consider a disease surveillance series. Modeling disease counts is crucial in understanding disease dynamics like incubation period, and is important for allowing timely interventions and resource allocations to curb the spread of a disease [[McCleary et al., 1980](#)]. As such, we use  $\text{GINAR}(p)$  and  $\text{TV-GINAR}(p)$  processes to study meningococcal

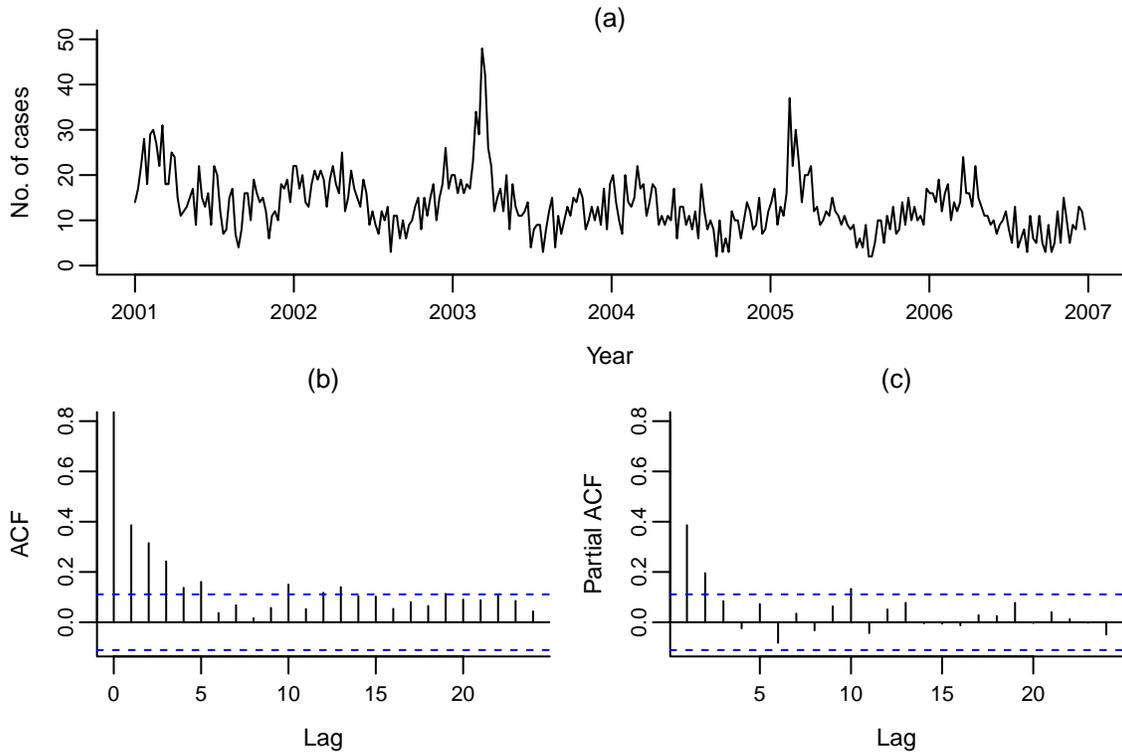


Figure 7.1: (a) Time series plot of the weekly meningococcal disease cases; (b) sample ACF and (c) sample PACF of residuals after removing a yearly sinusoidal seasonality term.

disease cases in Germany, as it is a low-count time series dataset depicting overdispersion. Figure 7.1 shows a summary of a time series of  $n = 313$  weekly meningococcal disease cases in Germany, over the period 2001–2006. The data was studied by Pedeli et al. [2015].

Examining the time series plot in Figure 7.1(a) we see a strong yearly seasonal component, but also evidence of a mean-variance relationship with possible overdispersion over time. Ignoring possible serial dependence, we remove the seasonal component by fitting a linear model with a yearly sinusoidal term. Figure 7.1(b) and (c) shows plots of the sample ACF and sample partial ACF up to lags 25 for the residuals, indicating that there is indeed time series dependence data after we account for this seasonal term. Using standard diagnostics for

Table 7.1: A comparison of different model selection and diagnostic criteria for different GINAR processes fit to the meningococcal disease series. All fitted processes contain a yearly seasonality component. Values in bold indicate the processes for which the AIC and RMSE values are the smallest.

GINAR process	AIC	RMSE	Forecast coverage	Ljung-Box P-value
Po-INAR(1)	1857.9	4.72	90.0%	< 0.0001
NB-INAR(1)	1814.9	4.68	96.5%	< 0.0001
Geom-INAR(1)	1825.2	4.65	93.3%	0.003
Po-INAR(2)	1837.7	4.62	90.4%	0.019
NB-INAR(2)	1800.1	4.77	95.5%	0.368
Geom-INAR(2)	1806.9	4.57	93.0%	0.331
Po-INAR(3)	1817.2	4.60	90.6%	0.172
NB-INAR(3)	1788.9	4.58	96.8%	0.370
Geom-INAR(3)	1799.3	4.55	92.3%	0.589
Po-INAR(4)	1818.5	4.58	88.3%	0.130
NB-INAR(4)	<b>1782.0</b>	4.56	96.8%	0.226
Geom-INAR(4)	1792.4	<b>4.53</b>	90.0%	0.455

autoregressive moving average processes [e.g. [Brockwell and Davis, 2016](#)] an autoregressive process of order 2 or 3 might be reasonable this time series. However, autoregressive processes are a poor approximation for count series, and fail to account for possible overdispersion in the series. Instead, we fit a number of different GINAR( $p$ ) processes to the series, accounting for a time-varying seasonality using the following model for the time-varying log innovation mean:

$$\log \mu_{\epsilon,t} = b_0 + b_1 \sin(2\pi t/52) + b_2 \cos(2\pi t/52).$$

Table 7.1 shows a summary of different GINAR( $p$ ) processes fit to the count series, where we let  $p = 1, 2, 3$ , and 4. We choose  $p = 4$  as the maximum order as that suggests

Table 7.2: Parameter estimates and 95% confidence intervals for the NB-INAR(4) model shown in Table 7.1.

Parameter	Estimate	Confidence Interval
$\alpha_1$	0.19	(0.05, 0.34)
$\alpha_2$	0.15	(-0.01, 0.30)
$\alpha_3$	0.11	(-0.05, 0.26)
$\alpha_4$	0.04	(-0.17, 0.20)
$\beta_0$	1.88	(1.60, 2.15)
$\beta_1$	0.34	(0.24, 0.44)
$\beta_2$	0.24	(0.13, 0.35)
$r$	0.15	(0.04, 0.26)

a maximum of a four week or approximately one month dependence. As in the simulation study, we consider Po-INAR processes (binomial thinning with Poisson innovations), NB-INAR processes (binomial thinning with negative binomial innovations), and Geom-INAR processes (negative binomial thinning with Poisson innovations). We fit each process using the CML method, with the Davies method to calculate the transition probabilities. For each process we tabulate the AIC, root mean squared error, one step ahead forecast coverage (using the Monte Carlo method described in Section 3.3 and averaged over all time points), and the P-value of Ljung-Box test to examine whether or not the estimated innovations are a sample of IID noise (we use 20 lags for each test).

Table 7.1 indicates that the NB-INAR(4) process is preferred with reference to minimizing the AIC and achieving good forecast coverage, whereas the Geom-INAR(4) process yields a slightly smaller RMSE value. On balance the NB-INAR(4) process is preferable. All these processes give a better goodness of fit relative to using processes of order 3, as Pedeli et al.

[2015] considered – it is worth noting however that they preferred the NB-INAR process as we do. In Table 7.2 we show the parameter estimates and 95% confidence intervals for this NB-INAR(4) model. We see that among the dependence parameters only  $\alpha_1$  does not have zero in the confidence interval. The overdispersion parameter,  $r$ , does not have zero in the 95% confidence interval implying that it is significant and that a negative binomial assumption on the innovation sequence is appropriate. The intervals for  $\alpha_2$ ,  $\alpha_3$ , and  $\alpha_4$  contain zero, implying that the model could possibly be simplified to have a smaller model order, however in terms of AIC this NB-INAR(4) model is preferred. Considering GINAR( $p$ ) processes with  $p = 5, 6, 7$  and  $8$  did not greatly improve the fit and seemed to suggest we were overfitting to the data – the confidence intervals for model parameters were too wide, as compared to using simpler processes. In summary, we learn that meningococcal disease cases in Germany from 2001–2006 exhibit strong dependencies over four-weekly (monthly) and yearly scales, and that there is significant overdispersion that should be accounted for.

