Change point tests for gradual changes in the Poisson-INARCH(1) process

Florian Schirra

Vom Fachbereich Mathematik der Rheinland-Pfälzischen Technischen Universität Kaiserslautern-Landau zur Verleihung des akademischen Grades Doktor der Naturwissenschaften (Doctor rerum naturalium, Dr. rer. nat.) genehmigte Dissertation.

Gutachter
Prof. Dr. Jörn Saß, RPTU in Kaiserslautern
Prof. Dr. Stefanie Schwaar, Hochschule für Technik und Wirtschaft Berlin
apl. Prof. Dr. Gaby Schneider, Goethe-Universität Frankfurt

Datum der Disputation: 11. Dezember 2025
DE-386

Acknowledgments

First I want to warmly thank Professor Jörn Saß and Professor Stefanie Schwaar for their close supervision of my PhD. The input on the topics of change point detection, statistics and stochastics in general was invaluable and the results of this theses would not be possible in the same manner without them. I am also very grateful that they always found time to give well-informed feedback on results that I presented them. Lastly, I could count on them finding the right words in times when the hardships of doing a PhD weighed harder on me.

Moreover, I thank my family and friends for creating a supportive environment that helped me keep progressing with my work. And perhaps more importantly, they made it possible to have time where I didn't think about my PhD at all and kept my head free of it for at least some moments. Foremost this means my girlfriend Lola, who is always able to lift up my mood, my brother Tobias and my mother Dorothe. I could also not imagine finishing this thesis without the best office neighbor Ria and the best friends possible: Zotti, Max, Jan, Jan, Andi, Jonas, Alex and Eric.

Lastly, I thank the financial department of the Fraunhofer ITWM for giving me the possibility of doing my PhD as an employee, providing me with the necessary time to accomplish this and the funds to visit conference and present my work. I am happy for every single one of my colleagues who make this the best working environment.

Abstract

Change point tests are a common tool to identify structural changes in the distribution of time series. In recent years, there has been progress in detecting changes within times series in countable spaces, e.g. the natural numbers. Such time series can be modeled by using Poisson-INARCH processes. Generally, it is assumed that changes occur as abrupt changes between stationary distributions. But in practice such changes often occur as a smooth transition, called gradual changes. To include such a behavior, we propose a new intensity function for Poisson-INARCH processes which makes gradual changes possible. We moreover propose a gradual change by adding a deterministic time dependent term on the intensity function. Properties as stationarity and a strong mixing property are investigated for both approaches. Under the alternative, convergence of estimators is ensured. We use those properties to prove consistency of tests based on both approaches under the null hypothesis and the alternative. We then analyze the quality of the models based on a comparative simulation study.

Zusammenfassung

Methoden zur Erkennung von Change-Points sind ein gängiges Instrument zur Ermittlung struktureller Veränderungen in der Verteilung von Zeitreihen. In den letzten Jahren gab es Fortschritte bei der Erkennung von Change-Points in Zeitreihen in abzählbaren Räumen, z. B. den natürlichen Zahlen. Solche Zeitreihen können mit Hilfe von Poisson-INARCH(1) Prozessen modelliert werden. Im Allgemeinen wird angenommen, dass Change-Points als Sprünge zwischen stationären Verteilungen auftreten. In der Praxis treten Change-Points jedoch oft als kontinuierliche Entwicklung zwischen beiden auf, die als graduelle Change-Points bezeichnet werden. Um ein solches Verhalten zu berücksichtigen, schlagen wir eine neue Intensitätsfunktion für Poisson-INARCH(1) Prozesse vor, die graduelle Change-Points möglich macht. Darüber hinaus schlagen wir vor, durch Hinzufügen eines deterministischen, zeitabhängigen Terms zur Intensitätsfunktion einen graduellen Change-Point in das Modell zu integrieren. Wir untersuchen beide Ansätze auf Eigenschaften wie Stationarität und eine Strong-Mixing-Eigenschaft. Unter der Alternative, stellen wir sicher, dass die Schätzer konvergieren. Das nutzen wir, um jeweils die Konsistenz eines auf beiden Ansätzen basierenden Tests unter der Nullhypothese und der Alternative nachzuweisen. Anschließend analysieren wir die Qualität der Modelle auf der Grundlage einer vergleichenden Simulationsstudie.

Contents

List of Figures

List of Tables

Acronyms

a.s.	Almost surely.
ACF	Autocorrelation function.
AMOC	At most one change.
AUEC	Asymptotically uniformly equicontinuous.
CDF	Cumulative distribution function.
CLS	Conditional least squares.
CUSUM	Cumulative sum.
g.c.d.	Greatest common denominator.
i.i.d	Independent and identically distributed.
M-Estimator	Maximum likelihood type estimator.
P-a.s.	Almost surely with respect to the measure P.
PACF	Partial autocorrelation function.
SLLN	Strong law of large numbers.
ULLN	Uniform law of large numbers.

Selected Notation

$\lVert\cdot\rVert$	Euclidian norm.
$\lVert\cdot\rVert_p$	L_p-norm.
$\lVert\cdot\rVert_{TV}$	Total variational norm.
$(x)_+$	$\max(0, x)$.
∇	Gradient operator with respect to θ.
∇^2	Hessian operator with respect to θ.
$\sigma(Y_i, ..., Y_{i+n})$	σ-algebra generated by the random variables $Y_i, ..., Y_{i+n}$.
$\mathcal{B}(X)$	Borel σ-algebra generated by the random variable X.
$\xrightarrow{d}$	Convergence in distribution.
$\mathbb{1}_A(x)$	Indicator function defined being 1 if $x \in A$ or 0 if $x \notin A$.
$\overline{X}_n$	Sample mean of $X_1, ..., X_n$.
$\mathrm{E}[X]$	Expectation of the random variable X.
$\mathrm{E}[X\|\mathcal{F}]$	Conditional expectation of the random variable X with respect to the σ-algebra $\mathcal{F}$.
$\mathrm{Var}(X)$	Variance of the random variable X.
$\mathrm{Var}(X\|\mathcal{F})$	Conditional variance of the random variable X with respect to the σ-algebra $\mathcal{F}$.
$X\|\mathcal{F} \sim Poi(\lambda)$	The conditional distribution of the random variable X with respect to the σ-algebra $\mathcal{F}$ is a Poisson distribution with intensity λ.
$\partial\Theta$	Boundary of the set Θ

1 Introduction

Change point tests are a statistical tool to decide whether data observed over time follows only one probability distribution or if it can be divided in at least two segments with different distributions. Consider a production process where the final product is measured. Although these values might not all be the same, they are distributed around the desired value. A sudden defect or an incorrectly adjusted machine could change the distribution of the size of these products. The producer is interested in deciding if the process at a given time is well functioning or not, in order to reduce the amount of mismanifactured products. As a result of such considerations, among others, change point testing as a field was introduced in Page [1954], by examining the cumulative sums of observations. In this introduction we start with a literature review on change point testing in different settings. We then see, that the topic of change point testing for gradual changes in count time series remains an open area. On this basis, we formulate our research question and briefly introduce our main contributions. Thereafter, in Section 1.2, we describe the structure of the thesis and conceptualize the approach to the topic of change point testing for gradual changes in count time series. We conclude the introduction with a more in depth explanation of our contributions.

1.1 Motivation and literature review

Since the publication of Page [1954], change point testing has been established as a valid approach to tackle problems of this kind with applications in finance (Chen and Gupta [2012]), medicine (Yang and Dumont [2006] Staudacher et al. [2005]), biology (Straub and Schneider [2020], Plomer et al. [2025]), image processing (Aach and Kaup [1995]) or software performance evaluation (Daly et al. [2020]). Over the years, the mathematical approaches have grown, e.g. by using maximum likelihood methods (Horvath [1993]), methods based on Bayesian statistics (Adams and MacKay [2007]) or machine learning algorithms like decision trees (Zheng et al. [2008]). The book R. Csörgő and Horváth [1997], as well as Chen and Gupta [2012] provide well guided introductions to the field in general. For more recent overviews on the literature on change point testing, we refer to Aminikhanghahi and Cook [2017], Truong et al. [2020] and van den Burg and Williams [2020].

A common assumption in change point testing is that the observations at the change point jump abruptly to a different distribution, e.g., the mean changes from one value to another. In practice, such changes might not present themselves as sudden jumps, instead a gradual transition between distributions may occur. Vogt and Dette [2015] and Quessy [2019] developed tests for such models and applied them to stock market data and global surface temperature data. Change points can even appear as just a gradually developing deviation from a stationary distribution, without reaching a new one. Such behavior can be seen in stock prices (Vogt and Dette [2015]) or also in macroeconomic data, where

Chu and White [1992] tested for a change in the trend of a time series. Hušková and Steinebach [2000] generalized this to a test for polynomially shaped gradual changes.

Often, the observed data are produced by counting processes, e.g., in traffic monitoring (Quddus [2008]), medicine (Franke and Seligmann [1993]) or in finance as the number of stock transactions (Brännäs and Quoreshi [2010]). Count time series display other properties compared to real valued time series and we refer to Weiß [2018] as an introduction. Models which allow for autocorrelation between subsequent observations are of particular interested. Count time series which fulfills this property are the INAR and the INARCH models. For both of these models, change point tests for abrupt change points were developed , see Franke et al. [2012] for the INARCH model and Weiß and Testik [2009] for the INAR model. Different applications of change point tests for count time series are e.g. the number of stock transactions (Hudecová et al. [2016], Doukhan and Kengne [2015], Kengne and Ngongo [2022]), economic data (Diop and Kengne [2016]) or crime monitoring (Cui et al. [2021]). Time series data can also display gradual changes instead of abrupt jumps. The population of invasive species displays a stationary behavior until the population increases gradually, see Cunze et al. [2025]. Another example is the number of infections regarding some contagious disease as e.g. Covid-19. When such a disease is already in a population there are always some daily cases but when an outbreak happens, the number of infections steeply increases. Chattopadhyay et al. [2022] introduced a test that detects change points in the number of infections of Covid-19. In a general setting, though, not much literature is available on the topic of testing for gradual change points in count time series.

This thesis aims to fill this gap by introducing two approaches for gradual changes in count time series. Moreover, we formulate a gradual change point model based on one of these approaches. We introduce two test statistics for this change point model and show that they produce consistent tests for gradual change points.

1.2 Outline

We proceed as follows: Section 2 familiarizes us with the general concept of change point testing in the setting of at most one change. In particular, we consider a hypothesis test, where we test the null hypothesis of no change against the alternative of exactly one change point. We formulate a test statistic of cumulative sums (CUSUM) for an example of independent normally distributed observations and present its asymptotics under both hypotheses. In Section 3 we get to know the underlying count time series model on which we base our gradual change point model. The so-called Poisson-INGARCH process is a variant of a GARCH process, based on innovations which are Poisson distributed. Moreover, we introduce the contraction property for Poisson-INGARCH processes which guarantees important properties of the process such as strong mixing and stationarity. Since gradually changing time series are not necessarily stationary, we conclude the third section with a property similar to the contraction property, yielding the strong mixing property for non-stationary time series. Section 4 introduces some Markov Chain theory. This will later be used to prove stationarity of a type of Poisson-INGARCH processes which do not fulfill the contraction property. Section 5 presents techniques to prove the convergence of estimators of time series. This is needed, because in most real-world settings, we do not know the model parameters of observations. Besides some more standard techniques for stationary processes, we are interested in results that hold in our gradual, and non-stationary setting. Therefore, we introduce the concept of triangular mixingales, which under certain conditions fulfill a strong law of large numbers, without being stationary. This strong law can then in return be used to assess the asymptotics of the estimates. In Section 6, we get to know a change point test for abrupt changes in time series of counts. The framework is from Franke et al. [2012] on the basis of a Poisson-INARCH(1) process. We are familiarized with the CUSUM test statistic as we use it later for the gradual change point model. Furthermore, we introduce assumptions on the process, which for the most part are also necessary in the gradual case. The section is concluded with a numerical analysis to get some intuition on the behavior of the test statistic. In Section 7, we present the first candidate for a gradual change point model. The approach is based on including the transition between two stationary distributions. To get a smooth gradual transition we introduce a new kind of Poisson-INGARCH(1) process based on a logistic function. This new process is not contractive, but we can use the results from Section 4 to prove stationarity and the strong mixing property. Section 8 introduces a different version of a gradual change point model. This variant is based on adding a time dependent function on the intensity of the process. We show that for a certain kind of such a function and additional assumptions, the conditional least squares (CLS) estimates for such a process still converge. Moreover, we prove these assumptions for a linear type of Poisson-INARCH(1) process and calculate the limits of the CLS estimates. Finally, in Section 9, we use the approach of the previous section to define a gradual change point model. We introduce a different kind of test statistic and prove that it converges in distribution. We prove consistency of both test statistics under the alternative, i.e., that they converge to infinity with probability one. We compare both test statistics in a numerical analysis. The thesis concludes with an outlook on further research questions.

1.3 Contributions

We propose two approaches for modeling gradual changes in count time series, in particular for the Poisson-INARCH(1) process. The first approach is based on the idea that in applications, observations often would not suddenly jump from one distribution into another. We expect there to be a transition period where the observations gradually shift from one distribution to another. We introduce a Poisson-INARCH(1) process with a logistic intensity function to have more flexibility in controlling the transition between distributions. Since the logistic function not necessarily yields the sufficient contraction property, we prove that the process is still stationary and converges to its stationary distribution for any starting value. The second model is characterized by a deterministic time-dependent function which is added on top of the intensity function after the change point. In this way, we can dictate the shape of the gradual change by the shape of the time-dependent function. Since the model is time-dependent after the change point, it can not be a stationary time series. We can use a more general concept than the contraction property to prove that the resulting time series is strongly mixing. The model can be described by a triangular and random array which is row-wise strongly mixing. We prove that the CLS estimates converge under some new assumptions. This result is achieved by showing that the CLS function fulfills a strong law of large numbers and an equicontinuity condition. We prove the new assumptions for an example of a time series following the second gradual change model. The limit of the estimates for this example is calculated and we compare numerically calculated estimates against their limit.

We then introduce a new change point test for gradual changes in the Poisson-INARCH(1) process. Under the null hypothesis of this test the process is a stationary Poisson-INARCH(1) process. This hypothesis is tested against the alternative that the process follows the second gradual change model. Thus, we extend the literature on change point models for the Poisson-INARCH(1) process to gradual changes. As described above those changes can have different shapes depending on the parametrization of the gradual change. To test for those hypotheses we introduce two test statistics. The first is the popular CUSUM test statistic as already used when testing for changes in the Poisson-INARCH(1) process. Since, the null hypothesis is the same as for an abrupt change point test, the resulting test is an asymptotic level α-test. Under the alternative we propose an assumption on the gradual change for which we proof consistency of the test statistic, i.e., it converges in probability to infinity. The second test statistic is based on a different model for gradual changes with independent observations. This test statistic is modified such that it is applicable to dependent observations. We generalize the theory for the asymptotic behavior of this test statistic to strongly mixing time series and get asymptotic level α under the null hypothesis. Under a different assumption regarding the alternative we get consistency of the second test statistic under the alternative. Finally, we prove both sets of the new assumptions for an example of the gradually changing Poisson-INARCH(1) process. We then carry out a simulation study based on this example to compare the strengths and weaknesses of both test statistics.

2 The Change Point problem

Suppose, we observe random variables $X_1, ..., X_n$ for some $n \in \mathbb{N}$. Fundamentally, a change point test for at most one change is a statistical hypothesis test. The null hypothesis of such a test is fulfilled, if for all $1 \leq i \leq n$ the X_i are identically distributed. This is tested against the alternative hypothesis, which states that there exists a number $m \in \mathbb{N}$ with $1 \leq m < n$ where $X_1, ..., X_m$ are identically distributed and also $X_{m+1}, .., X_n$ are identically distributed while X_m and X_{m+1} have distinct distributions. More formally this can be defined as:

Definition 2.1. Let $X_1, ..., X_n$, $n \in \mathbb{N}$, be random variables with distributions $\mathcal{D}_1, ..., \mathcal{D}_n$ where $X_i \sim \mathcal{D}_i$ for $1 \leq i \leq n$. Then, the change point problem (for at most one change) is to test the null hypothesis

$$H_0 : \mathcal{D}_1 = \mathcal{D}_2 = ... = \mathcal{D}_n$$

against the alternative

$$H_1 : \mathcal{D}_1 = ... = \mathcal{D}_{m^*} \neq \mathcal{D}_{m^*+1} = ... = \mathcal{D}_n$$

for $1 \leq m^* < n$.

If, moreover, we assume that one knows the cumulative distribution function $F_i(x|\theta)$ of X_i for some $\theta \in \mathbb{R}^p$, $p \in \mathbb{N}$, we can simplify the definition of the change point problem as follows:

Definition 2.2. Let $X_1, ..., X_n$, $n \in \mathbb{N}$, be random variables, where $F_i(\cdot|\theta_i)$ is the cumulative distribution function of X_i, for $1 \leq i \leq n$ and $\theta_i \in \mathbb{R}^p$. Then, the change point problem (for at most one change) is to test the null hypothesis

$$H_0 : \theta_1 = \theta_2 = ... = \theta_n$$

against the alternative

$$H_1 : \theta_1 = ... = \theta_{m^*} \neq \theta_{m^*+1} = ... = \theta_n$$

for $1 \leq m^* < n$.

This is a relatively broad definition. To get a better understanding of how changes, as defined above, look like and how to test for them. The next section explains a popular approach as an introductory example.

2.1 Example for a change in mean of independent normally distributed observations

Assume that the observed random variables $X_1, ..., X_n$ are independent and identically distributed (i.i.d.) following a normal distribution. Then, the definition of the change point problem for a change in mean is given by:

Definition 2.3. Let $X_1, ..., X_n$, $n \in \mathbb{N}$ be independent random variables, where $X_i \sim \mathcal{N}(\mu_i, \sigma^2)$ for $\sigma > 0$ and $\mu_i \in \mathbb{R}$, $i = 1, ..., n$. Then, the change point problem is to test the null hypothesis

$$H_0 : \mu_1 = \mu_2 = ... = \mu_n$$

against the alternative

$$H_1 : \mu_1 = ... = \mu_{m^*} \neq \mu_{m^*+1} = ... = \mu_n$$

for $1 \leq m^* < n$.

To get an idea on the behavior of a model following the definition above, consider the following example and visualization of the respective sample paths.