We next consider weekly meningococcal disease cases from 2001–2017. We look at a longer time series to study long-term changes in the weekly counts of this disease. A TV-GINAR( $p$ ) process with both the dependence and innovation mean parameter time-varying could potentially better capture these long-term changes. Figure 7.2 shows a summary of a time series of  $n = 887$  weekly meningococcal disease cases in Germany, over the period 2001–2017. Examining this time series plot in Figure 7.2(a) we see evidence of a mean-variance relationship and also a decaying number of weekly counts - the peaks keep getting smaller, with a downward trend overall. Figure 7.2(b) and (c) show the sample ACF and PACF for 25 lags for the residuals after we account for seasonality and trend, indicating that there is indeed time series dependence. These plots also suggest that an autoregressive process of

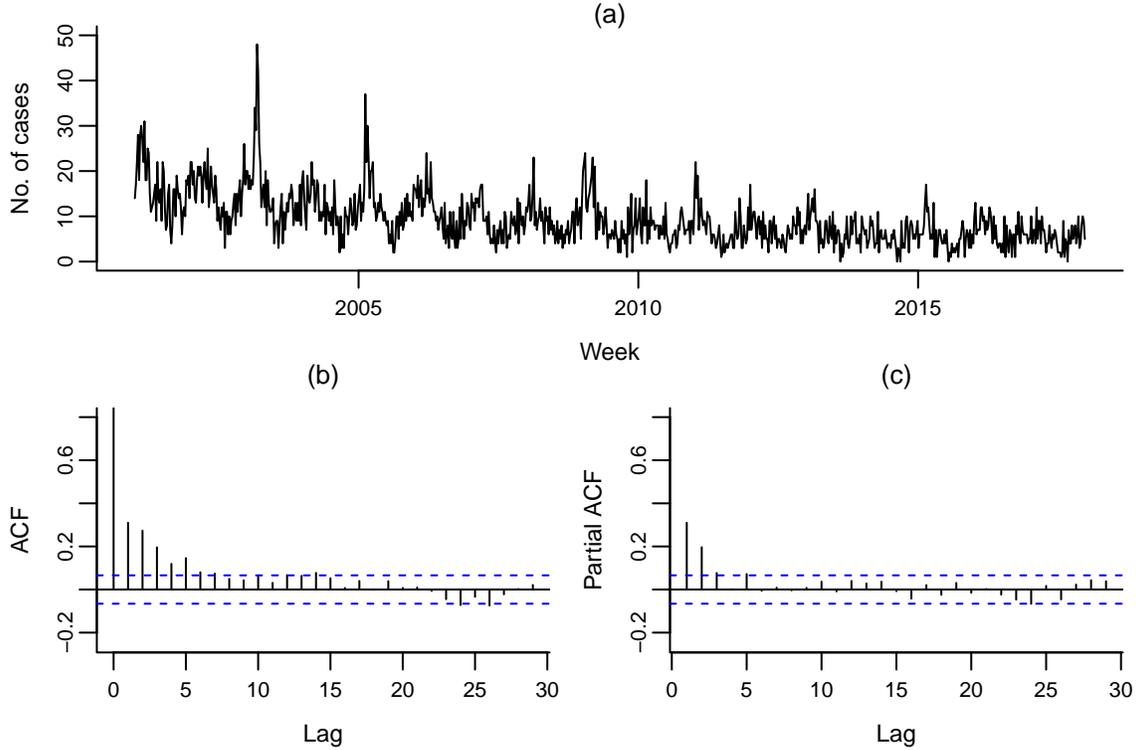


Figure 7.2: (a) Time series plot of the weekly meningococcal disease cases; (b) sample ACF and (c) sample PACF of residuals after removing a yearly sinusoidal seasonality and trend term.

order 2 or 3 might be reasonable. Although we are not using autoregressive processes these plots give us a good sense of potential model orders to consider.

Furthermore, we only consider processes with binomial thinning and negative binomial innovations since that was preferred for the shorter time series of these meningococcal disease counts, and because some overdispersion is evident through Figure 7.2(a). For all models considered we account for the time-varying seasonality and trend using the following model for the time-varying log innovation mean:

$$\log \mu_{\xi,t} = b_0 + b_1 \sin(2\pi t/52) + b_2 \cos(2\pi t/52) + b_3 t. \quad (7.1)$$

For the time-varying dependence parameter, we consider two variations. The first uses a polynomial basis and the second uses a b-spline basis function. A polynomial basis gives us

Table 7.3: A comparison of different model selection and diagnostic criteria for different GINAR processes fit to the meningococcal disease series. All fitted processes contain a yearly seasonality component and trend. Values in bold indicate the processes for which the AIC and RMSE values are the smallest.

Process	AIC	RMSE	Forecast coverage	Ljung-Box P-value
NB-INAR(2)	4620.9	3.65	96.4%	0.562
TV-NBINAR(2)-Poly	4616.0	3.64	96.4%	0.554
TV-NBINAR(2)-Basis	4617.4	3.62	96.2%	0.550
NB-INAR(3)	4612.6	3.64	96.7%	0.401
TV-NBINAR(3)-Poly	<b>4608.1</b>	<b>3.60</b>	96.2%	0.674
TV-NBINAR(3)-Basis	4612.6	3.61	96.5%	0.660

a flexible model to study non-linear relationships. Higher-degree polynomials could fit more complex data. While a b-spline basis allows even greater flexibility and can better fit data with varying local behavior.

The polynomial basis used for the models presented here is of degree two as follows:

$$\alpha_j(u) = \frac{\exp(\eta_{j,0} + \eta_{j,1}u + \eta_{j,2}u^2)}{1 + \sum_{k=1}^p \exp(\eta_{k,0} + \eta_{k,1}u + \eta_{k,2}u^2)}, \quad u \in [0, 1], \quad (7.2)$$

for  $j = 1, 2, \dots, p$ . Note that we tried higher-order polynomials however those models had larger AIC values and we do not consider them for this application. For the second case of the time-varying dependence parameter curve, we consider a cubic b-spline design matrix with three degrees of freedom. The degrees of freedom were again chosen using the AIC.

Table 7.3 shows a summary of the different models fit to this dataset. We consider  $p = 2$  and 3, and consider processes with binomial thinning and negative binomial innovations. We fit each process using the CML method. For each process we show the AIC, root mean squared error, one step ahead forecast coverage, and P-value of Ljung-Box test (we use 20

lags for each test). In Table 7.3, TV-NBINAR(3)-Poly refers to a time-varying NB-INAR(3) process with the polynomial basis for the dependence parameter curve, and TV-NBINAR(3)-Basis refers to a time-varying NB-INAR(3) process with the b-spline basis for the dependence parameter curve. The models NB-INAR(2) and NB-INAR(3) have constant dependence parameters with a time-varying innovation mean as described in (7.1). The results indicate that the TV-NBINAR(3)-Poly model is preferred with reference to minimizing the AIC, lowest RMSE, and achieving good forecast coverage. All processes considered do not have issues with goodness of fit, as shown by the Ljung-Box test P-values. Note that we did not show models of order  $p = 4$  because the AIC was almost the same as that of TV-NBINAR(3)-Poly (less than a point difference), and the estimate for  $\{\alpha_4(u) : u \in [0, 1]\}$  was approximately zero for both the time-varying and constant dependence parameter models.

Figure 7.3 shows the dependence parameter curves for the TV-NBINAR(3)-Poly model along with the 95% confidence intervals calculated using the method described in Section 6.2, along with the estimates of the NB-INAR(3) model. We see evidence that the  $\{\alpha_1(u) : u \in [0, 1]\}$  curve may be decreasing with time, and its confidence interval gets narrower for later time points. The  $\{\alpha_2(u) : u \in [0, 1]\}$  parameter curve also seems to be decreasing but has much larger confidence interval bands. This possibly suggests that this parameter may not benefit from being time-varying, and is constant as the larger confidence interval widths suggest that there is a lot of uncertainty in this relationship over time. This makes it hard to conclude whether or not you can tell apart a constant term from a time-varying curve. Also, the constant  $\alpha_2$  estimate from the NB-INAR(2) model is approximately equal to the  $\{\alpha_2(u) : u \in [0, 1]\}$  parameter curve for most time points, again suggesting that a constant model for this parameter could be more appropriate. The  $\{\alpha_3(u) : u \in [0, 1]\}$  parameter curve has a lot of uncertainty in the beginning and then is nearly zero, eventually increasing

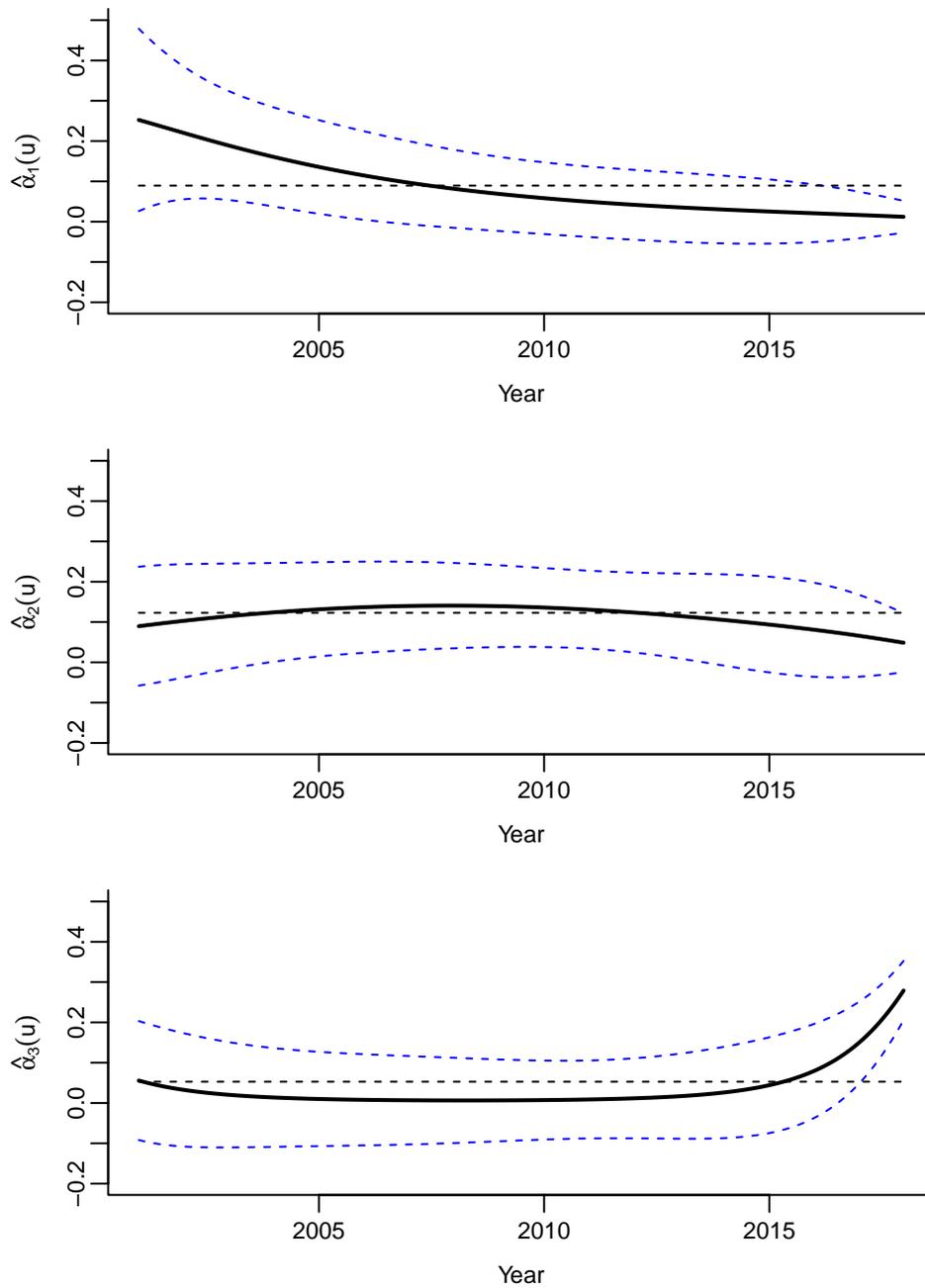


Figure 7.3: The estimated dependence parameter curves, with their 95% confidence intervals, for the TV-NBINAR(3)-Poly model fit to the meningococcal disease cases in Germany from 2001–2017. The dashed line represents the constant dependence parameter estimates from the NB-INAR(3) model.