Example 2.4. Let $X_1, ..., X_{200}$ be a sequence of independent random variables where $X_i \sim \mathcal{N}(\mu_i, \sigma)$, for $\sigma = 1$, $i = 1, ..., 200$. Then, if we set

$$\mu_1 = \mu_2 = ... = \mu_{200} = 0$$

we are under H_0 and if we set

$$\mu_1 = \mu_2 = ... = \mu_{100} = 0,\ \mu_{101} = \mu_{102} = ... = \mu_{200} = 2$$

we have an example for the alternative. In Figure 2.1, we see an example of no change in this setting. We notice, that they are effectively 200 observations of i.i.d. standard normally distributed random variables. Figure 2.2 shows a sample path with a change point at $m^* = 100$, where the expectation changes from 0 to 2, while the standard deviation stays constant at 1. Here, we have two sets of i.i.d. random variables which differ only in the expectation.

To construct a test statistic which is able to discern those two cases, at first there is need for a consistent estimator of the parameter that changes. In this

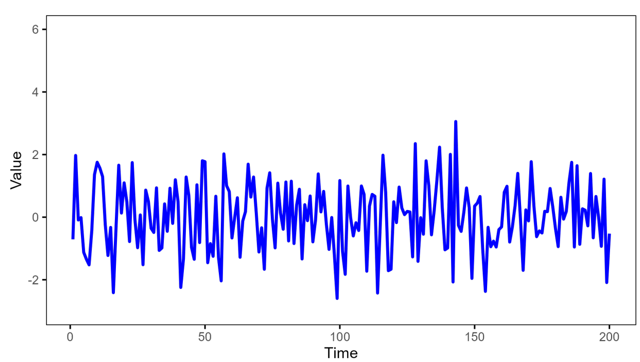

Figure 2.1: Example of no change for independent normal distributed random variables.

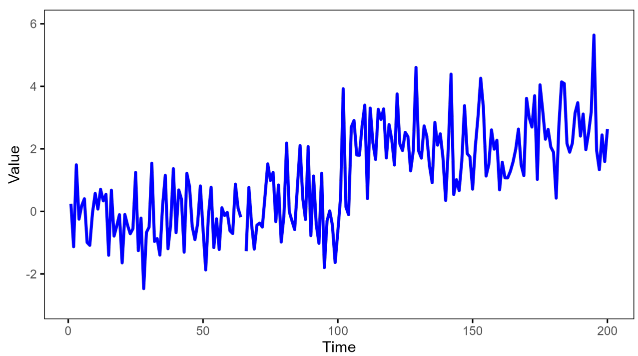

Figure 2.2: Example of a change point in mean at $m = 100$ for independent normal distributed random variables.

case, this is the sample mean $\overline{X}_n$, given by

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

An idea could be to use that estimator as a test statistic by comparing the deviation of each observation to the estimator, i.e., $X_i - \overline{X}_n$. In case of H_0, the expectation of that difference is equal to 0. Moreover, if all those differences are added up for $i = 1, ..., k$ and some $k \in \{1, ..., n\}$, the expectation of that sum is still 0. Under H_1 on the other hand, this is not necessarily the case. Looking back at the example in Figure 2.2, the expectation of $\overline{X}_n$ is 1, while the expectation of X_i is either 0 or two depending on i being smaller or larger than the position of the change point m^*. Thus, the expectation of the difference $X_i - \overline{X}_n$ is either 1 or -1, again depending on i being smaller or larger than m^*. Note that this behavior would be reversed if the mean would change from 2 to 0. Adding up those differences from 1 to k and taking the absolute value of this sum yields under H_1 a time series which has linearly increasing expectations from 0 to m^* and linearly decreasing expectation from $m^* + 1$ to n. Such cumulative sums could therefore stay uniformly close to 0 under H_0 while attaining large values under H_1, in particular the maximum of the cumulative sums. Hence, a test statistic based on the maximum over all $k = 0, ..., n$ of such cumulative sums should be able to discern between H_0 and H_1. This would happen either, by being close to 0, or by growing comparatively large. Such a test statistic is called CUSUM which is defined as:

Definition 2.5. Let $X_1, ..., X_n$, $n \in \mathbb{N}$ be a sequence of random variables, then

$$T_n = \max_{0<k<n} \sqrt{\frac{n}{k(n-k)}}|S_k| \tag{2.1}$$

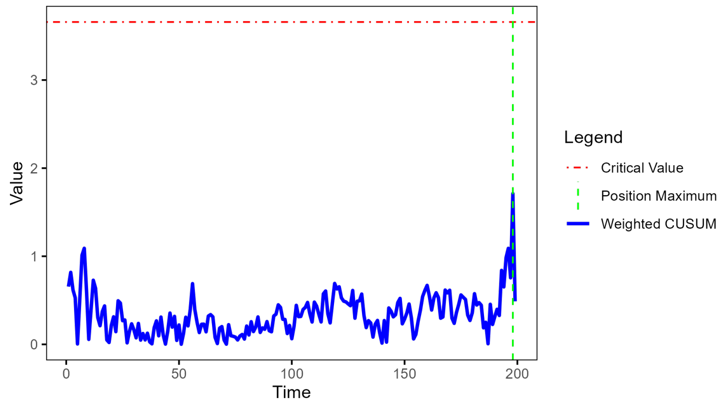

Figure 2.3: Weighted cumulative sums with their maximums T_n marked by the green line for the examples in Figure 2.1.

with

$$S_k = \sum_{i=1}^{k} (X_i - \overline{X}_n)$$

is called CUSUM test statistic.

Remark 2.6. Note, that we added a weighting factor in front of the cumulative sums, which is for a given $k \in \{1, ..., n\}$ of order $O(\sqrt{n}^{-1})$. This is done because the cumulative sums do in fact diverge to infinity for $n \to \infty$. Although having an expectation of 0, the sum of the differences $X_i - \overline{X}_n$ does not converge itself.

Example 2.7. We apply this CUSUM test statistic to the paths from Example 2.4. The results are displayed in Figure 2.3 and Figure 2.4. To be more precise, the blue line shows the values of $\sqrt{\frac{n}{k(n-k)}}|S_k|$ for $k = 1, ..., 200$ and the vertical green line highlights the position of the maximum of those values, which indicates that we can find the maximum T_n there. In Figure 2.3, the H_0 case, one can see that the values all stay relatively close to 0. There is no clear structure, just small deviations from 0, induced from the randomness of the process. For the H_1 case on the other hand, there is a clear structure identifiable, see Figure 2.4. Here, $\sqrt{\frac{n}{k(n-k)}}|S_k|$ increases until the position of the change point and then starts to decrease again. This is the expected behavior for the expectation of the cumulative sums. The maximum, i.e., the test statistic T_n, marked by the vertical green line is about 5 times higher than in the case of H_0.

This confirms the intuition about the test statistic, that it stays low if the observations are homogeneous, i.e., there is no change in the observations. If there is a change point, then the test statistic takes larger values. The behavior under H_0 can, at least asymptotically, be described as in the following, see for example Theorem 2.1.2 in R. Csörgő and Horváth [1997].

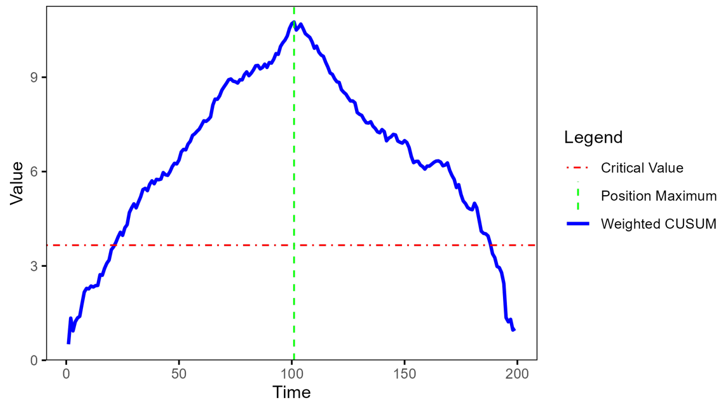

Figure 2.4: Weighted cumulative sums with their maximums T_n marked by the green line for the example in Figure 2.2.

Theorem 2.8. *Under H_0, i.e., $X_1, ..., X_n$, $n \in \mathbb{N}$ is a sequence of i.i.d. random variables where $X_i \sim \mathcal{N}(\mu, \sigma)$, for $\mu \in \mathbb{R}$ and $\sigma > 0$, we have*

$$P\left(a(\log(n))\frac{T_n}{\sigma} - b(\log(n)) < x\right) \overset{n\to\infty}{\longrightarrow} P(G_2 < x) = e^{-2e^{-x}} \tag{2.2}$$

and

$$P\left(a(\log(n))\frac{T_n}{\hat{\sigma}_n} - b(\log(n)) < x\right) \overset{n\to\infty}{\longrightarrow} P(G_2 < x) = e^{-2e^{-x}}, \tag{2.3}$$

where G_2 is a Gumbel extreme value distributed random variable, $a(x) = \sqrt{2\log(x)}$ and $b(x) = 2\log(x) + 1/2\log(\log(x)) - 1/2\log(\pi)$. Moreover, for

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

it holds that

$$\hat{\sigma}_n - \sigma = o(\log(\log(n))^{-1/2}).$$

Consequently, under the null hypothesis, the test statistic is asymptotically Gumbel distributed. Whereas, under the alternative, we get that asymptotically the test statistic converges to infinity in probability:

Theorem 2.9. *Let $m^* := m_n = \lfloor \tau n \rfloor$ for $n \in \mathbb{N}$ and $0 < \tau < 1$. Under H_1, i.e., $X_1, ..., X_n$, $n \in \mathbb{N}$ is a sequence of independent random variables where $X_i \sim \mathcal{N}(\mu_i, \sigma)$, for $\sigma > 0$ and $\mu_i \in \mathbb{R}$, $i = 1, ..., n$, where $\mu_1 = ... = \mu_m \neq \mu_{m+1} = ... = \mu_n$, we have for all $c > 0$*

$$P\left(a(\log(n))\frac{T_n}{\hat{\sigma}_n} - b(\log(n)) > c\right) \overset{n\to\infty}{\longrightarrow} 1. \tag{2.4}$$

Remark 2.10. Under the alternative we usually set the position of the change point in relation to the number of observations n, here given by the relative position τ. That is done to ensure that the amount of time points before and after the change point grows to infinity as n grows to infinity. Then, we are able to apply asymptotic results on the time series before and after the change point.

From (2.2), for $0 < \alpha < 1$ we can deduce the critical value

$$c(\alpha) = \frac{\hat{\sigma}_n}{a(\log(n))}(q_{1-\alpha} + b(\log(n))), \tag{2.5}$$

where $q_{1-\alpha}$ is the $(1-\alpha)$-quantile of G_2, i.e., the Gumbel extreme value distribution, and $a(\log(n))$, $b(\log(n))$ are as in Theorem 2.8. We call such a test an asymptotic level α-test. Moreover, since the test statistic converges with probability 1 to infinity under H_1, we say that the test has asymptotic power 1 under the alternative.

Example 2.11. Going back to Example 2.7, this critical value corresponds to the red horizontal lines in Figures 2.3 and 2.4. In both cases they follow (2.5) with a significance level of $\alpha = 0.05$. As expected, the test statistic is small under H_0 and large under H_1. Moreover, with these critical values, it is possible to accept H_0 in the first example and reject H_0 in the second the example.

If we find that H_0 has been rejected, the relative position of the maximum of the CUSUM statistic yields an estimator of the relative position of the change point, see Lemma 1.5.3 in R. Csörgő and Horváth [1997].

Theorem 2.12. *Let the conditions of Theorem 2.9 be fulfilled. Then, we have for*

$$\hat{m}_n = \underset{1 \le k < n}{\arg\max} \sqrt{\frac{n}{k(n-k)}} |S_k| \tag{2.6}$$

that

$$\hat{\tau}_n - \tau = o(1) \ \ a.s. \tag{2.7}$$

for $\hat{\tau}_n := \frac{\hat{m}_n}{n}$.

Example 2.13. Revisiting Example 2.7 and the Figures 2.3 and 2.4, in both cases the green vertical lines are marking the position of the maximum, i.e., $\hat{m}_n$ from (2.6). For the example under the null hypothesis, this maximum is just at a random time point between the first and last observation. For the example with a change on the other hand, the position of this maximum contains information about the position of the change. In this case it is equal to the position of the change point with $\hat{m}_n = 100$. Note that in general it is not guaranteed that $\hat{m}_n$ converges to m^* itself. That only holds for the relative position, as stated in the previous theorem.

Remark 2.14. To conclude this introduction to the concept of change point testing, note that there are different versions of the CUSUM statistic, which use different weighting factors than in (2.1), namely:

$$T_{n,2} = \max_{\epsilon n \leq k \leq n-\epsilon n} \sqrt{\frac{n}{k(n-k)}} |S_k|, \tag{2.8}$$

$$T_{n,3} = \max_{1 \leq k \leq n} \left(\frac{n^2}{k(n-k)} \right)^{\beta} \frac{1}{\sqrt{n}} |S_k|, \tag{2.9}$$

where $0 < \epsilon < 1$ and $1 \leq \beta < 1/2$. Similar results to Theorems 2.8, 2.9 and 2.12 still hold for those test statistics. Here, under H_0, the test statistics converge in distribution to transformed versions of the supremum of a Brownian bridge $B(t)$, i.e.,

$$\frac{T_{n,2}}{\hat{\sigma}_n} \xrightarrow{d} \sup_{\epsilon \leq t \leq 1-\epsilon} \frac{|B(t)|}{\sqrt{t(1-t)}},$$

$$\frac{T_{n,3}}{\hat{\sigma}_n} \xrightarrow{d} \sup_{0 \leq t \leq 1} \frac{|B(t)|}{(t(1-t))^{\beta}},$$

where $\hat{\sigma}_n$ is given as in Theorem 2.8.

2.2 Numerical results

In this section, we analyze how different parameter combinations influence the outcome of the test. The theoretical results established in the section before tell us only that, for a large enough sample size, the results are indifferent to the parameters. In practice we usually do not have samples large enough to disregard, e.g., the effect of a large change in μ in comparison to a small change. In fact, we would expect for small sample sizes to obtain rejection rates, i.e., the amount of times the null hypothesis is rejected relative to the sample size, which are a bit above or below the significance level under the null hypothesis, and under the alternative to be lower than 100%. In Table 2.1, we can see the rejection rates for $\mu_1 = 0$, $\sigma = 1$, $\mu_n = 1/2$, $\alpha = 5\%$ and 10000 repetitions under H_1. The rejection rate under the null hypothesis with a maximum of 1.29% climbs for increasing observations just above 1. Although, there should be roughly 5% of rejections, meaning that the test displays a tendency to be conservative under H_0. Regarding the rejections under the alternative, they both climb to effectively 100% for 1000 observations. But they are lower before that, sitting at 18.47% and 11.95% for $\tau = 0.5$ and $\tau = 0.75$, respectively. It seems that for a given difference in μ before and after the change point, the test needs a certain number of observations to be able to correctly identify the alternative. The difference between the results for $\tau = 0.5$ and $\tau = 0.75$ suggests, that it is not just about the absolute number of observations. Both before and after the change point there need to be enough observations such that the asymptotic results show effect.

For the same experiments as in Table 2.1, we have calculated the estimated relative position $\hat{\tau}_n$ of the change point under the alternative, see Table 2.2.

Number of observations	100	200	500	1000
H_0	0.55%	0.75%	1.01%	1.29%
$H_1,\ \tau = 0.5$	18.47%	56.71%	98.44%	100%
$H_1,\ \tau = 0.75$	11.95%	38.36%	92.35%	99.97%

Table 2.1: Empirical rejection rates of the test statistic for $\mu_0 = 0$, $\sigma = 1$, α =5%, 10000 repetitions and $\mu_n = 1/2$ under H_1.

Number of observations	100	200	500	1000
$H_1,\ \tau = 0.5$	0.4996	0.501	0.4992	0.4998
$H_1,\ \tau = 0.75$	0.7133	0.7316	0.7453	0.7489

Table 2.2: Average estimated relative position of the change point for $\mu_0 = 0$, $\mu_n = 1/2$, $\sigma = 1$, $\alpha = 5\%$ and 10000 repetitions.

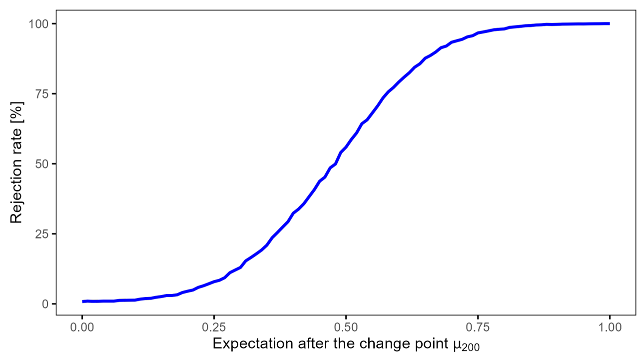

Figure 2.5: Empirical rejection rates of the test for $\mu_n \in [0, 1]$ with $\mu_0 = 0$, $\sigma = 1$, $\tau = 0.5$, $\alpha = 5\%$, $n = 200$ and 10000 repetitions.

This is done by taking the mean of the estimated relative positions in the cases, where the null hypothesis is rejected. For $\tau = 0.5$, the relative position $\hat{\tau}_n$ is close to constant around 50%. For $\tau = 0.75$, we get values for small sample sizes $\hat{\tau}_n$ with a deviation up to 0.04 from the real value 0.75. This is a smaller difference if viewed in absolute terms: For $n = 100$ observations, the position of the change is estimated with an error of being about 3 to 4 time steps too early. If we look at $n = 1000$, this estimated time of change is at 11 time steps before the change point. So, while recognizing this effect of reduced accuracy, it is of less concern than the effects on the rejection rate.

We want to further investigate the effects of changing parameters on the rejection rate. Intuitively, if all the other parameters stay constant, one would assume the following effects: first, an increasing difference $|\mu_1 - \mu_n|$ should increase the rejection rate. If the change is more pronounced, we assume the test to perform better. In Figure 2.5, we see the results of an experiment test-

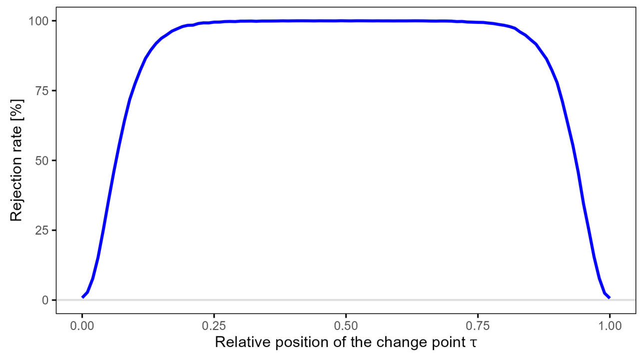

Figure 2.6: Empirical rejection rates of the test for $\tau \in [0, 1]$ with $\mu_0 = 0$, $\mu_n = 1$, $\sigma = 1$, $\alpha = 5\%$, $n = 200$ and 10000 repetitions.

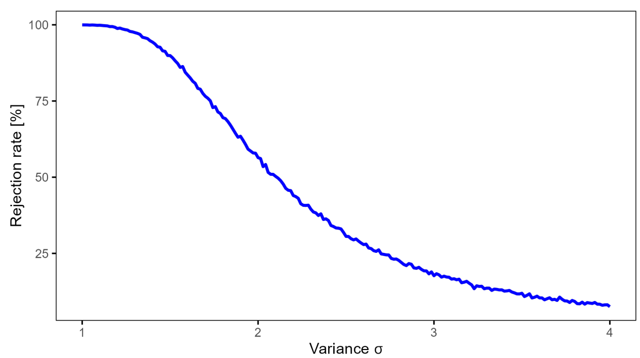

Figure 2.7: Empirical rejection rates of the test for $\sigma \in [0, 4]$ with $\mu_0 = 0$, $\mu_n = 1$, $\tau = 0.5$, $\alpha = 5\%$, $n = 200$ and 10000 repetitions.

ing for that effect. For otherwise constant parameters, we varied the difference $|\mu_1 - \mu_n|$ between 0 and 1. Resulting in rejection rates increasing from about 1% to 100% in an S-shaped curve. So, experimentally, the intuition holds true. Second, if the position of the change point moves further to the edges, the rejection rate should decrease. We have seen before, that the position of the change point tends to have this impact. In the next experiment, we applied the test to values of τ between 0 and 1, see Figure 2.6. The curve starts at 0 and increases steeply to arrive at a plateau of close to 100% near $\tau = 0.25$. This plateau lasts until around $\tau = 0.75$ and then begins to drop sharply to 0 again. The graph seems to be mirrored at $\tau = 0.5$. This not only supports our idea that not only the absolute amount of observations is important. We also need to consider the distribution before and after the change point. In an extreme case of $n = 1000$ and $m = 1$ there would be no way to discern between an anomaly and a change point. Moreover, by the symmetry of the curve, we can deduce that there is no

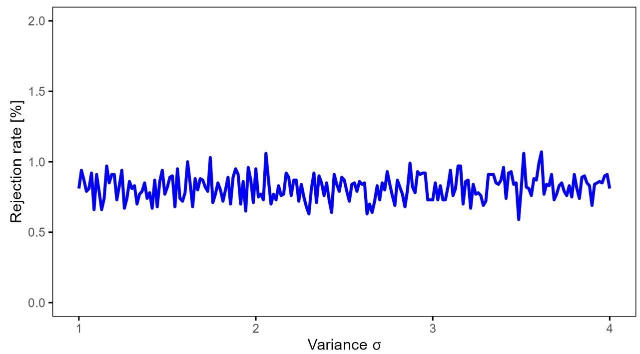

Figure 2.8: Empirical rejection rates of the test for $\sigma \in [0, 4]$ with $\mu_0 = \mu_n = 0$, $\alpha = 5\%$, $n = 200$ and 10000 repetitions.

difference between an early change point and a later one, if both have the same distance to 0 and n, respectively. Lastly, an increasing standard deviation σ could decrease the rejection rate under H_1. Since, we expect a larger σ to blur the difference between the distributions before and after the change point. In Figure 2.7 we see the results of an experiment where we set σ to values in $[1, 4]$. We get a result similar to a reversed version of Figure 2.5, where the rejection rate decreases in an S-shaped curve from 100% to close to 0%. This implies that a higher variance or noise in the data makes it more difficult to identify that the null hypothesis is not fulfilled. We did the same experiment for data following the null hypothesis, see Figure 2.8. Here, the results conflict with the intuition. The rejection rate seems to be relatively constant at around 0.82%. This does not show any evidence that large deviations produce time series which are more likely to reject the null hypothesis.

Now, we have an idea how we can define a change point test and know the CUSUM approach to formulate a test statistic. We illustrated how the test statistic behaves under H_0 and H_1, based on numerical examples. Furthermore, we know how to estimate the relative position of the change point, if the null hypothesis is rejected. Moreover, we got an understanding of how the parameters influence the outcome of a test in a setting where the number of observations is limited. We proceed by introducing the time series model central to this work.

3 Count time series models

Count time series - fundamentally - are time series $(Y_i)_{i\in\mathbb{N}_0}$ where the random variables Y_i take only values in countable sets. For our purposes, we are only interested in the natural numbers including 0. Generally, it is possible to consider other countable spaces as well. The aim of this section is to introduce a special type of count time series, the so called Poisson-INGARCH process. In the Section 3.1, we introduce the concept of autoregressive models for real valued time series. We will make use of one of those concepts to obtain an autoregressive structure for count time series. Then, we introduce the idea of strong mixing, a concept of asymptotically vanishing dependence between random variables in a time series. With strongly mixing time series it is possible to use e.g. laws of large numbers or central limit theorems for dependent time series. In Section 3.3 we introduce the Poisson-INGARCH process and an important property, which is called contraction property. Thereafter we present results for Poisson-INGARCH processes fulfilling this contraction property, e.g., that they are strongly mixing. We use an example to study the effects of these results. The section concludes with a more general concept than the contraction property, which still ensures the strong mixing property.

3.1 Introduction to autoregressive time series models

As a starting point for the introduction of autoregressive time series we consider a class of time series with cumulative distribution functions independent of time. We call such time series stationary and a more precise definition is the following.

Definition 3.1. Let $(Y_i)_{i\in\mathbb{N}_0}$ be a time series with cumulative distribution functions $F_{i_1,...,i_n} : \mathbb{R}^n \to [0,1]$ for $n, i_1, ..., i_n \in \mathbb{N}$ given by

$$F_{i_1,...,i_n}(y_1,...,y_n) = P(Y_{i_1} \leq y_{i_1}, ..., Y_{i_n} \leq y_{i_n})$$

for $y_1, ..., y_n \in \mathbb{R}$. The time series is called stationary, if

$$F_{i_1,...,i_n}(y_1,...,y_n) = F_{i_1+j,...,i_n+j}(y_1,...,y_n)$$

for any $n, i_1, ..., i_n, j \in \mathbb{N}$ and $y_1, ..., y_n \in \mathbb{R}$.

The consequence of stationarity is a constant expectation of a time series, which follows directly if we set n equal to 1. On the on hand, this shows that stationarity is in conflict with our goal of modeling gradually increasing time series. On the other hand, stationarity is accompanied with stochastic properties that we want to make use of. Moreover, we will test the case of gradually changing time series against a case where the time series is stationary. Hence, we need stationarity at least to define the null hypothesis. A gradually increasing time series can then be achieved by adding a time dependent term on this stationary model. Usually defined on real valued time series, autoregressive moving average (ARMA) type models can, under certain conditions, fulfill these properties. This leads to the idea of modeling count time series with this approach.

Definition 3.2. The real valued process $(Y_i)_{i\in\mathbb{N}_0}$ is called ARMA(p,q) if

$$Y_i = \omega + \sum_{j=1}^{p} \alpha_j Y_{i-j} + \sum_{j=1}^{q} \beta_j \varepsilon_{i-j} + \varepsilon_i,$$

where $\omega \in \mathbb{R}$, $p, q \in \mathbb{N}$, $\alpha_j \in \mathbb{R}$ for $j = 1, ..., p$, $\beta_j \in \mathbb{R}$ for $j = 1, ..., q$ and a process $(\varepsilon_i)_{i\in\mathbb{N}_0}$ with $E[\varepsilon_i] = 0$, $\mathrm{Var}(\varepsilon_i) = \sigma^2$ and $\mathrm{Cov}(\varepsilon_i, \varepsilon_j) = 0$ for $\sigma > 0$ and $i, j \in \mathbb{N}_0$ where $i \neq j$. The process $(\varepsilon_i)_{i\in\mathbb{N}_0}$ is called white noise.

The disadvantage is, that if we use an i.i.d. integer valued process instead of white noise, the resulting process is not necessarily integer valued itself. This is only ensured, if we restrict the parameters to be natural numbers themselves. Or else, the product of a parameter α_j with an integer valued Y_{i-j} is not an integer, and so would be the resulting value of Y_i. Consequently, the minimal positive value for both the α_j and β_j is 1. If just one single of those α_j would be equal to 1, then the process could not be stationary. Consider an ARMA(1,0), also called AR(1) process, with $\alpha_1 = 1$, then

$$\begin{aligned}\mathrm{Var}(Y_i) &= \mathrm{Var}(Y_{i-1}) + \mathrm{Var}(\varepsilon_i) = ... = \mathrm{Var}(Y_0) + \sum_{j=1}^{i} \mathrm{Var}(\varepsilon_j) = \mathrm{Var}(Y_0) + i\sigma^2 \\ &> \mathrm{Var}(Y_0) + (i-1)\sigma^2 = \mathrm{Var}(Y_{i-1}).\end{aligned}$$

Moreover, this implies a variance which grows to infinity. If $\alpha_1 = 2$, the variances would even grow with an exponential rate. Thus, we are limited to processes which are non-stationary with different growth rates to infinity. Or the process being equal to the innovations ε_i if all the parameters are 0. To overcome those limitations, we are using a model based on the concept of so called generalized autoregressive conditionally heteroskedastic (GARCH) time series. This concept is grounded on the observation that financial time series exhibit the property that variances form so called clusters. Variance clustering describes the behavior, that the variance is not constant over the complete time series but large prices changes tend to be followed by large price changes. And small price changes also tend to be followed by small price changes. This can be modeled by making the variances dependent on the observations.

Definition 3.3. The process $(Y_i)_{i\in\mathbb{N}_0}$ is called GARCH(p,q) if

- $\mathrm{E}[Y_i|\mathcal{F}_{i-1}] = 0$,
- $\mathrm{Var}(Y_i|\mathcal{F}_{i-1}) = \sigma_i^2$
- and $\sigma_i^2 = \omega + \sum\limits_{j=1}^{p} \alpha_j Y_{i-j}^2 + \sum\limits_{j=1}^{q} \beta_j \sigma_{i-j}^2$,

where $\omega > 0$, $p, q \in \mathbb{N}$, $\alpha_j \geq 0$ for $j = 1, ..., p$ and $\beta_j \geq 0$ for $j = 1, ..., q$. In the special case where $q = 0$ the process is called ARCH(p).

The ARMA-like form of the variance results in autocorrelation of the variance and the correlation to the process itself. Therefore, GARCH(p,q) processes yield the desired property of variance clusters. For more insights on ARMA and GARCH type processes, see Franke et al. [2010]. Modeling a parameter of an integer valued distribution as an ARMA-like process results in an integer valued GARCH (INGARCH) process. Since usually the parameters take values in the real numbers, the ARMA process of that parameter is not restricted to have only integer valued parameters itself. Hence, the issues that we have described beforehand are overcome and the variety of processes that such a model could produce are more promising.

3.2 Strong mixing conditions

Real-world phenomena which occur in close temporal proximity often have significant influence on each other. This means, that present observations are dependent on observations in the near past. But in the long run, the influence of a present observation to one in the far future is negligible. E.g., a water particle in a river, being heavily dependent on its position a few seconds earlier. But the position of that particle hours before is close to being irrelevant. One way of describing such properties are so called strong mixing conditions. They are defined on the basis that the dependence of observations should tend to 0 the more they are away from each other. Different measures of dependence are used to define different strong mixing conditions, two of whom we will get to know in this subsection. The concept was first introduced by Rosenblatt [1956], where the author - beyond just the definition - already formulated and proved a central limit theorem for strongly mixing processes. So while being a relatively general concept, it inherits important properties known for independent time series. We start with defining the mentioned measures of dependence.

Definition 3.4. Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\mathcal{A}, \mathcal{B} \subset \mathcal{F}$ be σ-algebras. Then, define

$$\alpha(\mathcal{A}, \mathcal{B}) := \sup_{A \in \mathcal{A}, B \in \mathcal{B}} |P(A \cap B) - P(A)P(B)| \tag{3.1}$$

$$\beta(\mathcal{A}, \mathcal{B}) := \sup_{A_1, \dots, A_I \in \mathcal{A}, B_1, \dots, B_J \in \mathcal{B}} \frac{1}{2} \sum_{i=1}^{I} \sum_{j=1}^{J} |P(A_i \cap B_j) - P(A_i)P(B_j)| \tag{3.2}$$

for arbitrary $I, J \in \mathbb{N}$.

Both these definitions have in common, that they consist of terms which are 0 if the σ-algebras are independent. Thus, they are usually called measures of dependence, since the "closer" such σ-algebras are to independence, the closer these measures are to 0. As stated before, there are many more such measures of dependence and thus more ways of formulating distinct mixing conditions. For more of those concepts and deeper insights of properties of strong mixing conditions not necessary for this work, consider Bradley [2002] or as a very brief summary Bradley [2005].

Remark 3.5. The term strong mixing arises from the fact that Rosenblatt [1956] wanted to discern it from the mixing in the ergodic-theoretic sense. We refer to Walters [1982] for an introduction to ergodic theory. Moreover, we generally call all mixing conditions based on the measures of dependence given in Bradley [2005] strong mixing conditions. This results from the fact, that all the other conditions are stronger in the sense that they imply α-mixing, i.e., the original strong mixing condition.

We now gather some properties of these two measures of dependence, which either give an idea about how they behave or will be helpful later in this work.

Lemma 3.6. *Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\mathcal{A}, \mathcal{B} \subset \mathcal{F}$ be σ-algebras. Then, we have*

(i) $0 \leq \alpha(\mathcal{A}, \mathcal{B}) \leq 1/4$, $0 \leq \beta(\mathcal{A}, \mathcal{B}) \leq 1$,

(ii) $\alpha(\mathcal{A}, \mathcal{B}) = \beta(\mathcal{A}, \mathcal{B}) = 0$ *if $\mathcal{A}$ and $\mathcal{B}$ are independent,*

(iii) $2\alpha(\mathcal{A}, \mathcal{B}) \leq \beta(\mathcal{A}, \mathcal{B})$.

For proofs we refer to Propositions 3.4 and 3.11 of Bradley [2002]. The second property follows directly from the definition of the suprema, being based on values which are all 0 if the σ-algebras are independent. As GARCH type processes are defined based on conditional distributions, we are interested in upper bounds of conditional expectations of such a process. One is given by the following result, based on Lemma 2.1 in McLeish [1975].

Lemma 3.7. *Let $(\Omega, \mathcal{F}, P)$ be a probability space, $\mathcal{A}, \mathcal{B} \subset \mathcal{F}$ be σ-algebras, X a random variable measurable with respect to $\mathcal{A}$ and $1 \leq p \leq r < \infty$. Then*

$$||\mathrm{E}[X|\mathcal{B}] - \mathrm{E}[X]||_p \leq 2(2^{1/p} + 1)\alpha(\mathcal{B}, \mathcal{A})^{1/p-1/r}||X||_r. \tag{3.3}$$

Now, we apply those measures of dependence on time series, which will result in the desired strong mixing conditions.

Definition 3.8. Let $Y = (Y_i)_{i \in \mathbb{N}_0}$ be a time series and define for $0 \leq J \leq L \leq \infty$ the σ-algebras

$$\mathcal{F}_J^L = \sigma(Y_i, J \leq i \leq L).$$

Then, for $n \in \mathbb{N}$

$$\alpha_Y(n) := \sup_{j \in \mathbb{N}_0} \alpha(\mathcal{F}_0^j, \mathcal{F}_{j+n}^\infty), \tag{3.4}$$

$$\beta_Y(n) := \sup_{j \in \mathbb{N}_0} \beta(\mathcal{F}_0^j, \mathcal{F}_{j+n}^\infty). \tag{3.5}$$

We call the time series $(Y_i)_{i \in \mathbb{N}_0}$ strongly mixing (or α-mixing) if $\alpha_Y(n) \overset{n\to\infty}{\longrightarrow} 0$ and absolutely regular (or β-mixing) if $\beta_Y(n) \overset{n\to\infty}{\longrightarrow} 0$.

So, the measures of dependence in this setting are not applied to finite parts of the time series, in contrast to, e.g., independent time series, where the random variables are pairwise independent. Here, we measure dependence by gathering all the information up to some time point $i \in \mathbb{N}_0$ and all the information from another time point $i+n$ for some $n \in \mathbb{N}$, i.e., the respective σ-algebras. And then in order to get a measure of dependence for time series, we apply the measures defined beforehand in (3.4) and (3.5) to those σ-algebras.

Remark 3.9. (i) A consequence of Lemma 3.6 (ii) is that an independent time series $Y = (Y_i)_{i \in \mathbb{N}_0}$ is mixing since then $\alpha_Y(n) = \beta_Y(n) = 0$.

(ii) From Lemma 3.6 (iii) we can deduce that absolute regularity is a stronger property than strong mixing, since if a time series is absolutely regular we have $\alpha_Y(n) \leq 1/2\beta_Y(n) \overset{n\to\infty}{\longrightarrow} 0$.

A property of strongly mixing time series which will come in handy later is the following. It shows that the strong mixing property is maintained when applying a continuous function on finitely many random variables of a mixing time series.

Lemma 3.10. *Let $Y = (Y_i)_{i \in \mathbb{N}_0}$ be a strongly mixing (absolutely regular) time series with $\alpha_Y(n) = O(a_n)$ ($\beta_Y(n) = O(a_n)$) with $a_n \overset{n\to\infty}{\longrightarrow} 0$. Then, for some continuous function $f : \mathbb{R}^k \to \mathbb{R}$, $k \in \mathbb{N}$ the time series $Z = (Z_i)_{i \in \mathbb{N}_0}$ defined by $Z_i := f(Y_i, ...Y_{i+k})$ is strongly mixing (absolutely regular) with $\alpha_Z(n) \leq \alpha_Y(n-k)$ ($\beta(n)_Z \leq \beta_Y(n-k)$).*

We refer to Remark 1.8(b) in Bradley [2005] for this result in the case of α-mixing and stationary time series. The general case can be found in Theorem 14.1 of Davidson [2021].

The next theorems motivate our interest in the mixing property: an invariance principle and a law of iterated logarithm. The original central limit theorem is formulated for independent time series, which is by definition not fulfilled for Poisson-INGARCH processes. For strongly mixing time series on the other hand, there exist similar results. The asymptotically vanishing dependence yields a kind of central limit theorem for mixing processes which we call invariance principle.