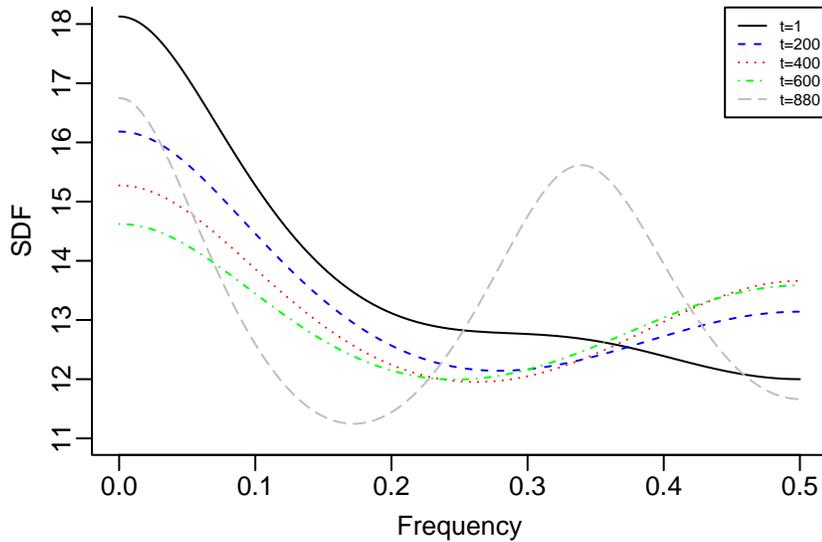


Figure 7.4: Spectral density for specified  $t$  values for TV-NBINAR(3)-Poly process not accounting for trend and seasonality.

sharply towards the last few weeks. This could suggest that the third-week lag becomes more prominent after 2015.

Furthermore, Figure 7.4 shows the spectral density of the TV-NBINAR(3)-Poly process for different time points; we see that spectral density at  $t = 880$  (one of the last time points in the series) is completely different from all other time points shown. This again suggests a shift in the dependence structure of the disease towards the last few weeks of the dataset, in particular, we see quasi-periodicity with period between 2 and 3 weeks. Lastly, Figure 7.5 shows the innovation mean parameter curve for the TV-NBINAR(3)-Poly model which is trying to capture the seasonality and trend; we also see that the confidence intervals get narrower as a function of time.

In summary, we learn that the meningococcal disease cases in Germany from 2001–2015 had some time-varying behavior and had strong dependencies over three-weekly and yearly scales. In the previous series from 2001–2006, we learned that the data showed

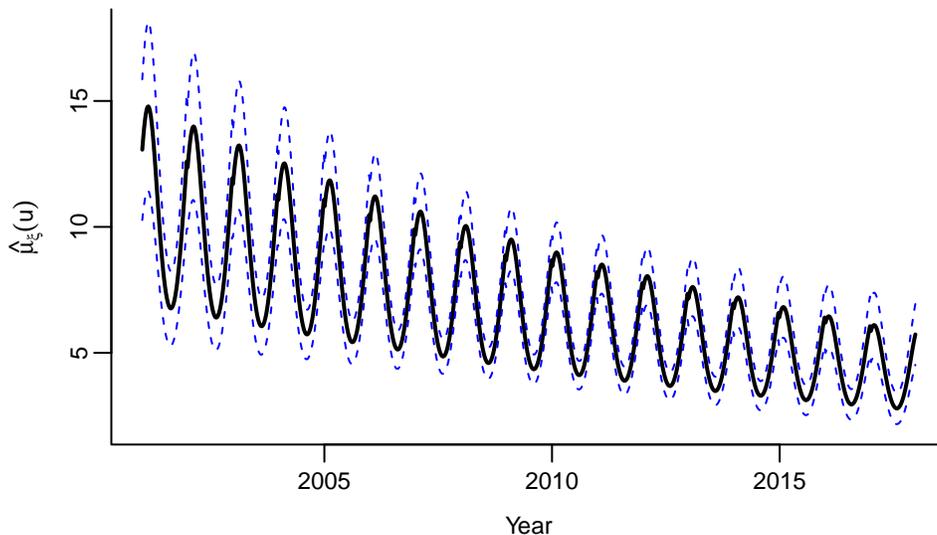


Figure 7.5: The estimated innovation mean parameter curve, with their 95% confidence intervals, for the TV-NBINAR(3)-Poly model fit to the meningococcal disease cases in Germany from 2001–2017.

strong dependencies over four-weekly scales. It is interesting that a model of order  $p = 4$  is not appropriate for the longer time series and perhaps suggesting a change in dependence structure as the rate of infections decreases.

## 7.2 Schizophrenic Patient Data

In this section, we consider a time series of perceptual speed scores for a schizophrenic patient. Modeling such time series can help diagnose the severity of a disorder and understand its progression. It can also help plan treatment and monitor how certain treatments affect perceptual speed. Also, accounting for a time series dependence parameter is important because a patient’s score is likely to be correlated to the score from previous days and the effects of certain medications also manifest over time. For these reasons, and given that the dataset is low-count, we consider  $\text{GINAR}(p)$  and  $\text{TV-GINAR}(p)$  processes to study the dataset. The time series dataset contains the daily observations ( $n = 120$ ) of the score

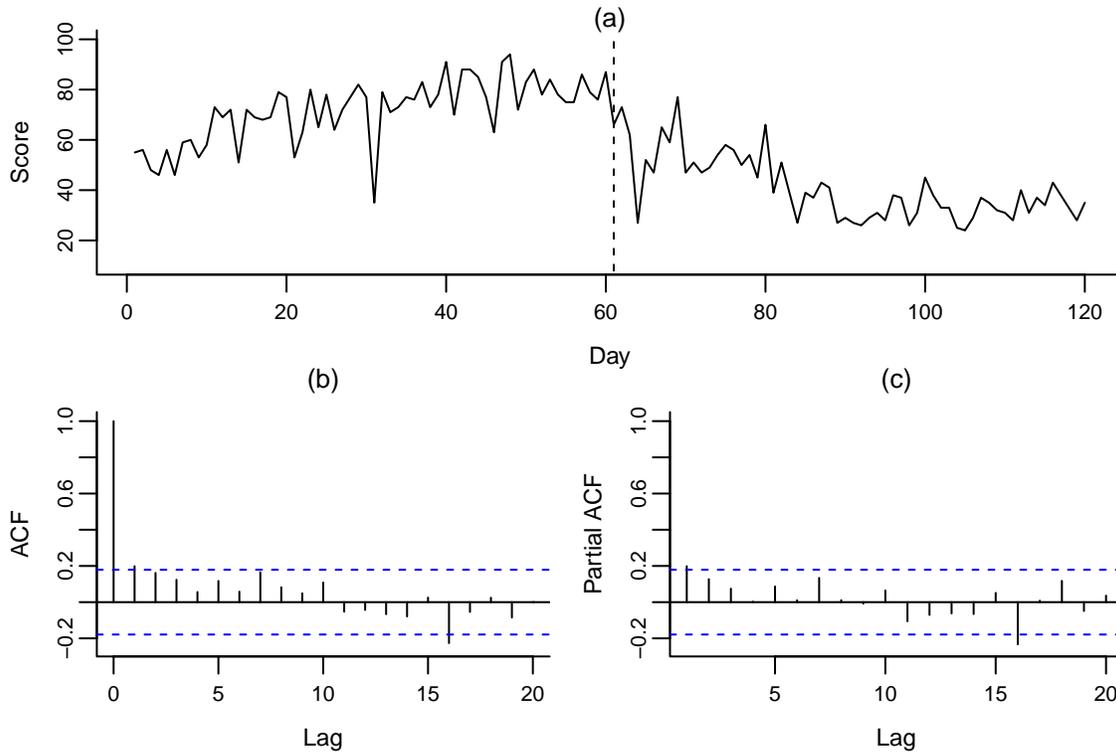


Figure 7.6: (a) Time series plot of the daily score achieved by the patient with a dashed line at day 61, when the patient received the treatment; (b) sample ACF and (c) sample PACF of residuals after removing the polynomial trend.

achieved by a schizophrenic patient on a test of perceptual speed [McCleary et al., 1980]. At day 61 the patient received a strong tranquilizer which could affect perceptual speed. This dataset was studied in Neal and Subba Rao [2007] and Kashikar et al. [2013].