Theorem 3.11. *Let $Y = (Y_i)_{i \in \mathbb{N}_0}$ be a stationary time series with $E[Y_i] = 0$ for $i \in \mathbb{N}_0$, having $(2+\delta)$th moments uniformly bounded by 1 for some $0 < \delta \leq 1$ and being strongly mixing, with*

$$\alpha_Y(n) = O\left(n^{-(1+\varepsilon)(1+2/\delta)}\right)$$

for some $\varepsilon > 0$. Then, the series

$$\sigma^2 = \mathrm{E}[Y_0^2] + 2\sum_{i=1}^{\infty} \mathrm{E}[Y_0 Y_i]. \tag{3.6}$$

converges absolutely. Then, we can redefine the time series $(Y_i)_{i \in \mathbb{N}_0}$ on a new probability space together with a Brownian motion W_t such that

$$\sum_{i=0}^{n} Y_i - W_n = O_P\left(n^{1/2-\lambda}\right) \tag{3.7}$$

for some $\lambda > 0$ depending on ε and δ only.

For a proof of this theorem, see Theorem 4 in Kuelbs and Philipp [1980]. Note that there the results are more general, given for weakly stationary and $\mathbb{R}^d$-valued time series for some $d \in \mathbb{N}$. For a definition of the stochastic Landau symbols and related properties, see Appendix B, and for a definition of the Brownian motion, see Appendix C. The law of iterated logarithm can also be generalized to strongly mixing time series.

Theorem 3.12. *Let $Y = (Y_i)_{i\in\mathbb{N}_0}$ be a time series with $E[Y_i] = 0$ for $i \in \mathbb{N}_0$, having $(2+\delta)$th moments uniformly bounded by 1 for some $0 < \delta \leq 1$ and being strongly mixing, with*

$$\alpha_Y(n) = O\left(n^{-(1+\varepsilon)(1+2/\delta)}\right)$$

for some $\varepsilon > 0$. Then,

$$\mathrm{E}\left[\left|\sum_{i=k+1}^{k+n} Y_i\right|^2\right] \leq n$$

for $k \in \mathbb{N}_0$ and $n \in \mathbb{N}$ and

$$\limsup_{n\to\infty}(n \log\log n)^{-1/2}\left|\sum_{i=0}^{n} Y_i\right| \leq 200(\varepsilon\delta)^{-1/2}.$$

We refer to Lemma 2.3 and Theorem 6 in Dehling and Philipp [1982] for a proof of this law of iterated logarithm. With these concepts all gathered we can now introduce the time series model that is used for change point problems considered in this work.

Remark 3.13. Theorem 3.11 in its original formulation needs the process to be weakly stationary. This weak form of stationarity means, that the process has a constant mean, a finite constant variance and time invariant covariances, i.e. $Cov(Y_i, Y_{i+j}) = \gamma_j$ for $i, j \in \mathbb{N}_0$ and $\gamma_j \in \mathbb{R}$. Since the process we are working with will be stationary in the sense of Definition 3.1 and has a finite variance, we get also weak stationarity, see the discussion of Definition 11.6 in Franke et al. [2010]. Thus, for simplicity, we only consider stationarity as in Definition 3.1.

3.3 The Poisson-INGARCH(p,q) process

As discussed in Section 3.1, we use a model based on the GARCH model. To achieve an integer valued time series, the model is defined based on a Poisson distribution. The intensity of the Poisson distribution is modeled as a time series which is dependent on previous observations.

Definition 3.14. Let $(Y_i)_{i\in\mathbb{N}_0}$ be a time series of counts, together with the process $(\lambda_i)_{i\in\mathbb{N}_0}$ and $\mathcal{F}_i = \sigma(\lambda_0, ..., \lambda_i, Y_0, .., Y_i)$ the σ-algebra generated by the two

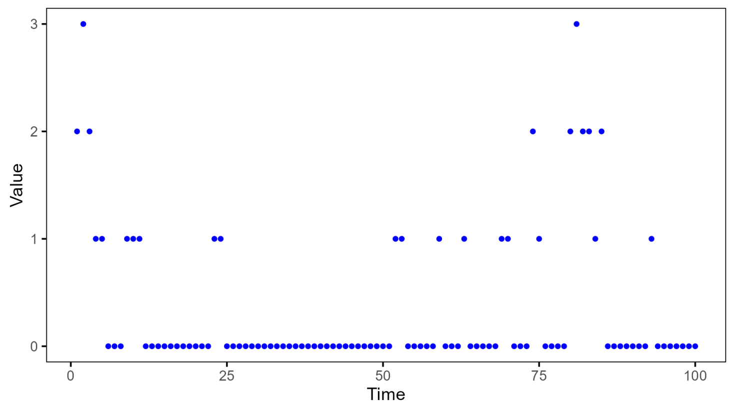

Figure 3.1: Sample path of a Poisson-INARCH(1) process with linear intensity $\lambda_i = 0.15 + 0.5 \cdot Y_{i-1}$

processes until time i. For $p, q \in \mathbb{N}$ the process is called Poisson-INGARCH(p,q) if

$$Y_i | \mathcal{F}_{i-1} \sim Poi(\lambda_i)$$

and

$$\lambda_i = g(Y_{i-1}, ..., Y_{i-p}, \lambda_{i-1}, ..., \lambda_{i-q}),$$

for some function $g : \mathbb{N}_0^p \times \mathbb{R}_+^q \to \mathbb{R}_+$. If $q = 0$, the process is called Poisson-INARCH(p) process.

The values of the process are by construction integer valued since every random variable is defined via a Poisson distribution. The dependence on previous realizations of both, the process itself and the intensities, is given by the function g which we will call intensity function. At first there are no restrictions on this intensity function. For the process to have some kind of regularity, there will be the need to define a specific property of this function. Why this is necessary, we will see by inspecting sample paths of Poisson-INARCH(1) processes with different linear intensities.

Example 3.15. We want to verify the assumed properties of the Poisson-INGARCH process in comparison with independent Poisson distributed observations. The sample path of a Poisson-INARCH(1) process is shown in Figure 3.1. Here, the intensity function is linear in Y_{i-1}, i.e., $g(y) = \omega + \alpha y$, with an additive part $\omega = 0.15$ and multiplicative part $\alpha = 0.5$. The result is a seemingly stationary process. The existence of autocorrelation is not obvious on first glance. Although, compared to independent Poisson-distributed samples, see Figure 3.2, the dependence structure of the Poisson-INARCH process gets more recognizable. Here, we set the intensity to 0.3 which is as of Lemma A.2 equal to the expectation of a Poisson distributed random variable. As we will see later, this is equal to the expectation of the process with an intensity function as given above,

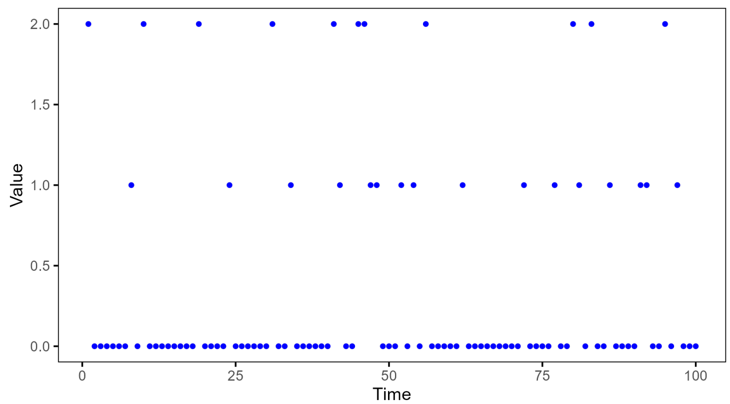

Figure 3.2: Samples of independent Poisson distributed random variables with intensity $\lambda = 0.3$

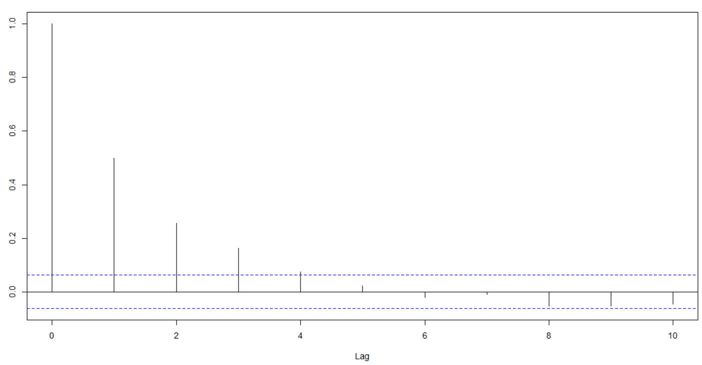

Figure 3.3: Sample autocorrelation of a Poisson-INARCH(1) process with linear intensity $\lambda_i = 0.15 + 0.5 \cdot Y_{i-1}$ and a sample length of 1000.

making both samples comparable. The independently distributed samples look more equally distributed in comparison to the Poisson-INARCH(1). There are no blocks of successive time points where the process is not equal to 0. This hints to clustering effects in the Poisson-INGARCH model, as would be expected from its real valued counterpart, the GARCH model. The sample autocorrelations and partial autocorrelations for ten lags of a process of this type are displayed in Figure 3.3 and Figure 3.4. It starts with a lag-one autocorrelation of 0.499 and starts decreasing in an exponential way. The partial autocorrelation hints to a time series, where each observations is depending mainly on the last observation and less on the ones before that. This gives the impression that a process of this form fulfills the desired properties of being a stationary count time series with a correlation between consecutive time points. Consider in comparison the

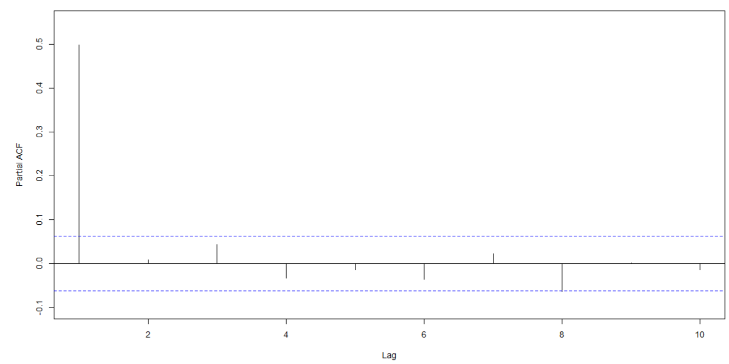

Figure 3.4: Sample partial autocorrelation of a Poisson-INARCH(1) process with linear intensity $\lambda_i = 0.15 + 0.5 \cdot Y_{i-1}$ and a sample length of 1000.

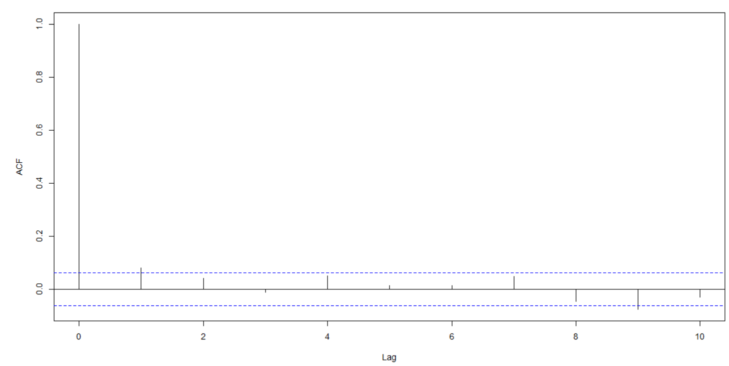

Figure 3.5: Sample autocorrelation of 1000 independent distributed Poisson-distributed random variables with intensity 0.3.

same plots regarding independent Poisson-distributed observations, the respective autocorrelations in Figure 3.5 and partial autocorrelations in Figure 3.6. There exists a clear difference, as the correlations for the independent random variables are all close to zero.

Example 3.16. Next we examine a different Poisson-INARCH(1) process, also with a linear intensity. Here, the multiplicative parameter has with $\alpha = 1.03$ a value of just above 1, with a sample path displayed in Figure 3.7. Resulting in a process which at some point starts increasing exponentially, since

$$\begin{aligned} \mathrm{E}[Y_i] &= \mathrm{E}[\mathrm{E}[Y_i|\mathcal{F}_{i-1}]] = \mathrm{E}[g_\theta(Y_{i-1})] = 0.15 + 1.03\mathrm{E}[Y_{i-1}] \\ &= 1.03^i + 0.15\frac{1 - 1.03^i}{0.03}. \end{aligned}$$

Thus, for a multiplicative parameter larger than 1, we get an explosive behavior

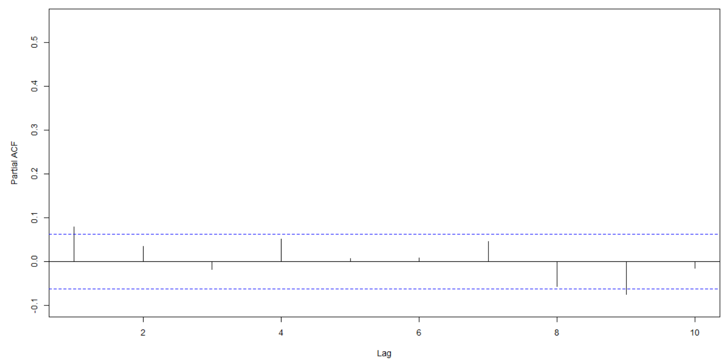

Figure 3.6: Sample partial autocorrelation of 1000 independent distributed Poisson-distributed random variables with intensity 0.3.

and obviously no stationarity.

To get stationarity, it seems that we need to restrict the intensity function to below 1 in a linear case. In a more general setting of the intensity function, this translates to the following property.

Definition 3.17. A function $g : \mathbb{N}_0^p \times \mathbb{R}_+^q \to \mathbb{R}$ is fulfilling the contraction property, if for all $y, \tilde{y} \in \mathbb{R}_+^p$ and $\lambda, \tilde{\lambda} \in \mathbb{N}_0^q$ it holds

$$\left| g(y, \lambda) - g(\tilde{y}, \tilde{\lambda}) \right| \leq \sum_{j=1}^{p} a_j \left| y_j - \tilde{y}_j \right| + \sum_{j=1}^{q} b_j \left| \lambda_j - \tilde{\lambda}_j \right| \tag{3.8}$$

for positive constants $a_1, .., a_p, b_1, ..., b_q$ with $\sum_{j=1}^{p} a_j + \sum_{j=1}^{q} b_j = L < 1$.

So, the intensity function has to be Lipschitz continuous in each argument and the sum of those Lipschitz constants has to be smaller than 1. In the linear case these Lipschitz constants are equal to the multiplicative parameter for each argument. Meaning that in the linear Poisson-INARCH(1) case, this parameter is restricted to values between 0 and 1. Two of the main consequences which are implied by this contraction property are stationarity and absolute regularity of the time series. Those results and other implications are studied in the Section 3.4.

Remark 3.18. Note that there are also other ways of modeling count time series with autocorrelation, for example INAR(1) type processes. There, the recursion is more similar to that in Definition 3.2, where the equivalents to the parameters α_j and β_j are not factors but parameters in a so called binomial thinning. That allows for the possibility to choose those parameters as real values while guaranteeing that the time series stays integer valued. For more information on other types of count time series models, see Weiß [2018].

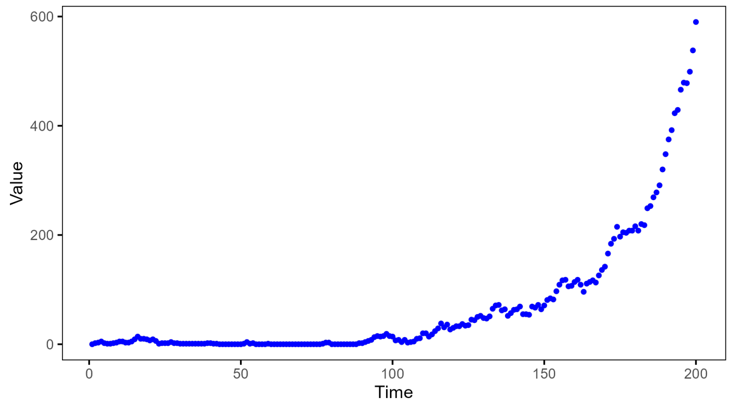

Figure 3.7: Sample path of a Poisson-INARCH(1) process with linear intensity $\lambda_i = 0.15 + 1.03 \cdot Y_{i-1}$

Remark 3.19. The Poisson-INGARCH process should not be confused with the Poisson process - although named alike. The Poisson-INGARCH is, in line of the arguments above, a type of an integer valued GARCH process where the integer valued distribution is a Poisson distribution. That is what distinguishes it from the Poisson process which is a process in continuous time with the difference of the process between two time points $t_1, t_2 \in \mathbb{R}_+$, $t_1 < t_2$ being Poisson distributed with intensity $\lambda(t_2 - t_1)$ for some $\lambda > 0$.

3.4 Consequences of the contraction property

We can use the contraction property (3.8) to deduce properties of the Poisson-INGARCH process. In this section, we present some results which are implied by the contraction property, starting with stationarity of the process.

Theorem 3.20. *Let $(Y_i)_{i \in \mathbb{N}_0}$ be a Poisson-INGARCH(p,q) process with intensity function $g : \mathbb{N}_0^p \times \mathbb{R}_+^q \to \mathbb{R}$ fulfilling the contraction property (3.8). Then, the time series $(Y_i)_{i \in \mathbb{N}_0}$ is stationary with finite mean $\mathrm{E}[Y_0] < \infty$.*

This result in the case of a linear intensity function with general $p, q \in \mathbb{N}$ has first been proven in Proposition 1 of Ferland et al. [2006]. The generalization to arbitrary intensity functions fulfilling the contraction property with $p = q = 1$ has been shown in Theorem 2.1 of Neumann [2011]. Finally, in Franke [2010] this result has been generalized to arbitrary intensity functions fulfilling the contraction property where $p, q \in \mathbb{N}$.