Figure 7.6(a) shows a summary of the time series plot of this dataset, where we see a significant drop in scores around day 61. We fit a linear model to this series with a cubic b-spline design matrix with three degrees of freedom, ignoring possible serial dependence. The sample ACF and sample PACF up to 25 lags for the residuals are shown in Figure 7.6(b) and (c) respectively. These plots show that after accounting for the polynomial trend we still have time-series dependence, and from the plots an autoregressive process of order 1 or 2 seems appropriate. Also, note that both, Neal and Subba Rao [2007] and Kashikar et al.

[2013], used GINAR(1) processes. Hence, we consider processes of order  $p = 1$  only in this section. We also see that there is evidence of a mean-variance relationship – the series has a mean of 56.3 and a variance of 402.6. This overdispersion suggests that a process with negative binomial innovations is appropriate, and this is what we use in the models presented in this section. Also, we use binomial thinning as the thinning operator.

For the time-varying models, we consider processes where both the innovation mean and dependence parameter are time-varying and when only the innovation mean parameter is time-varying. For the innovation mean parameter curve, we consider a cubic b-spline design matrix with three and four degrees of freedom. For the time-varying dependence parameter curve, we again consider three variations. The first is a polynomial basis of degree two as shown in (7.2), the second is a cubic b-spline basis with four and six degrees of freedom, and the third is defined as follows:

$$\alpha_1(u) = \frac{\exp(\eta_{1,0} + \eta_{1,1} \mathbb{1}_{t \geq 61})}{1 + \exp(\eta_{1,0} + \eta_{1,1} \mathbb{1}_{t \geq 61})}, \quad u \in [0, 1].$$

This basis function has an indicator variable based on the day the patient received the treatment. Note that the choices of degrees of freedom were made based on the model AIC. We chose four degrees of freedom for the b-spline basis for the dependence parameter curve because using that we obtained a lower AIC as compared to two, three, and five degrees of freedom. Similarly, for the polynomial basis, we considered degrees two, three, and four, and a polynomial basis of degree two had the lowest AIC.

In Table 7.4 we present six models, each fit using the CML method. For each process we show AIC, root mean squared error, one step ahead forecast, and P-value of Ljung-Box test (20 lags for each test). For all models shown in Table 7.4 we assume a constant overdispersion parameter. The NB-INAR(1) model refers to a stationary GINAR(1) process with binomial thinning and negative binomial innovations. This is the only stationary model considered and

Table 7.4: A comparison of different model selection and diagnostic criteria for different GINAR processes fit to the perceptual score series. Values in bold indicate the processes for which the AIC and RMSE values are the smallest.

Process	AIC	RMSE	Forecast coverage	Ljung-Box P-value
NB-INAR(1)	957.0	12.48	94.1%	< 0.0001
TV-NBINAR(1)-1	880.8	9.60	95.8%	0.290
TV-NBINAR(1)-2	<b>870.4</b>	9.10	96.6%	0.494
TV-NBINAR(1)-Poly	872.7	9.22	96.6%	0.485
TV-NBINAR(1)-Basis	870.5	<b>9.00</b>	95.8%	0.381
TV-NBINAR(1)-Ind	878.5	9.37	95.8%	0.365
TV-NBINAR(1)-Basis-6	874.6	<b>9.00</b>	95.8%	0.381

from the results in Table 7.4 we see that it has the largest AIC, RMSE, and significant Ljung-Box p-value and is therefore not a good model for this dataset. The TV-NBINAR(1)-Poly, TV-NBINAR(1)-Basis, TV-NBINAR(1)-Basis-6, and TV-NBINAR(1)-Ind models assume a time-varying dependence parameter curve, modeled via a polynomial basis of degree two, cubic b-spline basis of degree four, cubic b-spline basis of degree six, and a basis with an indicator variable for the treatment period (day 61 onwards) respectively. These four models have a time-varying innovation mean parameter modeled via a cubic b-spline basis function of degree three. As mentioned earlier, choices of degrees of freedom for the b-spline basis and the degree of the polynomial basis were based on the AIC. Although TV-NBINAR(1)-Basis-6 did not have the lowest AIC, it was included because a cubic b-spline with six degrees of freedom was able to capture the period when the patient received the treatment. Models with increasing either of these hyperparameters saw increasing AIC as well. Lastly, the TV-NBINAR(1)-1 and TV-NBINAR(1)-2 models assume a constant dependence parameter

and a time-varying innovation mean parameter curve. The TV-NBINAR(1)-1 and TV-NBINAR(1)-2 processes model the time-varying innovation mean using a cubic b-spline basis with three and four degrees of freedom respectively (at five degrees of freedom the AIC started to increase hence we did not consider higher degrees). We see that both TV-NBINAR(1)-2 and TV-NBINAR(1)-Basis are almost identical in their performance and can both be considered the best models out of all the ones considered. Both these models have the lowest AIC and RMSE, preferred goodness of fit measures, and forecast coverages.

Let us consider the dependence parameter curve of the TV-NBINAR(1)-Basis, TV-NBINAR(1)-Basis-6, TV-NBINAR(1)-Poly, TV-NBINAR(1)-Ind, and TV-NBINAR(1)-2 models shown in Figure 7.7. The TV-NBINAR(1)-Basis and TV-NBINAR(1)-Poly models both suggest similar behavior of the dependence parameter curve as seen in Figure 7.7(a) and (c) – dependence on the past lag seems to be nearly 0 with high confidence between days 40–90, as suggested by the narrow confidence intervals. The dependence seems to be decreasing until day 40 and then increasing after day 90. Recall that day 61 is when the patient received the tranquilizer, so having a time-series dependence of approximately 0 for some days following the treatment and then a sudden spike can give insight into the effectiveness of the treatment. Also, Figure 7.9 shows the spectral density for the preferred TV-NBINAR(1)-Basis model at different time points, depicting the change in spectral properties over time. We see how the spectral density flattens for days  $t = 30$  suggesting a white noise process. Figure 7.7(g) shows the dependence parameter curve for the TV-NBINAR(1)-Ind model and we clearly see a drop in the dependence parameter curve after the patient received the treatment at day 61. Figure 7.8(a) shows the dependence parameter curve for the TV-NBINAR(1)-Basis-6 model which seems to start gradually increasing around day

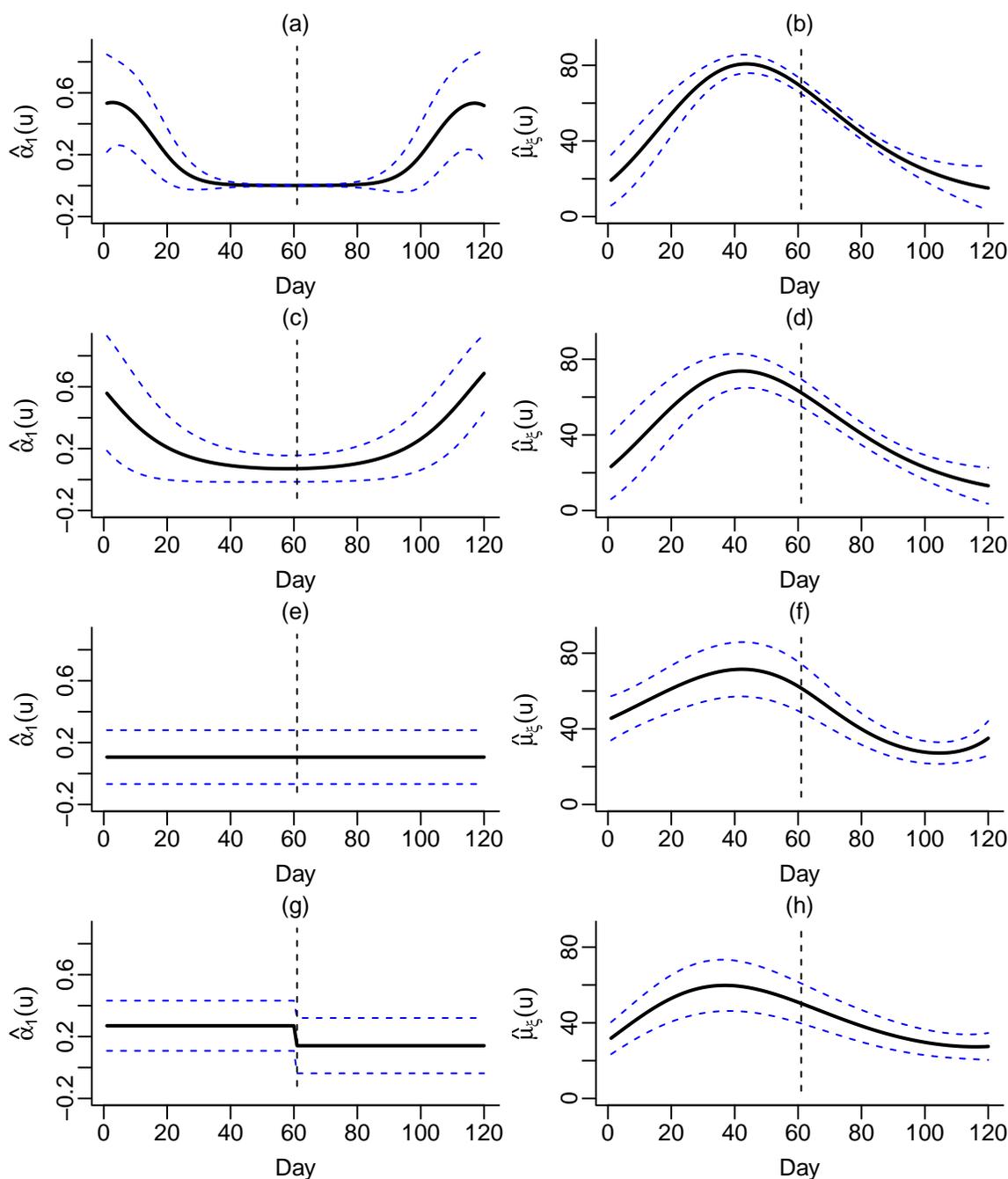


Figure 7.7: (a), (b) The estimated  $\alpha_1(u)$  and  $\mu_\xi(u)$  parameter curve for the TV-NBINAR(1)-Basis model respectively; (c), (d) The estimated  $\alpha_1(u)$  and  $\mu_\xi(u)$  parameter curve for the TV-NBINAR(1)-Poly model respectively ; (e), (f) The estimated  $\alpha_1(u)$  and  $\mu_\xi(u)$  parameter curve for the TV-NBINAR(1)-2 model respectively; (g), (h) The estimated  $\alpha_1(u)$  and  $\mu_\xi(u)$  parameter curve for the TV-NBINAR(1)-Ind model respectively.

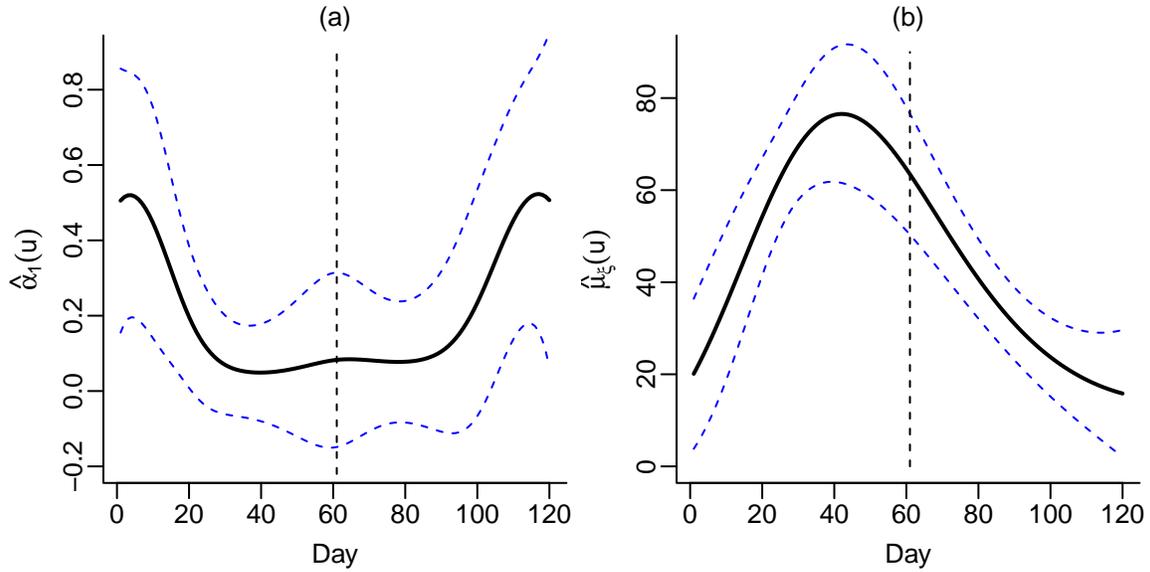


Figure 7.8: (a), (b) The estimated  $\alpha_1(u)$  and  $\mu_\xi(u)$  parameter curve for the TV-NBINAR(1)-Basis-6 model respectively.

80, after the treatment is administered. Figure 7.7(b) and (f) show the time-varying innovation mean for the TV-NBINAR(1)-Basis and TV-NBINAR(1)-2 models. We see that TV-NBINAR(1)-Basis has a much narrower confidence interval suggesting a better fit as compared to TV-NBINAR(1)-2. For all innovation mean parameter curves shown in Figure 7.7 we see that the innovation mean goes down after the treatment at day 61 indicating a drop in scores on average.

In summary, we learn that the perceptual speed scores for a schizophrenic patient had some time-varying behavior and strong dependencies over one-day periods. We see the merits of using a time-varying GINAR( $p$ ) process as compared to the stationary GINAR( $p$ ) process as all the time-varying models had much better performance in terms of AIC, RMSE, forecast coverage, and goodness of fit measures. We are also able to use these parameter curves for further inference. For instance, the dependence parameter curves for this application can be useful in understanding the effect of the tranquilizer or other dynamics about the patient.

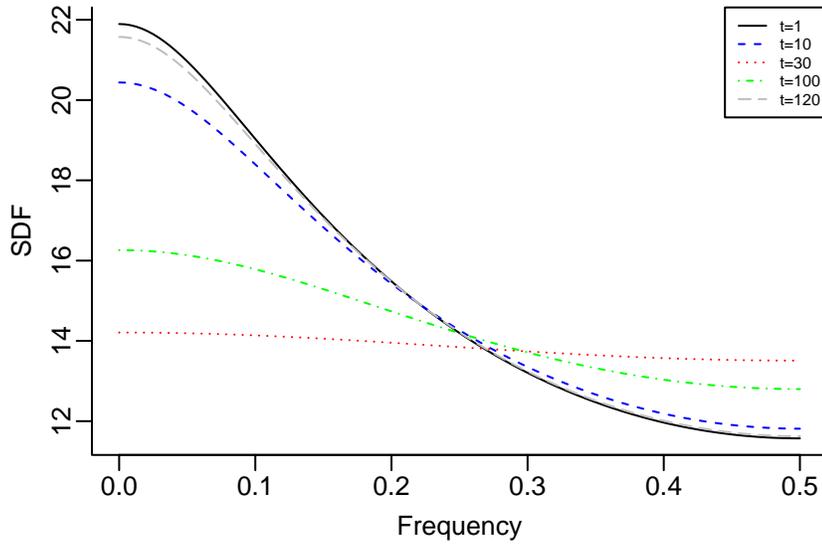


Figure 7.9: Spectral density for specified days values for TV-NBINAR(1)-Basis process not accounting for polynomial trend.

### 7.3 Conclusions

In this chapter we explored two applications of  $\text{GINAR}(p)$  and  $\text{TV-GINAR}(p)$  processes – a disease surveillance series and a series pertaining to a schizophrenic patient’s perceptual speed scores. Both series had characteristics like low-count values and overdispersion which violated the Gaussianity assumption. Hence, models like  $\text{GINAR}(p)$  and  $\text{TV-GINAR}(p)$  are more appropriate and also respect the integer nature of the data.

We explored many models for the disease surveillance series, with varying thinning operators and innovation distributions, eventually deciding on a  $\text{NB-INAR}(4)$  model as the best one. From this, we also see how  $\text{GINAR}(p)$  models are flexible enough to allow for different data structures simply by changing the thinning operator and/or innovation distribution. Furthermore, we also showed the merits of the new  $\text{TV-GINAR}(p)$  process for a longer series of meningococcal disease cases which is better able to capture time-varying trends in the dataset. We also see that models that have both the dependence and innovation

mean parameter curve time-varying perform better than models where only the innovation mean parameter curve is time-varying, suggesting that the model benefits from having a time-varying dependence structure.

In the second application studying the time series of perceptual speed scores for a schizophrenic patient we explored a range of stationary  $\text{GINAR}(p)$  and  $\text{TV-GINAR}(p)$  models as well. The patient received a treatment at day 61 leading to changes in the perceptual speed scores after the treatment. Therefore, a stationary  $\text{GINAR}(p)$  model does not perform well compared to the time-varying models. The time-varying models show how the dependence structure of the process changes with time which can particularly be useful in understanding the effects of a treatment.

In conclusion, through these applications, we are able to see the merits of the time-varying process which gives us much more flexibility in modeling different non-stationary count time series processes.

## Chapter 8: Discussion and Future Work

### 8.1 Discussion

In this dissertation, we focused on the modeling of discrete-values time series processes. In particular, we focused our attention on generalized integer autoregressive processes. While these processes share many characteristics with non-integer autoregressive processes, such as being a Markov process of a given order, through the selection of the thinning operator and innovation, we can accurately model positively valued count series with a variety of distributional and dispersion properties.

We first introduced these processes in Chapter 2 and provided a literature review of different examples of these processes, demonstrating their flexibility for a wide range of applications. We further derived the statistical properties and presented two methods for calculating the transition probabilities.

In Chapter 3 we presented various estimation methods that have been used for specific examples of these  $\text{GINAR}(p)$  processes and extended them to the class of  $\text{GINAR}(p)$  processes; we also proposed a simple estimation method called the pseudo maximum likelihood which approximates the transition probabilities using a normal distribution. We extended the asymptotic results for different estimators of the process parameters to apply to  $\text{GINAR}$  processes, not just  $\text{INAR}$  processes defined with binomial thinning, as is commonplace in the

literature. Note that we do not have asymptotic results for the saddlepoint approximation technique or the Whittle likelihood estimate which can be a potential area of future research. We studied the small sample performance of these estimation methods in Chapter 4, our goal was to understand the performance of these estimation methods under varying conditions. In the literature, the performance for some of these estimation methods is provided for certain examples only – for example, processes with binomial thinning and/or Poisson innovation, low model orders, or not a wide range of parameter values. We presented simulations for different thinning operators and innovation distributions, varying parameter values, and higher model orders. The main takeaway from these simulations was that CML performs well over all other methods, even when changing the model order  $p$ , the values of the parameters, thinning operation, and innovation distribution. However, it is the most computationally intensive. If further speed increases are desired we can refer to the simulations as a guide to decide which estimation method to use. For instance, a comparable estimation method to CML is pseudo maximum likelihood when the innovation distribution is Poisson, but not so much when the innovation distribution is negative binomial. We also presented coverage simulations to better understand the performance of these estimation methods.