Example 3.21. If $(Y_i)_{i \in \mathbb{N}_0}$ is a linear Poisson-INARCH(1) process we are now able to calculate the mean and variance. For $g(y) = \omega + \alpha y$ with $\omega > 0$ and

$0 < \alpha < 1$, we get a stationary time series. Let $i \in \mathbb{N}$, then

$$\begin{aligned}\mathrm{E}[Y_i] &= \mathrm{E}[\mathrm{E}[Y_i|\mathcal{F}_{i-1}]] = \mathrm{E}[\lambda_i] = \mathrm{E}[g(Y_{i-1})] = \omega + \alpha\mathrm{E}[Y_{i-1}] \\ &= \omega + \alpha\mathrm{E}[Y_i]\end{aligned}$$

which we can use to calculate

$$\mathrm{E}[Y_i] = \frac{\omega}{1-\alpha}. \tag{3.9}$$

With Lemma A.2(i), we have

$$\begin{aligned}\mathrm{Var}(Y_i) &= \mathrm{Var}(\mathrm{E}[Y_i|\mathcal{F}_{i-1}]) + \mathrm{E}[\mathrm{Var}(Y_i|\mathcal{F}_{i-1})] = \mathrm{Var}(g(Y_{i-1})) + \mathrm{E}[Y_i] \\ &= \mathrm{Var}(\alpha Y_{i-1} + \omega) + \mathrm{E}[Y_i] = \alpha^2\mathrm{Var}(Y_i) + \frac{\omega}{1-\alpha},\end{aligned}$$

implying that

$$\mathrm{Var}(Y_i) = \frac{\omega}{(1-\alpha)(1-\alpha^2)}. \tag{3.10}$$

So firstly, we can now confirm what we proposed in Example 3.15: for parameters $\omega = 0.15$ and $\alpha = 0.5$ we get an expectation of 0.3, equal to the expectation for the independent Poisson-distributed random variables. Secondly, the variance is larger than the expectation for linear Poisson-INARCH(1) processes. Consequently, the stationary distribution is overdispersed and can not be a Poisson distribution. Since for Poisson distributions the variance and expectation are equal. Lastly, the covariance of consecutive random variables can be calculated as

$$\begin{aligned}\mathrm{Cov}(Y_i, Y_{i+1}) &= \mathrm{E}[Y_iY_{i+1}] - \mathrm{E}[Y_i]\mathrm{E}[Y_{i+1}] = \mathrm{E}[\mathrm{E}[Y_iY_{i+1}|\mathcal{F}_i]] - \mathrm{E}[Y_i]\mathrm{E}[Y_i] \\ &= \mathrm{E}[Y_i\mathrm{E}[Y_{i+1}|\mathcal{F}_i]] - \mathrm{E}[Y_i]^2 = \mathrm{E}[Y_ig(Y_i)] - \mathrm{E}[Y_i]^2 \\ &= \omega\mathrm{E}[Y_i] + \alpha\mathrm{E}[Y_i^2] - \mathrm{E}[Y_i]^2 = \omega\mathrm{E}[Y_i] + \alpha\left(\mathrm{Var}(Y_i) + \mathrm{E}[Y_i]^2\right) - \mathrm{E}[Y_i]^2 \\ &= \omega\mathrm{E}[Y_i] + \alpha\mathrm{Var}(Y_i) - (1-\alpha)\mathrm{E}[Y_i]^2 \\ &= \frac{\omega^2}{1-\alpha} + \frac{\alpha\omega}{(1-\alpha)(1-\alpha^2)} - \frac{\omega^2}{1-\alpha} = \frac{\alpha\omega}{(1-\alpha)(1-\alpha^2)}.\end{aligned} \tag{3.11}$$

Dividing by the variance then leads us to the correlation of subsequent observations, being

$$\mathrm{Corr}(Y_i, Y_{i+1}) = \alpha. \tag{3.12}$$

Next, we see the second important implication of the contraction property, namely that it ensures the time series fulfills one of the mixing conditions from the preceding section. This has been proven in Neumann [2011].

Theorem 3.22. *Let $Y = (Y_i)_{i\in\mathbb{N}_0}$ be a Poisson-INGARCH(1,1) process with intensity function $g : \mathbb{N}_0 \times \mathbb{R}_+ \to \mathbb{R}$ fulfilling the contraction property (3.8). Then, the process is absolutely regular with*

$$\beta_Y(n) = O(L^n). \tag{3.13}$$

So, the count time series resulting from a Poisson-INGARCH(1,1) process fulfilling the contraction property is absolutely regular with an exponentially decreasing mixing rate. Since absolute regularity is a stronger condition than strong mixing, we can apply both Theorems 3.11 and 3.12. Note that the prerequisites regarding the mixing rate of these two theorems hold, because the exponentially decreasing mixing rate is always $O(n^{-c})$ for an arbitrary $c > 0$. For general intensity functions the mixing property only holds for the case of $p = q = 1$. That is because for larger values, mixing is hard to verify, if at all possible. Hence, in the following we will restrict ourselves to the case of both p and q being at most 1.

Remark 3.23. (i) Neumann [2021] showed such a result for linear Poisson-INGARCH(p,q) processes with $p, q \geq 1$. Since we are more interested in general intensity functions of Poisson-INARCH(1) processes, we only refer here to this result.

(ii) For Poisson-INGARCH(p,q) processes with a general intensity function and $p, q \geq 1$, there exists such a result based on another concept of asymptotically diminishing dependence, called weak dependence. For more on weakly dependent Poisson-INGARCH(p,q) processes we refer to Franke [2010].

We conclude this subsection with two more properties of a stationary Poisson-INGARCH(1,1)-process. The first one is about the consequences of the value of the intensity function having a constant part equal or unequal to 0.

Lemma 3.24. *Let $(Y_i)_{i\in\mathbb{N}_0}$ be a Poisson-INGARCH(1,1) process with intensity function $g : \mathbb{N}_0 \times \mathbb{R}_+ \to \mathbb{R}$ fulfilling the contraction property (3.8) and π the probability mass function, which coincides with the stationary distribution of $(Y_i)_{i\in\mathbb{N}_0}$. Then, the following statements hold true.*

(i) If $g(0, 0) = 0$ we have that $\pi(0) = 1$.

(ii) If $g(0, 0) > 0$ we have that $\pi(y) < 1$ for all $y \in \mathbb{N}_0$.

This Lemma is based on Theorem 2.1 (iii) in Neumann [2011]. The consequence of this Lemma is, that if the constant part of the intensity function is 0, the stationary distribution is constantly 0 with probability 1. Since we are not interested in constant time series, we will consider only intensity functions with $g(y, \lambda) \geq C > 0$. In the linear case this is fulfilled by assuming $\omega > 0$. Since by looking at the expectations in Example 3.21, we see that they would be equal to 0 if $\omega = 0$. For linear Poisson-INGARCH(1,1) processes there is an additional result about the existence of moments, see Ferland et al. [2006].

Theorem 3.25. *Let $(Y_i)_{i\in\mathbb{N}_0}$ be a Poisson-INGARCH(1,1) process with intensity function $g : \mathbb{N}_0 \times \mathbb{R}_+ \to \mathbb{R}$ fulfilling the contraction property (3.8). Then, all moments $\mathrm{E}[Y_0^k]$ for $k \in \mathbb{N}$ exist.*

This will later be important, since already from Theorems 3.11 and 3.12, we can conclude that the existence of higher moments is necessary at some point.

With this theorem, it will be much easier to show that such assumptions are fulfilled if we inspect the linear case.
Lastly, to wind up this section about count time series, we present some conditions under which even non-stationary Poisson-INGARCH(1,1) processes can be strongly mixing.

3.5 A more general case of mixing in Poisson-INARCH(1) processes

Recalling Theorem 3.20 and Theorem 3.13, we observe that the requirements for absolute regularity and stationarity overlap for Poisson-INGARCH(1,1) processes. If the intensity function is contractive we always get both. Thus, in the setting of the section beforehand, there cannot be a Poisson-INGARCH(1,1) process which is mixing but non-stationary. Since non-stationary time series are exactly what we want to model, we introduce another setting for a process $(Y_i)_{i\in\mathbb{N}_0}$ from Doukhan et al. [2022] on a probability space $(\Omega, \mathcal{F}, P)$ where for all $i \geq 0$

$$Y_i|\mathcal{F}_{i-1} \sim Poi(\lambda_i) \tag{3.14}$$

$$\lambda_i = g_i(Y_{i-1}, \lambda_{i-1}, Z_{i-1}) \tag{3.15}$$

with $\mathcal{F}_i = \sigma(\lambda_0, ..., \lambda_i, Y_0, .., Y_i, Z_0, ..., Z_i)$, some functions $g_i : \mathbb{N}_0 \times \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}_+$ and $(Z_i)_{i\in\mathbb{N}_0}$ is a sequence of $\mathbb{R}^d$-valued covariates. For $i \in \mathbb{N}_0$, Z_i is assumed to be independent of $\mathcal{F}_{i-1}$ and Y_i. This formulation of a Poisson-INGARCH(1,1) type process allows for non-stationarity and in particular potentially unbounded trends. For the strong mixing condition to hold, we require some conditions on the functions g_i. We will see, that if they are all fulfilled, we still have absolute regularity of the process.

(C1) There exists some $L_1 < 1$, such that the following condition is fulfilled: if $\lambda, \lambda' \geq 0$, $Y \sim Poi(\lambda)$ being independent of Z_i, then

$$\mathrm{E}\left[|g_i(Y, \lambda, Z_i) - g_i(Y, \lambda', Z_i)|\right] \leq L_1|\lambda - \lambda'|. \tag{3.16}$$

(C2) There exists some $L_2 < 1$, such that the following condition is fulfilled. If $\lambda, \lambda' \geq 0$, then there exists a coupling of (Y, Z) and (Y', Z'), with $Y \sim Pois(\lambda)$, $Y' \sim Pois(\lambda')$, $Z, Z' \overset{d}{=} Z_i$, Z being independent of Y and Z' being independent of Y', such that

$$\mathrm{E}\left[|g_i(Y, \lambda, Z) - g_i(Y', \lambda', Z')|\right] \leq L_2|\lambda - \lambda'|. \tag{3.17}$$

(C3) Let $(\tilde{\lambda}_i)_{i\in\mathbb{N}_0}$ and $(\tilde{\lambda}'_i)_{i\in\mathbb{N}_0}$ be two independent processes on a probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$ which have the same distribution as $(\lambda_i)_{i\in\mathbb{N}_0}$. Suppose there exist constants $L_3 < 1$ and $M_0, M_1 < \infty$ such that

  (i) $\tilde{\mathrm{E}}[|\tilde{\lambda}_0 - \tilde{\lambda}'_0|] \leq M_0$

  (ii) $\tilde{\mathrm{E}}\left[|\tilde{\lambda}_{i+1} - \tilde{\lambda}'_{i+1}||\tilde{\lambda}_i, \tilde{\lambda}'_i\right] \leq L_3|\tilde{\lambda}_i - \tilde{\lambda}'_i| + M_1.$

These conditions above are quite technical and were derived particularly to prove the strong mixing condition. Note that the second condition (C2) is analogous to the contraction property (3.8).

Remark 3.26. For simplicity we use the absolute value as a distance measure in these conditions. In Doukhan et al. [2022] this is formulated more generally for a set of distance measure fulfilling a certain property. The distance measure induced by the absolute value is one of them.

With these conditions all fulfilled, a Poisson-INGARCH(1,1) process of the form above still is absolutely regular with exponentially decreasing mixing rate.

Theorem 3.27. *Let $(Y_i)_{i\in\mathbb{N}_0}$ be a Poisson-INGARCH(1,1) process with an intensity as in (3.15) and (C1)-(C3) be satisfied with constants $L_1, L_2, L_3 < 1$ and $M_0, M_1 < \infty$. Then, $X = (Y_i, Z_i)_{i\in\mathbb{N}_0}$ is absolutely regular with*

$$\beta_X(n) \le L_2^{n-1} \frac{1}{1-L_1} \left(\frac{M_1}{1-L_3} + M_0 \right) \tag{3.18}$$

For a proof of this result, see Theorem 2.1 of Doukhan et al. [2022]. To get a perspective on the power of this theorem and what kind of intensity functions still yield a strongly mixing time series, consider the following corollary.

Corollary 3.28. *Let $(Y_i)_{i\in\mathbb{N}_0}$ be a Poisson-INGARCH(1,1) process with intensity $\lambda_{i+1} = a_i Y_i + b_i \lambda_i + Z_i$ for $a_i, b_i \ge 0$ and $i \in \mathbb{N}_0$. Moreover, suppose $L_2 = \sup_{i\in\mathbb{N}_0} a_i + b_i < 1$, $\sup_{i\in\mathbb{N}_0} \mathrm{E}[\sqrt{Z_i} - \mathrm{E}[\sqrt{Z_i}]] < \infty$ where Z_i is a non-negative random variable (covariate) which is independent of $\lambda_0, ..., \lambda_i, Y_0, .., Y_i, Z_0, ..., Z_{i-1}$ and $\mathrm{E}[\lambda_0] < \infty$. Then, $X = (Y_i, Z_i)_{i\in\mathbb{N}_0}$ is absolutely regular with*

$$\beta_X(n) = O(\rho^n). \tag{3.19}$$

for some $0 < \rho < 1$.

This can also be found in Doukhan et al. [2022] as Corollary 3.2. The only restriction on the Z_i's here is that the absolute mean of the centralized expectation of it's square root is uniformly bounded. Since this is always 0 for deterministic Z_i, we can see that this model even allows for unboundedly increasing deterministic time dependent functions.
With all these properties of a Poisson-INGARCH(1,1) process with a contractive intensity function gathered, we briefly review another way of proving stability and mixing properties of time series by treating them as Markov chains.

4 Mixing and stationarity of Markov chains

Besides the contraction property, there are other ways of showing a strong mixing property or stationarity of Poisson-INARCH processes. By applying results from Markov chain theory we can circumvent the contraction property. Thus, we are able to prove those two properties for Poisson-INARCH processes with intensity functions that are not contractive. We start this by defining what Markov chains are.

Definition 4.1. The time series $(Y_i)_{i\in\mathbb{N}_0}$ taking values in the probability space $(\Omega, \mathcal{F}, P)$ with $\Omega = \prod_{i=0}^{\infty} X$ is called (countable space) Markov chain if for every $n \in \mathbb{N}_0$ and any sequence of states $\{y_0, ..., y_n\} \in \mathbb{N}_0^n$,

$$P_\mu(Y_0 = y_0, ..., Y_n = y_n) = \mu(y_0)P_{y_0}(Y_1 = y_1)...P_{y_{n-1}}(Y_y = y_n). \tag{4.1}$$

The probability space $(\Omega, \mathcal{F}, P)$ is called path space.
The probability measure μ is called initial probability of the chain.
The process $(Y_i)_{i\in\mathbb{N}_0}$ is a time-homogeneous Markov chain if the probabilities $P_{y_i}(Y_{i+1} = y_{i+1}) := P(Y_{i+1} = y_{i+1}|Y_i = y_i)$ depend only on the values of $y_i, y_{i+1} \in X$ and are independent of the time points $i \in \mathbb{N}_0$.

From this definition, in particular (4.1), we get an important property of Markov chains: the joint distribution of $Y_0, .., Y_n$ is given by the respective distributions given the random variable one step before. If we assume a time-homogeneous Markov Chain and set $P(x, y) := P_x(Y_1 = y)$, (4.1) can be written as

$$P_\mu(Y_0 = y_0, ..., Y_n = y_n) = \mu(y_0)P(y_0, y_1)...P(y_{n-1}, y_n), \tag{4.2}$$

or for the conditional probability of the process, we have

$$P_\mu(Y_{n+1} = y_{n+1}|Y_n = y_n, ..., Y_0 = y_0) = P(y_n, y_{n+1}), \tag{4.3}$$

what we call the Markov property. Formally, we can define the conditional probabilities $P(x, y)$ as in the following.

Definition 4.2. The matrix $P = \{P(x, y)|x, y \in X\}$ is called Markov transition matrix if for $x, y \in X$

$$P(x, y) \geq 0, \qquad \sum_{z\in X} P(x, z) = 1. \tag{4.4}$$

If we set $P^0(x, y) = \mathbb{1}_x(y)$ and inductively define

$$P^n(x, z) = \sum_{y\in\mathbb{N}_0} P(x, y)P^{n-1}(y, z) \tag{4.5}$$

we can compute the distribution of the process at time n given Y_0 as

$$P^n(x, y) = P_\mu(Y_n = y|Y_0 = x). \tag{4.6}$$

Hence, we call P^n the n-step transition matrix. For an arbitrary set $A \subset X$ we set $P^n(y, A) := \sum_{x \in A} P^n(y, x)$. If we do not condition the process on a single state, i.e., $Y_0 = x \in X$, but an initial distribution, we set

$$P_\mu(Y_n \in A) = \sum_{y \in X} \mu(y) P^n(y, A). \tag{4.7}$$

Before we continue with stating the results necessary for proving stationarity of a Markov chain, we need to gather some useful properties of Markov chains. Beginning with smallness and petiteness of a set.

Definition 4.3. A set $C \in \mathcal{B}(X)$ is called a small set if there exists an $m > 0$, and a non-trivial measure ν_m on $\mathcal{B}(X)$, i.e. $\nu_m(X) > 0$, such that for all $y \in C$, $B \in \mathcal{B}(X)$,

$$P^m(y, B) \geq \nu_m(B). \tag{4.8}$$

When (4.8) holds, we say that C is ν_m-small.

Definition 4.4. Let μ be a probability measure on $\mathbb{N}$ and define for $y \in X$ and $A \in \mathcal{B}(X)$

$$K_\mu(y, A) := \sum_{n=0}^{\infty} P^n(y, A) \mu(n). \tag{4.9}$$

We call a set $C \in \mathcal{B}(X)$ ν_μ-petite if for all $y \in C$, $B \in \mathcal{B}(X)$

$$K_\mu(y, B) \geq \nu_\mu(B) \tag{4.10}$$

where ν_μ is a non-trivial measure on $\mathcal{B}(X)$.

Both are technical properties of Markov Chains and we refer the reader to Meyn and Tweedie [2012] for further insights on Markov chain theory. One property that we will make use of is the relation of both properties to one another, as can be seen in Proposition 5.5.2 of Meyn and Tweedie [2012].

Proposition 4.5. *If $C \in \mathcal{B}(X)$ is ν_m-small, then C is ν_{δ_m} petite for $\delta_m(n) = \mathbb{1}_m(n)$ for $n, m \in \mathbb{N}$.*

So, if we have a small set with respect to a Markov chain it is also petite. The next necessary property is irreducibility.

Definition 4.6. The Markov chain $(Y_i)_{i \in \mathbb{N}_0}$ is called φ-irreducible, if there exists a measure φ on the Borel σ-algebra $\mathcal{B}(X)$ such that, whenever $\varphi(A) > 0$ for $A \in \mathcal{B}(\mathbb{N}_0)$, we have

$$L(y, A) := P\left(\min_{i \in \mathbb{N}} \{Y_i \in A | Y_0 = y\} < \infty\right) > 0 \tag{4.11}$$

for all $y \in \mathbb{N}_0$.

If this definition is fulfilled, the probability of reaching a set $A \in \mathcal{B}(\mathbb{N}_0)$ from any given value $y \in X$ of the chain in finite time is positive. With the restriction that A is not a null set with respect to the measure φ. The inversion of that property maybe helps with the intuition: If there is a set that can not be reached with positive probability in finite time, we could reduce X by that set A without changing the chain itself. Hence, the state space could be reduced, or in other words: the state space is not irreducible. Next, we introduce a property, which is concerned with the amount of time a chain returns to a certain value, called recurrence.