GINAR( $p$ ) processes assume stationarity, which is not always possible or applicable in real-world settings. Hence, we turned our attention to modeling non-stationary discrete-valued time series for which we introduced the time-varying generalized integer autoregressive process (TV-GINAR( $p$ )) in Chapter 5. We allowed the parameter curves to vary with time via generalized functions of basis vectors. This gives us more flexibility in modeling different types of non-stationary processes. We derived the statistical properties and transition probabilities of this process. In Chapter 6 we showed how to extend three estimation methods, including CML, from Chapter 3 to the TV-GINAR( $p$ ) process. The biggest challenge

here is the asymptotic theory for these estimation methods, which is no longer straightforward without the assumptions of stationarity and ergodicity. Hence, we showed how to derive the confidence intervals using the Delta method assuming asymptotic normality of the estimators. We also presented simulation studies to study finite sample performance of the estimation methods and consistent with the GINAR( $p$ ) simulations, we see that the CML method outperforms both the pseudo likelihood and CLS estimation methods. We further provided a discussion of model selection techniques and showed that the AIC could be used as a metric to decide whether a process is TV-GINAR( $p$ ) or not.

Lastly, in Chapter 7 we presented two applications, a disease surveillance series, and a schizophrenic patient scores dataset, to demonstrate the merits of GINAR( $p$ ) and TV-GINAR( $p$ ) processes. We particularly showed the advantages of using a time-varying process as TV-GINAR( $p$ ) processes outperform stationary GINAR( $p$ ) processes in both applications.

## 8.2 Future Work

In this section we provide a discussion of areas of future research. In particular we discuss forecasting for TV-GINAR( $p$ ) processes, extensions of this dissertation to multivariate processes, and model selection.

### 8.2.1 Forecasting

An area of future research is forecasting. For the TV-GINAR( $p$ ) process, we have not spent much time discussing forecasting strategies. The techniques described for the GINAR( $p$ ) process do not apply naturally to the TV-GINAR( $p$ ) situation because we have time-varying parameters and do not necessarily know the future parameter values. In the stationary case, forecasting has been extensively studied as discussed in Section 3.3. Forecasting for non-stationary time series has been discussed in [Whittle \[1965\]](#), [Abdrabbo and Priestley](#)

[1967], and Fryzlewicz et al. [2003] among others. (Although these do not specifically deal with GINAR processes or discrete-valued time series.)

Hence, we need a way of determining or approximating the future parameters which can then be used for forecasting. Some naive approaches could be to use the parameter values at time  $T$  for parameter values at  $T + 1, T + 2, \dots, T + h$ , where  $h$  is the forecast horizon. We could also use the mean or median of the parameters at all time points. These approaches are simple, however they do not take into account the time-varying nature of the parameters.

Another possible strategy is local modeling of the parameter estimates. For example, one possibility is to choose a time window for the parameter estimates,  $W$ , and then fit a low order polynomial model to the parameter values. The results can then be used to forecast future parameter values for the forecast horizon  $h$ . We can use cross-validation strategies or AIC for choosing the window length. A drawback of this approach is that for more complex trends this methodology might be too simplistic. This can be investigated further in future research.

Palma et al. [2013] suggest redefining the sample size to  $T^* = T + h$ , where  $h$  is the desired forecast horizon. The observations,  $T + 1, T + 2, \dots, T^*$  can then be treated as missing values. They propose using the Kalman filter equations, but for our purposes we can use the forecasting methodology defined in Section 3.3 or the minimum variance predictor.

## 8.2.2 Time-varying parameter curves

In this dissertation, we have carried out model selection for TV-GINAR( $p$ ) processes using AIC, which has been proposed as a viable model selection technique for GINAR( $p$ ) processes (e.g. Weiß and Feld [2020]) and locally stationary processes (e.g. Dahlhaus [1996] and Dahlhaus [2012]). Dahlhaus [1996] also propose a non-stationary information criteria

defined as follows:

$$NIC(k) = -\frac{1}{T}\ell_T(\hat{\eta}_T) + \frac{k}{T},$$

where  $\ell_T(\hat{\eta}_T)$  is the likelihood function evaluated at the estimated parameter values for a sample of size  $T$ , and  $k$  is the total number of parameters in the model. Note that the TV-GINAR( $p$ ) process does not satisfy all the assumptions described in [Dahlhaus \[1996\]](#) to derive this NIC criterion. However, this can be a starting point to develop similar methodology for the TV-GINAR( $p$ ) process.

In Chapter 7 when introducing time-varying parameter curves for the application we allow all the dependence parameter curves to vary. For example, when the order of the model was  $p = 4$  we let all four dependence parameter curves have the same time-varying structure, i.e. same basis function definition, for simplicity. This may not always be appropriate as each dependence parameter could have a different time-varying structure with varying basis functions. For instance, in [Figure 7.3](#), it seems as though  $\{\alpha_2(u) : u \in [0, 1]\}$  does not change much over time and could have been modeled using a simple intercept only basis function. This is an observation after we have already fit the model and looked at the parameter curve estimates. We could also perform exploratory analysis to determine appropriate basis functions for the parameter curves, for example, calculating windowed statistics for the process, such as sample autocorrelation or partial autocorrelation. However, the potential issue here is determining the time window length and we would have to experiment with a few different values. All these methods rely on exploratory analysis which may not always be feasible for larger model orders or more model variations to consider. Hence, another methodology is employing grid search techniques over a range of basis functions for each parameter curve and choose the basis function and/or model order which minimizes the AIC. However, this does have the disadvantage of being computationally intensive.

Another promising method for model selection for these parameter curves is penalized regression. [Lin and Guo \[2020\]](#) introduce lasso for model selection and estimation in the GINAR( $p$ ) model with binomial thinning and Poisson innovation. They argue that adaptive lasso, introduced by [Zou \[2006\]](#), is more suitable. They propose getting the estimate of the  $\boldsymbol{\theta} = (\alpha_1, \alpha_2, \dots, \alpha_p, \mu_\epsilon)^T$  through the following optimization:

$$\hat{\boldsymbol{\theta}} = \arg \max \left( -\frac{1}{n} \ell(\boldsymbol{\theta}) + \lambda \sum_{j=1}^p |\theta_j| \tau_j \right),$$

where  $\ell(\boldsymbol{\theta})$  is the conditional log likelihood of the process, and  $\lambda$  is a tuning parameter that can be chosen through AIC or cross-validation. [Lin and Guo \[2020\]](#) discuss different choices of  $\tau_j$ . Another method was proposed by [Wang \[2020\]](#) which uses a penalized conditional least squares approach for the GINAR(1) process with binomial thinning and Poisson innovations and with covariates. They consider two penalty functions for the estimation, and show how this aids in deleting insignificant covariates. Considering such extensions for the TV-GINAR( $p$ ) process would be interesting, as it could help solve the issue of determining what basis function is suitable for which parameter curve and has a relatively simple implementation.

### 8.2.3 Multivariate Processes and Other Extensions

In this dissertation, we only considered univariate GINAR( $p$ ) processes. A natural extension is multivariate GINAR( $p$ ) and TV-GINAR( $p$ ) processes. Multivariate processes are useful when we observe several related features simultaneously over time. The main difference when discussing multivariate processes lies in the definition of a multivariate thinning operator. The most widely used approach is matrix-binomial thinning introduced by [Franke and Rao Subba \[1995\]](#). A generalized version of this operator which uses generalized thinning (rather than binomial thinning) was introduced by [Latour \[1997\]](#). Other definitions

of the multivariate INAR process have been considered, e.g. [Heinen and Rengifo \[2007\]](#), [Pedeli and Karlis \[2011\]](#), [Pedeli and Karlis \[2013\]](#), among others. Given that we have used [Latour \[1998\]](#)'s definition of the GINAR( $p$ ) process in this dissertation, we present their representation of the multivariate GINAR( $p$ ) process as follows:

**Definition 8.2.1** *Let  $\mathbf{X}$  be a non-negative integer-valued random  $r$ -vector, and let  $\mathbf{A}$  be an  $r \times r$  matrix of the generalized thinning operators with independent elements  $\{\alpha_{j,k}\}_{1 \leq j,k \leq r}$ . Then the MGINAR( $p$ ) process can be defined as*

$$\mathbf{X}_t = \sum_{j=1}^p \mathbf{A}_j \odot \mathbf{X}_{t-j} + \boldsymbol{\epsilon}_t, \quad t \in \mathbb{Z}$$

where the elementwise thinning operation is defined as follows:

$$\mathbf{A} \odot \mathbf{X} = \mathbf{A} \odot \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_r \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^r \alpha_{1,k} \odot X_k \\ \sum_{k=1}^r \alpha_{2,k} \odot X_k \\ \vdots \\ \sum_{k=1}^r \alpha_{r,k} \odot X_k \end{bmatrix}.$$

Let  $\{\mathbf{A}_j \odot\}$  for  $j = 1, 2, \dots, p$ , be  $p$  mutually independent operators and the innovation process  $\{\boldsymbol{\epsilon}_t : t \in \mathbb{Z}\}$  be a set of IID non-negative integer-valued RVs with finite mean vector  $\boldsymbol{\mu}_\epsilon = (\mu_{1,\epsilon}, \mu_{2,\epsilon}, \dots, \mu_{t,\epsilon})^T$  and finite variance vector  $\boldsymbol{\Sigma}_\epsilon = (\sigma_{1,\epsilon}^2, \sigma_{2,\epsilon}^2, \dots, \sigma_{t,\epsilon}^2)^T$ . Also,  $\{\boldsymbol{\epsilon}_t : t \in \mathbb{Z}\}$  is assumed independent of the thinning operators. A strictly stationary solution exists when  $\det(\mathbf{I} - \mathbf{A}_1 z - \dots - \mathbf{A}_p z^p) = 0$  has all its roots outside the unit circle.

It would be interesting to consider this MGINAR( $p$ ) process and its parameter estimation for future research, like we did in this dissertation for the GINAR( $p$ ) process. It is important to note that likelihood theory even for the univariate case is cumbersome, and even more so for the multivariate case leading to prediction and estimation challenges [[Davis et al., 2021](#)].