Definition 4.7. (i) The number of times a Markov chain visits a set $A \in \mathcal{B}(\mathbb{N}_0)$ is defined as the random variable

$$\eta_A := \sum_{n=1}^{\infty} \mathbb{1}_A(Y_n). \tag{4.12}$$

(ii) A state $\boldsymbol{\alpha} \in X$ of a Markov chain $(Y_i)_{i \in \mathbb{N}_0}$ is called recurrent if

$$E_\alpha[\eta_\alpha] := \mathrm{E}[\eta_\alpha | Y_0 = \alpha] = \infty. \tag{4.13}$$

The chain is called recurrent if all states are recurrent.

(iii) The set $A \in \mathcal{B}(X)$ is called Harris recurrent if for all $y \in A$

$$P_y(\eta_A = \infty) = 1. \tag{4.14}$$

A Markov chain $(Y_i)_{i \in \mathbb{N}_0}$ is called Harris (recurrent) if it is φ-irreducible and every set in $\mathcal{B}^+(X) := \{A \in \mathcal{B}(X) | \varphi(A) > 0\}$ is Harris recurrent.

The difference of those two definitions is mainly that normal recurrence requires that, in expectation, the number of times a chain arrives at each state is infinite. Harris recurrence on the other hand requires that the probability of accessing each set infinitely often is 1, with the restriction that the sets should not be null sets with respect to ν_m. Moreover, Harris recurrence can be applied on general state spaces which are not necessarily countable.

Remark 4.8. If $(Y_i)_{i \in \mathbb{N}_0}$ is a Harris recurrent Markov chain on a countable state space X and $\varphi(\alpha) > 0$ for all $\alpha \in X$, then it is also recurrent, since, if $P_\alpha(\eta_\alpha = \infty) = 1$, then $E_\alpha[\eta_\alpha] = \infty$ for all $\alpha \in X$.

These properties of Markov chains are connected, as demonstrated by the following Proposition, see Proposition 9.1.7 (ii) in Meyn and Tweedie [2012].

Proposition 4.9. *Let $(Y_i)_{i \in \mathbb{N}_0}$ be a φ-irreducible Markov chain. If there exists some petite set $C \in \mathcal{B}(X)$ such that for all $y \in X$, $L(y, C) = 1$, then $(Y_i)_{i \in \mathbb{N}_0}$ is Harris recurrent.*

With these rather technical properties gathered, we now move on to the concept which guarantees stationarity, or at least convergence to a stationary distribution.

Definition 4.10. A σ-finite measure π on $\mathcal{B}(X)$ with the property that for all $A \in \mathcal{B}(X)$

$$\pi(A) = \int_X \pi(dy)P(y, A) = \sum_{y \in X} \pi(y)P(y, A) \tag{4.15}$$

is called invariant.

If we assume that the initial distribution of a Markov chain $(Y_i)_{i \in \mathbb{N}_0}$ is an invariant probability distribution π of X, we get for $A \in \mathcal{B}(X)$ and $n \in \mathbb{N}$

$$\begin{aligned}
\pi(A) &= \sum_{y \in X} \pi(y)P(y, A) = \sum_{y \in X} \sum_{x \in X} \pi(x)P(x, y)P(y, A) \\
&= \sum_{x \in X} \pi(x) \sum_{y \in X} P(x, y) \sum_{z \in A} P(y, z) = \sum_{x \in X} \pi(x) \sum_{z \in A} \sum_{y \in X} P(x, y)P(y, z) \\
&= \sum_{x \in X} \pi(x) \sum_{z \in A} P^2(x, z) = \sum_{x \in X} \pi(x)P^2(x, A) \\
&\dots \\
&= \sum_{x \in X} \pi(x)P^n(x, A) = P_\pi(Y_n \in A).
\end{aligned} \tag{4.16}$$

Using (4.15) and the definition of the n-step probability transition matrix (4.5) for a set $A \in \mathcal{B}(X)$. The Markov property for $y_n, ..., y_{n+k} \in X$ where $n, k \in \mathbb{N}$ on the joint distributions of $Y_n, ..., Y_{n+k}$ yields

$$P_\pi(Y_n = y_n, ..., Y_{n+k}) = P_\pi(Y_n = y_n)P(y_n, y_{n+1})...P(y_{n+k-1}, y_{n+k}). \tag{4.17}$$

Equality for all $n, k \in \mathbb{N}$ holds, if the distribution of Y_n is independent of n. In the case of the invariant probability distribution being the initial distribution, the distribution of Y_n is equal to π. In this case the chain is stationary. To prove that a Markov chain is stationary, it is thus sufficient to prove existence and uniqueness of an invariant distribution and that the initial distribution is equal to the invariant distribution. Conditions under which this holds are provided in the following Theorem.

Theorem 4.11. *Let $(Y_i)_{i \in \mathbb{N}_0}$ be a recurrent Markov chain. Then, it admits a unique (up to constant multiples) invariant measure π. The invariant measure is finite (rather than merely σ-finite), if there exists a petite set $C \in \mathcal{B}(X)$ such that*

$$\sup_{y \in C} E_y[\tau_C] < \infty$$

for $\tau_C = \min\{n \geq 1 : Y_n \in C\}$.

We refer to Theorem 10.0.1 in Meyn and Tweedie [2012] for a proof. If the latter part of the theorem holds true, finiteness of an invariant measure $\tilde{\pi}$ implies that we can define a probability measure

$$\pi(A) := \frac{\tilde{\pi}(A)}{\tilde{\pi}(X)}, \tag{4.18}$$

which is still invariant, since we can factor constant multiples out in equation(4.15). So, there must exist a unique invariant probability measure. Now, we know such an invariant probability measure exists. To get stationarity as in Definition 3.1, we needed the initial distribution to already be the stationary distribution. To get more properties of the invariant probability measure, we need another concept which excludes cyclic behavior.

Definition 4.12. A φ-irreducible Markov chain on a countable space is called aperiodic, if

$$d(x) := g.c.d.\{n \geq 1 : \; P^n(x,x) > 0\} = 1 \tag{4.19}$$

for all $x \in X$.

If a state $y \in X$ would only be attainable every d-th time step for $d \in \mathbb{N}$, a chain could not be stationary. Since the distribution of Y_d and Y_{d+1} would not be equal. Requiring aperiodicity on a stationary Markov chain yields the convergence to an invariant measure, see Theorem 13.0.1 in Meyn and Tweedie [2012].

Theorem 4.13. *Let $(Y_i)_{i \in \mathbb{N}_0}$ be an aperiodic Harris recurrent chain, with invariant measure π. Then, the invariant measure π is finite if and only if π is an unique invariant probability measure such that for every initial condition $x \in X$,*

$$\sup_{A \in \mathcal{B}(X)} |P^n(x,A) - \pi(A)| \overset{n \to \infty}{\longrightarrow} 0. \tag{4.20}$$

Besides just knowing that the invariant probability measure exists and is unique, we get that for every initial condition, the n-step probability transition matrix converges to the invariant probability measure. That helps us for example when we want to simulate stationary time series and we do not know the invariant measure. Just starting with a random initial value, the time series will asymptotically be in the stationary distribution, after a sufficient number of steps. We then let the sample begin from this step and have simulated a stationary time series.
Also, in this Markov chain setting we are interested in mixing properties of Markov chains. In the introduction of the strong mixing properties, we have already mentioned the relation to the theory of ergodicity. In Meyn and Tweedie [2012] different kinds of ergodicity for Markov chains are discussed. Here, we are concerned with the kind of ergodicity which is closest related to the strong mixing properties, called uniform ergodicity.

Definition 4.14. (i) Let ν be a signed measure, i.e., a measure that can take on negative values. Then, the total variation norm is defined as

$$\|\nu\|_{TV} := \sup_{A \in \mathcal{B}(X)} \nu(A) - \inf_{A \in \mathcal{B}(X)} \nu(A). \tag{4.21}$$

(ii) Based on the total variation norm we can define the distance between a

Markov transition matrix P and a measure π as

$$|||P - \pi||| := \sup_{y \in X} \|P(x, \cdot) - \pi\|_{TV} . \tag{4.22}$$

(iii) A Markov chain $(Y_i)_{i\in\mathbb{N}_0}$ with transition matrix P and an invariant measure π with

$$|||P^n - \pi||| \overset{n\to\infty}{\longrightarrow} 0 \tag{4.23}$$

is called uniformly ergodic.

To use uniform ergodicity, we first need a condition which is easier to prove than the ergodicity itself, given by the following result which is part of Theorem 16.0.2 in Meyn and Tweedie [2012].

Theorem 4.15. *Let $(Y_i)_{i\in\mathbb{N}_0}$ be a Markov chain. Then, the chain is uniformly ergodic if and only if the state space X is ν_m-small for some $m \in \mathbb{N}$.*

With this condition we can reduce uniform ergodicity to smallness of the state space X. This is helpful, since by showing this for a Markov chain, we also get finiteness in Theorem 4.11. The connection of uniform ergodicity is given by the following equivalence, see Theorem 16.1.5 from Meyn and Tweedie [2012].

Theorem 4.16. *Let $(Y_i)_{i\in\mathbb{N}_0}$ be an uniformly ergodic Markov chain. Then, there exist $0 < R < \infty$ and $0 < \rho < 1$ such that for any measurable $g, h : X \to \mathbb{R}$ with $g^2, h^2 \leq 1$, $y \in X$ and $k, n \in \mathbb{N}$,*

$$|\mathrm{E}[g(Y_k)h(Y_{n+k})] - \mathrm{E}[g(Y_k)]\,\mathrm{E}[h(Y_{n+k})]| \leq R\rho^n(1 + \rho^k). \tag{4.24}$$

The chain is then called geometrically mixing.

Since we cannot use geometrically mixing itself as a property further on, we prove the following relationship between geometrically mixing and strongly mixing.

Lemma 4.17. *Let $Y = (Y_i)_{i\in\mathbb{N}_0}$ be a geometrically mixing Markov chain. Then, Y is strongly mixing with*

$$\alpha_Y(n) \leq R\rho^n(1 + \rho^n) \tag{4.25}$$

for some $0 < \rho < 1$ and $0 < R < \infty$.

Proof. Let $i, n \in \mathbb{N}$. For $A \in \mathcal{F}_0^i$ and $B \in \mathcal{F}_{i+n}^\infty$, $\mathbb{1}_A$, $\mathbb{1}_B$ are measurable, non-negative and smaller or equal than 1. Hence, the mixing condition from (4.24) can be applied with $P(A) = \mathrm{E}[\mathbb{1}_A]$:

$$\begin{aligned}
\alpha_Y(n) &= \sup_{k\in\mathbb{N}} \sup_{A\in\mathcal{F}_0^i, B\in\mathcal{F}_{i+n}^\infty} |P(A\cap B) - P(A)P(B)| \\
&= \sup_{k\in\mathbb{N}} \sup_{A\in\mathcal{F}_0^i, B\in\mathcal{F}_{i+n}^\infty} |\mathrm{E}[\mathbb{1}_A\mathbb{1}_B] - \mathrm{E}[\mathbb{1}_A]\mathrm{E}[\mathbb{1}_B]| \\
&\leq \delta(n) \leq R\rho^n(1 + \rho^k) \xrightarrow{n\to\infty} 0
\end{aligned}$$

for some $0 < \rho < 1$ and $0 < R < \infty$. Hence, the α-mixing condition from Bradley [2002] holds. □

With these results gathered, we are able to show similar properties of Poisson-INGARCH processes as in the section before, for some intensity functions that are not necessarily contractive. So, we can now move on to the next section, reviewing the theory on consistency of M-estimators.

5 Identifiability and laws of large numbers

Another challenge that we will have to handle to guarantee asymptotics of our test statistics will be parameter estimation. In our example in Section 2 we effectively already estimated a parameter: the expectation of a normal distribution via the sample mean. Later we need to estimate parameters of a parametric intensity function, for which the convergence of the estimators is not as clear as in the introductory example. Thus, we will present some theory for parameter estimation in this section. We begin with the general idea and results for stationary time series. The following subsections will expand this theory to some results of which we can use in the non-stationary case.

5.1 The identifiability condition and consistency for stationary time series

Specifically, we are interested in the convergence of the so called conditional least squares (CLS) estimators which we get by minimizing a conditional least squares function. For a time series $(X_i)_{i\in\mathbb{N}_0}$ and an adapted filtration $(\mathcal{F}_i)_{i\in\mathbb{N}_0}$ the conditional least squares function is given by

$$\begin{aligned} Q_n :\ &\Theta \to \mathbb{R} \\ &\theta \mapsto n^{-1}\sum_{i=1}^{n}(X_i - \mathrm{E}[X_i|\mathcal{F}_{i-1},\theta])^2 \end{aligned} \tag{5.1}$$

as introduced in Lawrence and Paul [1978]. More generally, estimators of this type are part of a class of M-estimators, a term coined by Huber [1964]. Then, the function Q_n is generalized to

$$\begin{aligned} Q_n :\ &\Theta \to \mathbb{R} \\ &\theta \mapsto n^{-1}\sum_{i=1}^{n} q(\theta, X_i), \end{aligned} \tag{5.2}$$

for a suitable measurable function $q : \Theta \times \mathbb{R} \to \mathbb{R}$. If we set for example $q(\theta, X_i) = -\log(f(X_i|\theta))$ where, the function f is a probability density function parameterized by θ, we get maximum likelihood estimators. Proving the convergence of such minimizers almost surely or in probability is usually not an easy task. Hence, a large toolbox was developed to approach this problem and for further reading on this topic, we refer to Pötscher and Prucha [1997]. Firstly, we have to be sure that there exists an unique minimizer, at least asymptotically. The property that guarantees this is called identifiability, see Definition 3.1 in Pötscher and Prucha [1997]. It is defined for a function $\overline{Q}_n(\theta)$, usually being the expectation of $Q_n(\theta)$.

Definition 5.1. For a given sequence of functions $\overline{Q}_n : \Theta \to \mathbb{R}$ the sequence of minimizers $\tilde{\theta}_n$ of $\overline{Q}_n(\theta)$ is called identifiably unique, if for every $\varepsilon > 0$

$$\liminf_{n\to\infty} \left[\inf_{\theta\in\Theta: \left\|\theta-\tilde{\theta}_n\right\| \geq \varepsilon} \overline{Q}_n(\theta) - \overline{Q}_n(\tilde{\theta}_n) \right] > 0. \tag{5.3}$$

This definition yields two properties of the function $\overline{Q}_n$. One is the already stated uniqueness of the minimizer and the other is that the functions do not become too flat when n tends to infinity. Note that we implicitly assume the existence of the minimizers.

Remark 5.2. Assume, we have that $\overline{Q}_n$ is constant over all $n \in \mathbb{N}$, $\tilde{\theta} = \theta_1$ is a minimizer of that function and moreover, Θ being compact. Then, $\overline{Q}_n$ is always identifiably unique if θ_1 is unique. This is for example the case for stationary time series. See the commentary to Definition 3.1 in Pötscher and Prucha [1997].

The following theorem introduces the basic result on how we can describe the asymptotic behavior of the CLS estimators. We will formulate this result for a function $R_n(\omega, \theta) = Q_n(X_1, .., X_n, \theta)$ and consequently also $\overline{R}_n(\theta) = \overline{Q}_n(\theta)$.

Theorem 5.3. *Let* $R_n : \Omega \times \Theta \to \mathbb{R}$ *and* $\overline{R}_n : \Theta \to \mathbb{R}$ *be two sequences of functions such that*

$$\sup_{\theta\in\Theta} |R_n(\omega, \theta) - \overline{R}_n(\theta)| \overset{n\to\infty}{\longrightarrow} 0 \;\; P\text{-}a.s. \tag{5.4}$$

Let $\tilde{\theta}_n$ *be an identifiably unique sequence of minimizers of* $\overline{R}_n(\theta)$*, then for any sequence* $\hat{\theta}_n$ *such that eventually*

$$R_n(\omega, \hat{\theta}_n) = \inf_{\theta\in\Theta} R_n(\omega, \theta) \tag{5.5}$$

holds, we have

$$\left\|\hat{\theta}_n - \tilde{\theta}_n\right\| \overset{n\to\infty}{\longrightarrow} 0 \;\; P\text{-}a.s. \tag{5.6}$$

We refer to Lemma 3.1 in Pötscher and Prucha [1997] for a proof of this theorem. This result reduces the asymptotic behavior of the CLS estimates $\hat{\theta}_n$ to the behavior of the minimizers of a CLS function in expectation, i.e., $\tilde{\theta}_n$. Of course, that does not guarantee the convergence of $\hat{\theta}_n$, which only holds in the case, where $\tilde{\theta}_n$ itself converges. In particular, if the $\tilde{\theta}_n$ are constant, which is usually the case if the underlying time series is stationary, $\hat{\theta}_n$ converges.

Remark 5.4. The word *eventually* in the theorem says, that there exists a set $\Omega_0 \in \mathcal{F}$ with $P(\Omega_0) = 1$, such that for all $\omega \in \Omega_0$ there exists a $N(\omega) \in \mathbb{N}$ such that (5.3) holds for $n \geq N(\omega)$.

A large part of a consistency proof now reduces to two parts: the identifiability condition and a uniform law of large numbers (ULLN) as in (5.4). We use this term since it is a law of large numbers, i.e., a sample mean converges to the expectation of the random variables and does this uniformly over a set of functions. This set of functions is given by $\{q(\theta,\cdot)|\theta \in \Theta\}$ with the function q from (5.2). Regarding the second part, we will now introduce a setting in which ULLN's hold for stationary time series. We begin by introducing the definition of equicontinuity for a set of functions which is a necessary condition for the ULLN we are using.

Definition 5.5. A family $\mathcal{D}$ of functions from $\mathbb{R}$ to $\mathbb{R}$ is called equicontinuous, if for all $x \in \mathbb{R}$ and $\varepsilon > 0$, there exists a $\delta > 0$ such that for all $y \in \mathbb{R}$ with $|x-y| < \delta$ and for all $f \in \mathcal{D}$, it holds that $|f(x) - f(y)| < \varepsilon$.

If the set of functions fulfills this equicontinuity condition, we get the following ULLN from Theorem 6.4 in Rao [1962].

Theorem 5.6. *Let $(X_i)_{i\in\mathbb{N}}$ be a stationary time series, $\mathcal{D}$ be an equicontinuous family of functions from $\mathbb{R}$ to $\mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ with $|f(x)| < g(x)$ for each $f \in \mathcal{D}$ and $x \in \mathbb{R}$. Suppose $\mathrm{E}[g(X_1)^{1+\delta}] < \infty$ for some $\delta \geq 0$, then, for*

$$\eta_n := \sup_{f\in\mathcal{D}} \left| n^{-1} \sum_{i=1}^{n} f(X_i) - \mathrm{E}[f(X_1)] \right|$$

we have

(i) $P(\lim_{n\to\infty} \eta_n = 0) = 1$,

(ii) $\lim_{n\to\infty} \mathrm{E}[\eta_n^{1+\delta}] < \infty$.