Future research can consider extensions of the MGINAR( $p$ ) process to the TV-MGINAR( $p$ ) process. In particular, studying estimation and forecasting procedures for the

TV-MGINAR( $p$ ) process and considering applications. Multivariate INAR processes models have already been used for various applications ranging from modeling hurricane and earthquake counts to mutual fund purchase and redemption order counts (e.g., [Boudreault and Charpentier \[2011\]](#), [Livsey et al. \[2018\]](#), [Darolles et al. \[2019\]](#)). Non-stationarity is prevalent in these applications, and hence research in non-stationary models will be relevant. Some recent research in the field has already looked at incorporating non-stationarity into these multivariate INAR processes, for example, [Bakouch et al. \[2020\]](#), [Santos et al. \[2021\]](#), [Yang et al. \[2023\]](#), [Sunecher et al. \[2024\]](#). Hence, it would be worthwhile to see how a time-varying multivariate GINAR process, using the time-varying structure defined in this dissertation, might perform in comparison to the models introduced in the aforementioned research.

Other extensions to the GINAR( $p$ ) process can also be considered for future research. Inspired by time series processes for non-count series, we can introduce dependence in the innovations or in the variance structure to obtain the INARMA and INGARCH processes, respectively [see e.g., [Weiß, 2018](#), Chapters 3 and 4, for definitions]. Due to the extra latent structure that is introduced to define these processes, estimating parameters is more involved. It would be interesting to learn how the different estimation and forecasting methods perform, especially with other thinning operators and more complicated model orders.

## Bibliography

- N. Abdrabbo and M. Priestley. On the prediction of non-stationary processes. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 29:570–585, 1967.
- M. A. Al-Osh and E.-E. A. Aly. First order autoregressive time series with negative binomial and geometric marginals. *Communications in Statistics-Theory and Methods*, 21:2483–2492, 1992.
- M. A. Al-Osh and A. A. Alzaid. First-order integer-valued autoregressive (INAR(1)) process. *Journal of Time Series Analysis*, 8:261–275, 1987.
- M. S. Aleksić and M. M. Ristić. A geometric minification integer-valued autoregressive model. *Applied Mathematical Modelling*, 90:265–280, 2021.
- N. Alzahrani, P. Neal, S. E. Spencer, T. J. McKinley, and P. Touloupou. Model selection for time series of count data. *Computational Statistics & Data Analysis*, 122:33–44, 2018.
- A. Alzaid and M. Al-Osh. An integer-valued  $p$ th-order autoregressive structure (INAR( $p$ )) process. *Journal of Applied Probability*, 27:314–324, 1990.
- T. Amemiya. *Advanced Econometrics*. Harvard University Press, Cambridge, Massachusetts, MA, 1985.
- D. W. K. Andrews. Consistency in nonlinear econometric models: A generic uniform law of large numbers. *Econometrica*, 55:1465–1471, 1987.
- H. S. Bakouch, Y. Sunecher, N. Mamode Khan, and V. Jowaheer. A non-stationary bivariate INAR(1) process with a simple cross-dependence: Estimation with some properties. *Australian & New Zealand Journal of Statistics*, 62:25–48, 2020.
- W. Barreto-Souza, S. Ndreca, R. B. Silva, and R. W. Silva. Non-linear INAR(1) processes under an alternative geometric thinning operator. *TEST*, 32:695—725, 2023.
- L. Bisaglia and A. Canale. Bayesian nonparametric forecasting for INAR models. *Computational Statistics & Data Analysis*, 100:70–78, 2016.
- L. Bisaglia and M. Gerolimetto. Model-based INAR bootstrap for forecasting INAR( $p$ ) models. *Computational Statistics*, 34:1815–1848, 2019.
- U. Böckenholt. Mixed INAR(1) Poisson regression models: Analyzing heterogeneity and serial dependencies in longitudinal count data. *Journal of Econometrics*, 89:317–338, 1998.
- P. Borges, F. F. Molinares, and M. Bourguignon. A geometric time series model with inflated-parameter bernoulli counting series. *Statistics & Probability Letters*, 119:264–272, 2016.
- M. Boudreault and A. Charpentier. Multivariate integer-valued autoregressive models applied to earthquake counts. *arXiv preprint arXiv:1112.0929*, 2011.

- M. Bourguignon and K. L. Vasconcellos. Improved estimation for Poisson INAR(1) models. *Journal of Statistical Computation and Simulation*, 85:2425–2441, 2015a.
- M. Bourguignon and K. L. P. Vasconcellos. First order non-negative integer valued autoregressive processes with power series innovations. *Brazilian Journal of Probability and Statistics*, 29:71 – 93, 2015b.
- M. Bourguignon, J. Rodrigues, and M. Santos-Neto. Extended Poisson INAR(1) processes with equidispersion, underdispersion and overdispersion. *Journal of Applied Statistics*, 46:101–118, 2019.
- K. Brannas. *Estimation and testing in integer-valued AR(1) models*. University of Umeå, Umeå, Sweden, 1993.
- K. Brannas. Explanatory variables in the AR(1) count data model. Technical report, Umea University, Umea Economics Studies 381, Umea, Sweden, 1995.
- P. Brockwell and R. Davis. *Introduction to Time Series and Forecasting (Third Edition)*. Springer, New York, NY, 2016.
- R. Bu and B. McCabe. Model selection, estimation and forecasting in INAR( $p$ ) models: A likelihood-based Markov chain approach. *International Journal of Forecasting*, 24:151–162, 2008.
- M. Cardinal, R. Roy, and J. Lambert. On the application of integer-valued time series models for the analysis of disease incidence. *Statistics in Medicine*, 18:2025–2039, 1999.
- I. M. M. da Silva. *Contributions to the analysis of discrete-valued time series*. PhD thesis, Universidade do Porto. Reitoria, Portugal, 2005.
- R. Dahlhaus. Maximum likelihood estimation and model selection for locally stationary processes. *Journal of Nonparametric Statistics*, 6:171–191, 1996.
- R. Dahlhaus. Locally stationary processes. In T. Subba Rao, S. Subba Rao, and C. Rao, editors, *Handbook of Statistics*, volume 30, pages 351–413. Elsevier, Heidelberg, Germany, 2012.
- H. E. Daniels. Saddlepoint approximations in statistics. *The Annals of Mathematical Statistics*, 25:631–650, 1954.
- S. Darolles, G. Le Fol, Y. Lu, and R. Sun. Bivariate integer-autoregressive process with an application to mutual fund flows. *Journal of Multivariate Analysis*, 173:181–203, 2019.
- R. B. Davies. Numerical inversion of a characteristic function. *Biometrika*, 60:415–417, 1973.
- R. A. Davis, K. Fokianos, S. H. Holan, H. Joe, J. Livsey, R. Lund, V. Pipiras, and N. Ravishanker. Count time series: A methodological review. *Journal of the American Statistical Association*, 116:1533–1547, 2021.
- A. C. Davison. *Statistical Models*. Cambridge University Press, Cambridge, England, 2003.
- V. Enciso-Mora, P. Neal, and T. S. Rao. Integer valued AR processes with explanatory variables. *Sankhyā: The Indian Journal of Statistics, Series B*, 71:248–263, 2009.
- M. Forughi, Z. Shishebor, and A. Zamani. Portmanteau tests for generalized integer-valued autoregressive time series models: Portmanteau tests for GINAR models. *Statistical Papers*, 63: 1163–1185, 2022.

- J. Franke and T. Rao Subba. Multivariate first-order integer-valued autoregressions. Technical report, Department of Mathematics, UMIST, Manchester, England, 1995.
- J. Franke and T. Seligmann. Conditional maximum likelihood estimates for INAR(1) processes and their application to modelling epileptic seizure counts. In T. Subba Rao, editor, *Developments in Time Series Analysis*, volume 55, pages 310–330. Chapman and Hall London, Boca Raton, FL, USA, 1993.
- R. K. Freeland. *Statistical analysis of discrete time series with application to the analysis of workers' compensation claims data*. PhD thesis, University of British Columbia, Vancouver, Canada, 1998.
- R. K. Freeland. True integer value time series. *AStA Advances in Statistical Analysis*, 94:217–229, 2010.
- R. K. Freeland and B. P. McCabe. Analysis of low count time series data by Poisson autoregression. *Journal of time series analysis*, 25:701–722, 2004a.
- R. K. Freeland and B. P. McCabe. Forecasting discrete valued low count time series. *International Journal of Forecasting*, 20:427–434, 2004b.
- P. Fryzlewicz, S. Van Bellegem, and R. Von Sachs. Forecasting non-stationary time series by wavelet process modelling. *Annals of the Institute of Statistical Mathematics*, 55:737–764, 2003.
- A. Graham. *Kronecker Products and Matrix Calculus: with applications*. Ellis Horwood, Mineola, New York, NY, 1981.
- K. Hadri. Maximum likelihood estimation of higher-order integer-valued autoregressive processes. *Journal of Time Series Analysis*, 30:259–259, 2009.
- A. Heinen and E. Rengifo. Multivariate autoregressive modeling of time series count data using copulas. *Journal of Empirical Finance*, 14:564–583, 2007.
- J. Huang and F. Zhu. A new first-order integer-valued autoregressive model with Bell innovations. *Entropy*, 23:713, 2021.
- M. A. Jazi, G. Jones, and C.-D. Lai. First-order integer valued AR processes with zero inflated poisson innovations. *Journal of Time Series Analysis*, 33:954–963, 2012.
- C. Jentsch and C. H. Weiß. Bootstrapping INAR models. *Bernoulli*, 25:2359–2408, 2019.
- D. Jin-Guan and L. Yuan. The integer-valued autoregressive (INAR( $p$ )) model. *Journal of Time Series Analysis*, 12:129–142, 1991.
- H. Joe. Time series models with univariate margins in the convolution-closed infinitely divisible class. *Journal of Applied Probability*, 33:664–677, 1996.
- H. Joe. Likelihood inference for generalized integer autoregressive time series models. *Econometrics*, 7:43, 2019.
- R. C. Jung and A. R. Tremayne. Coherent forecasting in integer time series models. *International Journal of Forecasting*, 22:223–238, 2006.
- R. C. Jung, G. Ronning, and A. R. Tremayne. Estimation in conditional first order autoregression with discrete support. *Statistical papers*, 46:195–224, 2005.
- H. Karlsen and D. Tjøstheim. Consistent estimates for the NEAR(2) and NLAR(2) time series models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 50:313–320,

1988.

- A. S. Kashikar, N. Rohan, and T. Ramanathan. Integer autoregressive models with structural breaks. *Journal of Applied Statistics*, 40:2653–2669, 2013.
- N. M. Khan, A. D. Soobhug, N. Youssef, S. Fedally, S. Nadarajah, and Z. Heetun. Re-visiting the covid-19 analysis using the class of high ordered integer-valued time series models with harmonic features. *Healthcare Analytics*, 2:100086, 2022.
- H. Kim and S. Lee. On first-order integer-valued autoregressive process with Katz family innovations. *Journal of Statistical Computation and Simulation*, 87:546–562, 2017.
- H.-Y. Kim and Y. Park. A non-stationary integer-valued autoregressive model. *Statistical Papers*, 49:485–502, 2008.
- L. A. Klimko and P. I. Nelson. On conditional least squares estimation for stochastic processes. *The Annals of Statistics*, 6:629–642, 1978.
- C. C. Kokonendji. Over-and underdispersion models. *Methods and Applications of Statistics in Clinical Trials: Planning, Analysis, and Inferential Methods*, 2:506–526, 2014.
- N. Kolev, L. Minkova, and P. Neytchev. Inflated-parameter family of generalized power series distributions and their application in analysis of overdispersed insurance data. *ARCH Research Clearing House*, 2:295–320, 2000.
- A. Latour. The Multivariate GINAR(p) process. *Advances in Applied Probability*, 29:228–248, 1997.
- A. Latour. Existence and stochastic structure of a non-negative integer-valued autoregressive process. *Journal of Time Series Analysis*, 19:439–455, 1998.
- X. Lin and Z. Guo. Model selection and estimation in INAR( $p$ ) model based on lasso class. In C. Huang, Y.-W. Chan, and N. Yen, editors, *Data Processing Techniques and Applications for Cyber-Physical Systems (DPTA 2019)*, pages 2053–2059, Singapore, 2020. Springer.
- X. Liu, D. Wang, D. Deng, J. Cheng, and F. Lu. Maximum likelihood estimation of the DDRCINAR( $p$ ) model. *Communications in Statistics-Theory and Methods*, 50:6231–6255, 2021.
- T. Lívio, N. M. Khan, M. Bourguignon, and H. S. Bakouch. An INAR(1) model with Poisson–Lindley innovations. *Economics Bulletin*, 38:1505–1513, 2018.
- J. Livsey, R. Lund, S. Kechagias, and V. Pipiras. Multivariate integer-valued time series with flexible autocovariances and their application to major hurricane counts. *The Annals of Applied Statistics*, 12:408–431, 2018.
- F. Lu and D. Wang. A new estimation for INAR(1) process with Poisson distribution. *Computational Statistics*, 37:1185–1201, 2022.
- B. P. McCabe and G. M. Martin. Bayesian predictions of low count time series. *International Journal of Forecasting*, 21:315–330, 2005.
- B. P. McCabe, G. M. Martin, and D. Harris. Efficient probabilistic forecasts for counts. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 73:253–272, 2011.
- R. McCleary, R. A. Hay, E. E. Meindinger, and D. McDowall. *Applied time series analysis for the social sciences*. Sage Publications, Beverly Hills, CA, 1980.

- E. McKenzie. Some simple models for discrete variate time series. *JAWRA Journal of the American Water Resources Association*, 21:645–650, 1985.
- A. S. Nastić, P. N. Laketa, and M. M. Ristić. Random environment integer-valued autoregressive process. *Journal of Time Series Analysis*, 37:267–287, 2016.
- P. Neal and T. Subba Rao. MCMC for integer-valued ARMA processes. *Journal of Time Series Analysis*, 28:92–110, 2007.
- W. Palma, R. Olea, and G. Ferreira. Estimation and forecasting of locally stationary processes. *Journal of Forecasting*, 32:86–96, 2013.
- Y. Park and H.-Y. Kim. Diagnostic checks for integer-valued autoregressive models using expected residuals. *Statistical Papers*, 53:951–970, 2012.
- X. Pedeli and D. Karlis. A bivariate INAR(1) process with application. *Statistical modelling*, 11:325–349, 2011.
- X. Pedeli and D. Karlis. On composite likelihood estimation of a multivariate INAR(1) model. *Journal of Time Series Analysis*, 34:206–220, 2013.
- X. Pedeli, A. C. Davison, and K. Fokianos. Likelihood estimation for the INAR( $p$ ) model by saddlepoint approximation. *Journal of the American Statistical Association*, 110:1229–1238, 2015.
- L. Piegl and W. Tiller. B-spline basis functions. In *The NURBS Book*, pages 47–79. Springer, Berlin, Germany, 1995.
- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2024. URL <https://www.R-project.org/>.
- N. Reid. Saddlepoint Methods and Statistical Inference. *Statistical Science*, 3:213 – 227, 1988.
- J. Rice. On the estimation of the parameters of a power spectrum. *Journal of Multivariate Analysis*, 9:378–392, 1979.
- M. M. Ristić, H. S. Bakouch, and A. S. Nastić. A new geometric first-order integer-valued autoregressive (NGINAR(1)) process. *Journal of Statistical Planning and Inference*, 139:2218–2226, 2009.
- M. M. Ristić, A. S. Nastić, and H. S. Bakouch. Estimation in an integer-valued autoregressive process with negative binomial marginals (NBINAR(1)). *Communications in Statistics-Theory and Methods*, 41:606–618, 2012.
- M. M. Ristić, A. S. Nastić, and A. V. Miletić Ilić. A geometric time series model with dependent bernoulli counting series. *Journal of Time Series Analysis*, 34:466–476, 2013.
- A. Roy and S. Karmakar. Time-varying auto-regressive models for count time-series. *Electronic Journal of Statistics*, 15:2905–2938, 2021.
- C. Santos, I. Pereira, and M. G. Scotto. On the theory of periodic multivariate INAR processes. *Statistical Papers*, 62:1291–1348, 2021.
- I. Silva and M. E. Silva. Asymptotic distribution of the Yule–Walker estimator for INAR( $p$ ) processes. *Statistics & Probability Letters*, 76:1655–1663, 2006.

- M. E. Silva and V. L. Oliveira. Difference equations for the higher order moments and cumulants of the INAR( $p$ ) model. *Journal of Time Series Analysis*, 26:17–36, 2005.
- F. W. Steutel and K. van Harn. Discrete analogues of self-decomposability and stability. *The Annals of Probability*, 7:893–899, 1979.
- Y. Sunecher, N. Mamode Khan, H. S. Bakouch, and V. Jowaheer. On some non-stationary bivariate INAR( $p$ ) models with applications to intra-day stock transaction series. *Journal of the Indian Society for Probability and Statistics*, pages 1–20, 2024.
- A. M. Sykulski, S. C. Olhede, A. P. Guillaumin, J. M. Lilly, and J. J. Early. The debiased Whittle likelihood. *Biometrika*, 106:251–266, 2019.
- X. Wang. Variable selection for first-order poisson integer-valued autoregressive model with covariables. *Australian & New Zealand Journal of Statistics*, 62:278–295, 2020.
- X. Wang, D. Wang, K. Yang, and D. Xu. Estimation and testing for the integer-valued threshold autoregressive models based on negative binomial thinning. *Communications in Statistics-Simulation and Computation*, 50:1622–1644, 2021.
- C. H. Weiß. Thinning operations for modeling time series of counts – a survey. *AStA Advances in Statistical Analysis*, 92:319–341, 2008.
- C. H. Weiß. Simultaneous confidence regions for the parameters of a Poisson INAR(1) model. *Statistical Methodology*, 8:517–527, 2011.
- C. H. Weiß. Integer-valued autoregressive models for counts showing underdispersion. *Journal of Applied Statistics*, 40:1931–1948, 2013.
- C. H. Weiß. A Poisson INAR(1) model with serially dependent innovations. *Metrika*, 78:829–851, 2015.
- C. H. Weiß. *An Introduction to Discrete-valued Time Series*. John Wiley & Sons, Hoboken, NJ, 2018.
- C. H. Weiß and C. Jentsch. Bootstrap-based bias corrections for INAR count time series. *Journal of Statistical Computation and Simulation*, 89:1248–1264, 2019.
- C. H. Weiß and M. H.-J. Feld. On the performance of information criteria for model identification of count time series. *Studies in Nonlinear Dynamics & Econometrics*, 24:20180012, 2020.
- P. Whittle. Recursive relations for predictors of non-stationary processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, 27:523–532, 1965.
- K. Yang, N. Xu, H. Li, Y. Zhao, and X. Dong. Multivariate threshold integer-valued autoregressive processes with explanatory variables. *Applied Mathematical Modelling*, 124:142–166, 2023.
- H. Zhang, D. Wang, and F. Zhu. Inference for INAR( $p$ ) processes with signed generalized power series thinning operator. *Journal of Statistical Planning and Inference*, 140:667–683, 2010.
- R. Zhu and H. Joe. Count data time series models based on expectation thinning. *Stochastic models*, 26:431–462, 2010a.
- R. Zhu and H. Joe. Negative binomial time series models based on expectation thinning operators. *Journal of Statistical Planning and Inference*, 140:1874–1888, 2010b.

H. Zou. The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101:1418–1429, 2006.