This is a very useful result for consistency proofs, since for every stationary time series, we only need the set of functions to be equicontinuous. Then, we get the ULLN immediately, if the functions in $\mathcal{D}$ are dominated by a function g and the $(1+\delta)$-th moment of $g(X_1)$ exists. If the parameter set is compact and $\overline{Q}_n(X_1, ..., X_n, \theta) = \mathrm{E}\left[Q_n(X_1, ..., X_n, \theta)\right]$ we have the setting from Remark 5.2, because of the stationarity. Then, also the identifiability condition is fulfilled and $\hat{\theta}_n$ converges to a constant $\tilde{\theta}_n = \theta_1 \in \Theta$. As we are also interested in non-stationary time series, we will not be able to use this result under the alternative. One result for such non-stationary time series is given in Theorem 1 of Pötscher and Prucha [1989]. Still, this theorem is not applicable to one of the non-stationary time series models we introduce later. Thus, we have to take a detour by using a pointwise law of large numbers for so called mixingale arrays which we then generalize to a uniform convergence which yields one part of the consistency in Theorem 5.3.

5.2 L_p-mixingales

Besides the idea of the strong mixing property, there have been different approaches to generalize results like the strong law of large numbers from independent random variables to settings which allow for different dependence structures. One of those approaches are the so called martingale differences. If a time series $(X_i)_{i\in\mathbb{N}}$ adapted to the filtration $(\mathcal{F}_i)_{i\in\mathbb{N}}$ fulfills among other properties that $\mathrm{E}[X_i|\mathcal{F}_{i-1}] = 0$, it is a martingale difference. The name arises from the fact, that for a martingale $(Y_i)_{i\in\mathbb{N}}$ adapted to the same filtration, we have that $X_i := Y_i - Y_{i-1}$ is a martingale difference, since being a martingale in particular means that $\mathrm{E}[Y_i|\mathcal{F}_{i-1}] = Y_{i-1}$. Moreover, one can generalize this idea. Similar as in the case of mixing, instead of the difference being constant 0, one can require it to diminish asymptotically. From this, we get the following definition, consequently called mixingales.

Definition 5.7. Let $(X_i)_{i\in\mathbb{N}}$ be a sequence of random variables on the probability space $(\Omega, \mathcal{F}, P)$ and $(\mathcal{F}_i)_{i\in\mathbb{N}}$ be a filtration, i.e., $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq ... \subseteq \mathcal{F}$. Then, we call the sequence $(X_i, \mathcal{F}_i)_{i\in\mathbb{N}}$ an L_p-mixingale for a $p \geq 1$, if there exist non-negative constants $(c_i)_{i\in\mathbb{N}}$ and $(\psi(m))_{m\in\mathbb{N}}$ where $\psi(m) \overset{m\to\infty}{\longrightarrow} 0$ with

(i) $\|E[X_i|\mathcal{F}_{i-m}]\|_p \leq c_i\psi(m)$ and,

(ii) $\|X_i - \mathrm{E}[X_i|\mathcal{F}_{i+m}]\|_p \leq c_i\psi(m+1)$,

for $i \geq 1$ and $m \geq 0$.

This concept was first introduced in the case of $p = 2$ by McLeish [1975]. For a more general definition with arbitrary p we refer to chapter 16 of Davidson [2021], where the reader can also find a broader introduction to mixingales as well as martingale differences.

Remark 5.8. Note, that in the definition we do not require that the time series is adapted to the filtration. If the process is adapted, part (ii) of the definition always holds. Then, X_i is $\mathcal{F}_{m+i}$-measurable for all $m \geq 0$ and hence $\mathrm{E}[X_i|\mathcal{F}_{i+m}] = X_i$ which means that the difference in (ii) is equal to 0.

In fact, we need a related form of mixingales for so called triangular arrays. Those are time series of the form

$$\{X_{i,n}|i, n \in \mathbb{N},\ i \leq n\}. \tag{5.7}$$

For such a triangular array, McLeish [1977] introduced a concept of a mixingale which has been generalized by Andrews [1988] and de Jong [1996].

Definition 5.9. Let $\{X_{i,n}|i, n \in \mathbb{N}, i \leq n\}$ be a sequence of random variables on the probability space $(\Omega, \mathcal{F}, P)$ and $\{\mathcal{F}_i^n|i, n \in \mathbb{N}, i \leq n\}$ a family of σ-algebras for which $\mathcal{F}_1^n \subseteq \mathcal{F}_2^n \subseteq ... \subseteq \mathcal{F}$ for all $n \in \mathbb{N}$. Then, we call the sequence $\{(X_{i,n}, \mathcal{F}_i^n)|i, n \in \mathbb{N}, i \leq n\}$ an L_p-mixingale array for a $p \geq 1$, if there exist non-negative constants $(c_{i,n})_{i,n\in\mathbb{N}}$ and $(\psi(m))_{m\in\mathbb{N}}$ where $\psi(m) \overset{m\to\infty}{\longrightarrow} 0$ with

(i) $\left\|E[X_{i,n}|\mathcal{F}_{i-m}^n]\right\|_p \leq c_{i,n}\psi(m)$ and,

(ii) $\left\|X_{i,n} - \mathrm{E}[X_{i,n}|\mathcal{F}_{i+m}^n]\right\|_p \leq c_{i,n}\psi(m+1)$,

for $i, n \geq 1$ and $m \geq 0$.

In comparison to Definition 5.7, triangular mixingale arrays are defined as being row-wise mixingales. Meaning, that for all $n \in \mathbb{N}$, $(X_{i,n})_{i \leq n}$ has to fulfill the mixingale property. Also in a row-wise manner, we can formulate a strong law of large numbers for triangular mixingale arrays, as proved in Theorem 4 of de Jong [1996].

Theorem 5.10. *Let $\{(X_{i,n}, \mathcal{F}_i^n) | i, n \in \mathbb{N}, i \leq n\}$ be an L_p-mixingale array for $p \geq 1$. Let $B_{i,n}$ be a deterministic triangular array and a_n a positive integer valued sequence which fulfills the three following statements:*

(i) One of the following conditions holds:

(a) $\sum\limits_{n=1}^{\infty} n^{-1} \sum\limits_{i=1}^{n} c_{i,n}^{p} B_{i,n}^{1-p} < \infty,$

(b) $|X_{i,n}| \leq B_{i,n}$ *almost surely for all* $i \leq n$*,*

(c) $\sum\limits_{n=1}^{\infty} n^{-1} \sum\limits_{i=1}^{n} \mathrm{E}[|X_{i,n}|^r] B_{i,n}^{1-r} < \infty$ *for some* $r \geq 1$*.*

(ii) $\sum\limits_{n=1}^{\infty} \left(n^{-1} \sum\limits_{i=1}^{n} c_{i,n} \psi(a_n) \right)^p < \infty.$

(iii) For all $\beta > 0$*,* $\sum\limits_{n=1}^{\infty} a_n \exp(-\beta^2 n^2 a_n^{-2} (\sum_{i=1}^{n} B_{i,n}^2)^{-1}) < \infty.$

Then,

$$n^{-1} \sum_{i=1}^{n} X_{i,n} \overset{n \to \infty}{\longrightarrow} 0 \tag{5.8}$$

almost surely.

Remark 5.11. In the same paper the author discusses the assumptions of this theorem and compares them to the assumptions of similar SLLN for mixingales as given in Definition 5.7. He concludes that in comparison, the triangular structure necessitates for more restrictive assumptions, represented by the three statements in Theorem 5.10.

Example 5.12. Let $p = 1$ and make the reasonable assumption, that the expectations of the triangular random array are uniformly bounded from below for some value larger than 0. Then, the sum $\sum_{i=1}^{n} B_{i,n}^{1-r}$ must be of at most order $O(n^{-\varepsilon})$ for some $\epsilon > 0$ and $r > 1$. One possible solution is $B_{i,n} = n^{(1+\varepsilon)/(r-1)}$. The same order has to hold for $\sum_{i=1}^{n} \psi(a_n)$, which implies that $a_n \overset{n \to \infty}{\longrightarrow} \infty$ to profit from the convergence to 0 of $\psi(n)$ for $n \longrightarrow \infty$. If we assume $\psi(n)$ to decrease exponentially, a_n is allowed to increase with an arbitrary polynomial rate. Since a_n grows to infinity, the exponential function in the third assumption needs to converge to 0 such that the whole sum can be bounded. Therefore, the term inside the exponential function needs to converge to negative infinity. We can choose the growth rate of a_n arbitrarily small and thus $\sum_{i=1}^{n} B_{i,n}^2$ has to grow with less than polynomial order 2. With our proposal for $B_{i,n}$ we get

$$\sum_{i=1}^{n} B_{i,n}^2 = \sum_{i=1}^{n} n^{2(1+\varepsilon)/(r-1)} = n^{1+\frac{2(1+\varepsilon)}{r-1}}.$$

For $r \in (1, 3]$ the order of the sum can not be below 2. A value $r > 3$ in return would need moments higher than 3 to be uniformly bounded for the triangular array. Hence, one perspective on the restrictiveness of the assumptions is given by the need for the existence of higher moments.

Remark 5.13. In general, the mixingale definition has to be understood as a low level property, able for proving convergence results. It is not recommended to use mixingales for modeling purposes. One reason is that continuous transformations generally do not preserve the mixingale property, as remarked on page 249 of Davidson [2021].

5.3 A uniform law of large numbers for asymptotically uniformly equicontinuous functions

With Theorem 5.10, we get convergence also for non-stationary time series of Q_n to $\overline{Q}_n$, although only for a given $\theta \in \Theta$. So, it is in a sense a pointwise law of large numbers. To apply Theorem 5.3, we need a uniform convergence over all $\theta \in \Theta$. So in the following we present a result which lifts this pointwise convergence to a uniform convergence with a stochastic Arzela-Ascoli type theorem. For this we need a different kind of equicontinuity in a stochastic sense, as given by Pötscher and Prucha [1994].

Definition 5.14. Let $(\Omega, \mathcal{F}, P)$ be a probability space, (Θ, ρ) a metric space and $Q_n : \Omega \times \Theta \to \mathbb{R}$ a sequence of functions that are measurable in their first argument. Then, Q_n is almost surely asymptotically uniformly equicontinuous (AUEC) on Θ, iff

$$\limsup_{n\to\infty} \sup_{\theta' \in \Theta} \sup_{\theta \in \Theta : \rho(\theta, \theta') \leq \delta} |Q_n(\theta) - Q_n(\theta')| \overset{\delta \to 0}{\longrightarrow} 0 \ \ P\text{-}a.s. \tag{5.9}$$

Compared to the equicontinuity from Definition 5.5 one difference is the stochasticity, meaning that it holds almost surely. The other difference is that this condition is only required to be fulfilled asymptotically in n. The uniformity in the name of this condition mirrors the fact that the equicontinuity in Definition 5.5 has to hold uniformly for all functions $f \in \mathcal{D}$. In this case, the set of functions is given by $\{q(\theta, \cdot) | \theta \in \Theta\}$. To show that a function Q_n fulfills this equicontinuity condition can be complex, thus we will make use of the following sufficient condition for the AUEC condition.

Lemma 5.15. *Let $(\omega, \mathcal{F}, P)$ be a probability space, (Θ, ρ) a metric space and $Q_n : \Omega \times \Theta \to \mathbb{R}$ a sequence of functions that are measurable in their first* *argument. If there exist sequence* $B_n : \Omega \to \mathbb{R}$ *and* $h : (0, \infty) \to [0, \infty)$ *with* $h(x) \stackrel{x \to 0}{\longrightarrow} 0$ *and* $|Q_n(\theta) - Q_n(\theta')| \leq B_n h(\rho(\theta, \theta'))$ *for all* $\theta, \theta' \in \Theta$. *Moreover, if*

$$\limsup_{n \to \infty} B_n < \infty \ a.s. \tag{5.10}$$

then the Q_n *are a.s. AUEC.*

A proof for this can be found in Lemma 1 of Andrews [1988]. For functions fulfilling this AUEC condition we can now formulate a theorem which yields uniform convergence over a parameter set Θ, given that pointwise convergence holds for all $\theta \in \Theta$.

Theorem 5.16. *Let* $\overline{Q}_n : \Theta \to \mathbb{R}$ *be an asymptotically uniformly equicontinuous sequence of non-random functions. If* (Θ, ρ) *is totally bounded,* $Q_n(\theta) \stackrel{n \to \infty}{\longrightarrow} \overline{Q}(\theta)$ *almost surely for all* $\theta \in \Theta_0$ *for a dense subset* Θ_0 *of* Θ *and* Q_n *is a.s. AUEC, then*

$$\sup_{\theta \in \Theta} |Q_n(\theta) - \overline{Q}_n(\theta)| \stackrel{n \to \infty}{\longrightarrow} 0 \ \ P\text{-}a.s. \tag{5.11}$$

For a proof we refer to Theorem 3.1 in Pötscher and Prucha [1994]. Later we will only consider parameter sets Θ which are compact, so they are also totally bounded. And since $\Theta_0 = \Theta$ is a dense subset of itself, we will not have many problems with the assumptions of this theorem besides the AUEC condition.

Remark 5.17. In Pötscher and Prucha [1994] we can find more equicontinuity conditions in addition to Definition 5.14. Some of them imply others and under certain conditions they are even equivalent. Moreover, most of them can be deduced by a similar sufficient condition as in Lemma 5.15 and fulfill similar theorems as the one given in Theorem 5.16.

6 A change point test for the Poisson-INARCH(1) process

In this section we present a change point test for a count time series model. It is formulated for abrupt changes in the intensity function of Poisson-INARCH(1) processes. In particular, we want to test for changes in the intensity function. This is work from Franke et al. [2012] and all the results in Section 6.1 and their proofs can be found in that paper.

6.1 The change point model and asymptotics of the test statistic

In the setting of this paper, we do not want to compare arbitrary intensity functions with each other. Instead the intensity function is now defined by a parameter, i.e., $g_\theta : \mathbb{N}_0 \to \mathbb{R}_+$ for some $\theta \in \Theta \subset \mathbb{R}^d$ and $d \in \mathbb{N}$. To ensure that the stationary distribution is not constant at 0, we impose the following structure of the intensity function:

$$g_\theta(y) := \theta_1 + g_{\theta_2,\ldots,\theta_d}(y) \tag{6.1}$$

for $\theta_1 > 0$ and $y \in \mathbb{N}_0$. So we test for changes in the parameter of the intensity function of a Poisson-INARCH(1) process. Thus, we introduce two independent time series $(Y_i)_{i\in\mathbb{N}_0}$ and $(Y_i^*)_{i\in\mathbb{N}_0}$, both being Poisson-INARCH(1) processes with intensity functions g_{θ_0} and $g_{\theta_0^*}$ respectively, for $\theta_0, \theta_0^* \in \Theta$. The time series $(X_i)_{0\le i\le n}$ that we observe is defined as

$$X_i = \begin{cases} Y_i, & 0 \le i \le m^* \\ Y_i^*, & m^* < i \le n \end{cases} \tag{6.2}$$

for $0 \le i \le n$ and $0 \le m^* \le n$. If we have $m^* = n$ then $X_i = Y_i$ for $i = 0, \ldots, n$, meaning that the observed time series does not have a change point, since it equals one of the Poisson-INARCH(1) processes. The change point test can be formulated in the following way.

Definition 6.1. Let $(X_i)_{0\le i\le n}$ be count time series defined as in (6.2). Then, the change point problem (for at most one change) is to test the null hypothesis

$$H_0 : m^* = n$$

against the alternative

$$H_1 : 1 \le m^* < n.$$

Since we can not assume to know the real parameters, we will have to estimate them. We estimate the parameters by the CLS method. We denote the sample residuals given the CLS estimates $\hat{\theta}_n$ of the parameter of the intensity function as

$$\hat{\varepsilon}_i = X_i - g_{\hat{\theta}_n}(X_{i-1}) \tag{6.3}$$

for $i = 1, ..., n$ and based on those denote the cumulative sums as

$$\hat{S}_n(k) = \sum_{i=1}^{k} \hat{\varepsilon}_i \tag{6.4}$$

for $k = 1, ..., n$. The CUSUM test statistic for this case of an abrupt change in the Poisson-INARCH(1) model is then given by

$$T_n = \max_{1 \le k < n} \sqrt{\frac{n}{k(n-k)}} |\hat{S}_n(k)|. \tag{6.5}$$

This test statistic is similar to (2.1), the test statistic we used in Section 2. The main difference is that in the cumulative sums we do not add up the deviations between each observation and the sample mean. Instead, we add up the deviations between each observation and the conditional expectation of that observation given the prior observation. This means that in Section 2 we tested for an abrupt change in the expectation and here for an abrupt change in the conditional expectation. Recall that $g_\theta(Y_{i-1})$ is the conditional expectation of Y_i given Y_{i-1}. Next we want to check the test statistic for its asymptotic behavior. But before doing that, we need to guarantee the consistency of the CLS estimates. To do that, we first need some assumptions which the model has to fulfill.

(A1) $\Theta \in \mathbb{R}^d$ is compact.

(A2) $m^* := m_n = \lfloor \tau n \rfloor$ for some $0 < \tau \le 1$.

(A3) There is a unique $\tilde{\theta}_0 \in \Theta \setminus \partial\Theta$ such that

$$\tilde{\theta}_0 = \arg\min_{\theta \in \Theta} e(\theta),$$

where

$$e(\theta) = \tau \mathrm{E}\left[(Y_1 - g_\theta(Y_0))^2\right] + (1-\tau)\mathrm{E}\left[\left(Y_n^* - g_\theta(Y_{n-1}^*)\right)^2\right].$$

(A4) For all $\theta \in \Theta$, the function $g_\theta : \mathbb{N}_0 \to [0, \infty)$ fulfills the contraction property (3.8).

(A5) $g_\theta(y)$ is twice continuously differentiable with respect to θ for all $y \in \mathbb{N}_0$.

(A6) Let $(Y_i)_{i \in \mathbb{N}}$ be a Poisson INARCH(1) process with intensity function $g_\theta(y)$, $\theta \in \{\theta_0, \theta_0^*\}$, then

$$\mathrm{E}\left[\sup_{\theta \in \Theta} ||\nabla g_\theta(Y_i)(\nabla g_\theta(Y_i))^\top||\right] < \infty \quad , \quad \mathrm{E}\left[Y_i \sup_{\theta \in \Theta} ||\nabla^2 g_\theta(Y_{i-1})||\right] < \infty.$$

Here, ∇ and ∇^2 denote the gradient and the Hessian with respect to θ. A justification for the need of Assumptions (A1) and (A3) can be deduced from

Remark 5.2, which means we can use both to argue that identifiably uniqueness holds for this model. Assumptions (A2) has the same reasoning as in Remark 2.10, guaranteeing that we can apply asymptotic results before and after the change point. Stationarity and absolute regularity - as well as other properties - follow from the contraction property (A4). Lastly, (A5) and (A6) are regularity assumptions on the intensity function used in the proofs of the following theorems, in order to fulfill assumptions of the results used in the proofs. Based on these assumptions, consistency is guaranteed under the null hypothesis H_0 as well as under the alternative H_1.

Theorem 6.2. *Let $X_0, ..., X_n$ be generated by model (6.2) where $(Y_i)_{i\in\mathbb{N}_0}$ and $(Y_i^*)_{i\in\mathbb{N}_0}$ are independent Poisson-INARCH(1) processes with intensity functions $g_{\theta_0}, g_{\theta_0^*} : \mathbb{N}_0 \to \mathbb{R}$ with $\theta_0, \theta_0^* \in \Theta$. Moreover, let (A1)–(A6) be satisfied. Then, the CLS estimate $\hat{\theta}_n$ is strongly consistent for $\tilde{\theta}_0$:*

$$\hat{\theta}_n \overset{n\to\infty}{\longrightarrow} \tilde{\theta}_0 \;\; P\text{-}a.s.$$

If we impose an additional assumption we even get the convergence speed of those estimates.

(A7) The Hessian $\nabla^2 e(\tilde{\theta}_0)$ is positive definite.

Theorem 6.3. *Let $X_0, ..., X_n$ be as in Theorem 6.2 and additionally (A7) be satisfied. Then, we have*

$$\sqrt{n}(\hat{\theta}_n - \tilde{\theta}_0) \overset{d}{\longrightarrow} \mathcal{N}(0, A^{-1}BA^{-1}) \;\; \text{for } n \to \infty$$

with $A = \nabla^2 e(\tilde{\theta}_0)$ and, with $\nabla\tilde{Q}_n(\theta)$ denoting the gradient of $\tilde{Q}_n(\theta)$ w.r.t. θ,

$$B = \lim_{n\to\infty} \frac{1}{n} \mathrm{E}\left[\nabla\tilde{Q}_n(\tilde{\theta}_0)(\nabla\tilde{Q}_n(\tilde{\theta}_0))^\top\right]. \tag{6.6}$$

With the consistency of the CLS estimates ensured, the next step is the asymptotic behavior of the test statistic under H_0. Since the proof relies on an invariance principle as in Theorem 3.11, we need an additional assumption on the existence of moments to fulfill the assumptions of said theorem.

(A8) The $(2+\nu)$-th moment of $\nabla g_{\theta_0}(Y_i)$ exists for some $\delta > 0$

With the new assumption for the existence of a $(2+\nu)$-th moment of the gradient of the intensity before the change point, we can make a statement about the asymptotic behavior of the test statistic under the null hypothesis:

Theorem 6.4. *Let $X_0, ..., X_n$ be as in Theorem 6.3 and additionally (A8) be satisfied. Then, if the hypothesis H_0 holds, we have for all $x \geq 0$*

$$P\left(a(\log(n))\frac{T_n}{\sigma} - b(\log(n)) \leq x\right) \overset{n\to\infty}{\longrightarrow} exp(-2e^{-x}) \tag{6.7}$$

where $a(u) = \sqrt{\log(u)}$, $b(u) = 2\log(u) + 0.5(\log(\log(u)) - \log(\pi))$ and σ^2 being the variance of the residuals $(\hat{\varepsilon}_i)_{i\in\mathbb{N}}$.

Just as in the case of independent normally distributed random variables we now have a test statistic that converges to a Gumbel distribution. Also like in the normally distributed case, we have the same critical value (2.5). Moreover, we can define the same alternative test statistics as in Remark 2.14, which also converge to their respective versions of the supremum of a Brownian bridge. Since the theorem above is formulated for a known variance of the residuals, we give a result for an estimator of this variance below.

Corollary 6.5. *Let $X_0, ..., X_n$ be as in Theorem 6.4. Then*

$$\hat{\sigma}_n^2 = \frac{1}{n-d} \sum_{i=1}^{n} \hat{\varepsilon}_i^2 = \sigma^2 + \mathcal{O}_p(\frac{1}{\sqrt{n}}) \tag{6.8}$$

Naturally, we are also interested in the test statistic under the alternative. The question arises, how the parameters have to change, such that the test has asymptotic power 1 under H_1. In the case of normally distributed data, this was already ensured if the expectation before and after the change point were not equal. The analogous property for the Poisson-INARCH(1) process is, that the expectation of Y_i and the expectation of $g_{\tilde{\theta}_0}(Y_{i-1})$ are not the same. This implies in particular, that we require the parameters θ_0 and θ_0^* to be different from each other. Moreover, $g_{\tilde{\theta}_0}(Y_{i-1})$ must not coincide with the conditional expectation of Y_i given Y_{i-1} for the estimated $\tilde{\theta}_0$.

(A9) $|\mathrm{E}[Y_i] - \mathrm{E}[g_{\tilde{\theta}_0}(Y_{i-1})]| = C > 0$

This last assumptions ensures that the test statistic can distinguish between the distribution before and after the change point. It is necessary for the proofs of the following two results, beginning with the asymptotic behavior of the test statistic under the alternative H_1.

Theorem 6.6. *Let $X_0, ..., X_n$ be as in Theorem 6.4 and additionally (A9) be satisfied. Then, if the hypothesis H_1 holds, we have for all $c > 0$*

$$P\left(a(\log(n))\frac{T_n}{\sigma} - b(\log(n)) \geq c\right) \stackrel{n\to\infty}{\longrightarrow} 1. \tag{6.9}$$

So under the alternative the test statistic grows to infinity. And the estimator of the relative position of the change point is consistent, as of the following corollary.

Corollary 6.7. *Let $X_0, ..., X_n$ be as in Theorem 6.6. Then, we have for*

$$\hat{m}_n = \operatorname*{arg\,max}_{0<k<n} \sqrt{\frac{n}{k(n-k)}} |S_k| \tag{6.10}$$

that

$$\hat{\tau}_n - \tau = o(1) \;\; a.s. \tag{6.11}$$

for $\hat{\tau}_n := \frac{\hat{m}_n}{n}$.

To summarize, for observations as in (6.2) we can formulate a CUSUM test statistic (6.5). This results in a change point test with asymptotic level α under the null hypothesis H_0 and asymptotic power 1 under the alternative H_1. Note that here α is not the multiplicative factor of a linear intensity function as has been used in examples. It is meant as the standard term for the significance level of a test in statistics. Moreover, under H_1, we have a consistent estimator of the relative position of the change point.

6.2 Verifying the assumptions for linear Poisson-INARCH(1) processes

We want to verify, that these assumptions are fulfilled for at last one example. We start by showing that the assumptions (A1) to (A8) are in general fulfilled by a linear intensity function. This means the intensity function is given as

$$g_\theta(y) = \omega + \alpha y \tag{6.12}$$

for $\theta = (\omega, \alpha)^\top$, $\omega \in \mathbb{R}_+$, $0 < \alpha < 1$ and $y \in \mathbb{N}_0$. Then, we know that the consistency of the CLS estimates holds and that the test statistic under the null hypothesis converges to the Gumbel distribution. If the parameters before and after the change point are such that also (A9) holds, we get that the test statistic grows to infinity for $n \to \infty$ and that we have a consistent estimator for the relative position of the change point.

Proposition 6.8. *For an intensity function of the form (6.12), the assumptions (A1)-(A8) hold true.*

Proof. We go through the assumptions one after another.

(A1) By choosing an arbitrarily small $\delta > 0$ and arbitrarily large $\Delta > 0$, we get a compact parameter set $\Theta = [\delta, \Delta] \times [0, 1-\delta] \subset \mathbb{R}^2$ on which we constrain the parameters.

(A2) We set $m^* = \lfloor \tau n \rfloor$ for some $0 < \tau < 1$ and the assumption is fulfilled by definition independent of the choice of $g\theta$.

(A3) We can pull out the parameters in each expected value in $e(\theta)$ for all $\theta \in \Theta$ by linearity of g_θ:

$$\begin{aligned} &E[(Y_1 - g_\theta(Y_0))^2] \\ &\quad = \mathrm{E}[Y_1^2] - 2\mathrm{E}[Y_1 g_\theta(Y_0)] + \mathrm{E}[g_\theta(Y_0)^2] \\ &\quad = \mathrm{E}[Y_1^2] - 2\omega\mathrm{E}[Y_1] - 2\alpha\mathrm{E}[Y_1Y_0] + \omega^2 + 2\omega\alpha\mathrm{E}[Y_0] + \alpha^2\mathrm{E}[Y_0^2] \end{aligned}$$

The same holds for $\mathrm{E}\left[\left(Y_n^* - g_\theta(Y_{n-1}^*)\right)^2\right]$. Thus, we can, without any dominated convergence argument, calculate the first and second derivatives of $e(\theta)$:

$$\nabla_\theta e(\theta) = 2 \begin{pmatrix} \omega + \alpha c_1 - c_2 \\ \alpha c_3 + \omega c_1 - c_4 \end{pmatrix}$$

$$\nabla^2_\theta e(\theta) = 2\begin{pmatrix} 1 & c_1 \\ c_1 & c_3 \end{pmatrix}$$

where

- $c_1 = \tau \mathrm{E}[Y_0] + (1-\tau)\mathrm{E}[Y^*_{n-1}]$,
- $c_2 = \tau \mathrm{E}[Y_1] + (1-\tau)\mathrm{E}[Y^*_n]$,
- $c_1 = c_2$ by stationarity,
- $c_3 = \tau \mathrm{E}[Y_0^2] + (1-\tau)\mathrm{E}[{Y^*_{n-1}}^2]$
- and $c_4 = \tau \mathrm{E}[Y_1 Y_0] + (1-\tau)\mathrm{E}\left[Y^*_n Y^*_{n-1}\right]$.

The critical point $\tilde{\theta}_0 = (\tilde{\omega}_0, \tilde{\alpha}_0)^\top \in \Theta$ fulfilling $\nabla_\theta e(\tilde{\theta}_0) = 0$ is

$$\tilde{\alpha}_0 = \frac{c_4 - c_2^2}{c_3 - c_1^2} = \frac{c_4 - c_1^2}{c_3 - c_1^2} \tag{6.13}$$

$$\tilde{\omega}_0 = c_1(1 - \tilde{\alpha}_0). \tag{6.14}$$

Next we will check if we can construct a Θ as in (A1) such that $\tilde{\theta}_0 \in \Theta \backslash \partial\Theta$. For this we need

$$\begin{aligned} \mathrm{E}[Y_0^2] &= \mathrm{Var}(Y_0) + \mathrm{E}[Y_0]^2 = \alpha^{-1}\mathrm{Cov}(Y_0, Y_1) + \mathrm{E}[Y_0]\mathrm{E}[Y_1] \\ &> \mathrm{Cov}(Y_0, Y_1) + \mathrm{E}[Y_0]\mathrm{E}[Y_1] = \mathrm{E}[Y_0 Y_1] \end{aligned}$$

which holds by stationarity, (3.10) and (3.11). Since this holds also for Y^*_n and Y^*_{n-1}, we have that $c_3 > c_4$. Therefore, $\tilde{\alpha}_0$ must be smaller than 1. Also we have

$$\begin{aligned} c_4 - c_1^2 &= \tau \mathrm{E}[Y_0 Y_1] + (1-\tau)\mathrm{E}\left[Y^*_n Y^*_{n-1}\right] - \left(\tau \mathrm{E}[Y_0] + (1-\tau)\mathrm{E}[Y^*_{n-1}]\right)^2 \\ &= \tau(\mathrm{E}[Y_0 Y_1] - \mathrm{E}[Y_0]\mathrm{E}[Y_1] + \mathrm{E}[Y_0]\mathrm{E}[Y_1]) \\ &\quad + (1-\tau)(\mathrm{E}\left[Y^*_n Y^*_{n-1}\right] - \mathrm{E}\left[Y^*_n\right]\mathrm{E}\left[Y^*_{n-1}\right] + \mathrm{E}\left[Y^*_n\right]\mathrm{E}\left[Y^*_{n-1}\right]) \\ &\quad - \tau^2 \mathrm{E}[Y_0] - 2\tau(1-\tau)\mathrm{E}[Y_0]\mathrm{E}[Y^*_{n-1}] - (1-\tau)^2\mathrm{E}[Y^*_{n-1}]^2 \\ &= \tau \mathrm{Cov}(Y_0, Y_1) + (1-\tau)\mathrm{Cov}(Y^*_n, Y^*_{n-1}) + \tau \mathrm{E}[Y_0]^2 - \tau^2\mathrm{E}[Y_0]^2 \\ &\quad - 2\tau(1-\tau)\mathrm{E}[Y_0]\mathrm{E}[Y^*_{n-1}] + (1-\tau)\mathrm{E}[Y^*_n]^2 - (1-\tau)^2\mathrm{E}[Y^*_n]^2 \\ &= \tau \mathrm{Cov}(Y_0, Y_1) + (1-\tau)\mathrm{Cov}(Y^*_n, Y^*_{n-1}) + \tau(1-\tau)\left(\mathrm{E}[Y_0] - \mathrm{E}[Y^*_n]\right)^2 \\ &> 0 \end{aligned}$$

by stationarity and (3.11). Then, $c_3 - c_1^2 > c_4 - c_1^2 > 0$ and we get that $\tilde{\alpha}_0$ must be larger than 0. This means that $0 < \tilde{\omega}_0 < c_1$ and we can define $\delta > 0$ and $\Delta > 0$ with $\Theta = [\delta, \Delta] \times [0, 1-\delta]$ and $\tilde{\theta}_0 \in \Theta \setminus \partial\Theta$. Lastly, since $\det(\nabla^2_\theta e(\theta)) = c3 - c_1^2 > 0$ for all $\theta \in \mathbb{R}^2$, the Hessian is positive definite and $\tilde{\theta}_0$ is a global minimum.

(A4) By assumption (A2), all $\theta = (\omega, \alpha)^\top \in \Theta$ parameterize $g_\theta(y)$ for $y \in \mathbb{N}_0$ such that

$$|g_\theta(x) - g_\theta(y)| = \alpha|x - y| \leq L|x - y|$$

for $x \in \mathbb{N}_0$ and $L = 1 - \delta < 1$. This proves the contraction property for all $\theta \in \Theta$.

(A5) g_θ is linear in both, ω and α. And since linear functions are twice continuously differentiable, $g_\theta(y)$ is twice continuously differentiable with respect to θ and for all $n \in \mathbb{N}_0$.

(A6) We have

$$\nabla_\theta g_\theta(Y_i) = 2 \begin{pmatrix} 1 \\ Y_i \end{pmatrix}$$

$$\nabla_\theta^2 g_\theta(Y_i) = 2 \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

and since $\mathrm{E}[Y_i^2] < \infty$ for a stationary linear Poisson-INARCH(1) by Theorem 3.25, both expectations in (A6) are finite.

(A7) This is already shown in the proof of (A3).

(A8) The $(2+\nu)$-th moment of $\nabla g_{\theta_0}(Y_i)$ exists for some $\nu > 0$ by the same argument as in the proof of (A6) with Theorem 3.25.

□

Remark 6.9. For a different proof of the consistency and asymptotic normality of the CLS estimates we refer to Section 4.2 of Weiß [2010].

Now that we know all assumptions are fulfilled for the linear Poisson-INARCH(1) process, we can use the test on some simulated paths following model (6.2). Note that we did not include (A9) in the proof above, because it is not fulfilled for all $\theta_0, \theta_0^* \in \Theta$ with $\theta_0 \neq \theta_0^*$. Instead we identify the parameter constellations, for which (A9) is not fulfilled in the following remark.

Remark 6.10. With assumption (A9), we exclude the case where parameters are different but still yield a Poisson-INARCH(1) process with the same expectation. With the consequence of limiting this test of a change in the conditional expectation to a test of a change in the expectation itself. To see that, assume $(X_i)_{0\leq i\leq n}$ as in (6.2) and linear intensity with $\alpha_0 \neq \alpha_0^*$, $\omega_0 \neq \omega_0^*$ and $\frac{\omega_0}{1-\alpha_0} = \frac{\omega_0^*}{1-\alpha_0^*}$ (Consider for example $\alpha_0 = 0.5$, $\alpha_0^* = 0.25$, $\omega_0 = 0.15$ and $\omega_0^* = 0.225$). By (6.14) and the overall expectation being constant, we know that also $\frac{\omega_0}{1-\alpha_0} = \frac{\tilde{\omega}_0}{1-\tilde{\alpha}_0}$. For the equation in assumption (A9) that means

$$\begin{aligned}
\mathrm{E}[Y_i] - \mathrm{E}[g_{\tilde{\theta}_0}(Y_{i-1})] &= \mathrm{E}[Y_i] - \tilde{\omega}_0 - \tilde{\alpha}_0 \mathrm{E}[Y_{i-1}] \\
&= \frac{\omega_0}{1-\alpha_0} - \tilde{\omega}_0 - \tilde{\alpha}_0 \frac{\omega_0}{1-\alpha_0} = (1-\tilde{\alpha}_0)\frac{\omega_0}{1-\alpha_0} - \tilde{\omega}_0 \\
&= (1-\tilde{\alpha}_0)\frac{\tilde{\omega}_0}{1-\tilde{\alpha}_0} - \tilde{\omega}_0 = 0.
\end{aligned}$$

Hence, assumption (A9) can not be fulfilled if the parameters change while the expectation stays constant.

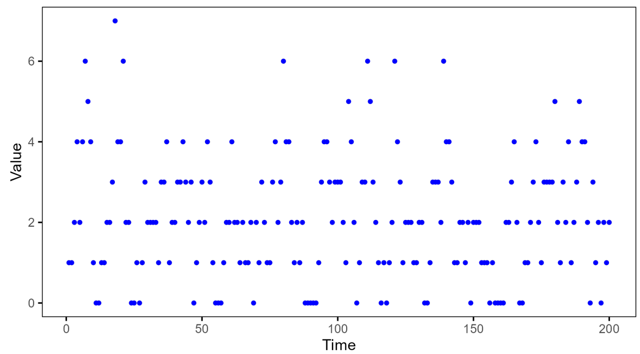

Figure 6.1: Sample path of linear Poisson-INARCH(1) without a change point where $\omega_0 = 1$, $\alpha_0 = 0.5$ and $n = 200$.

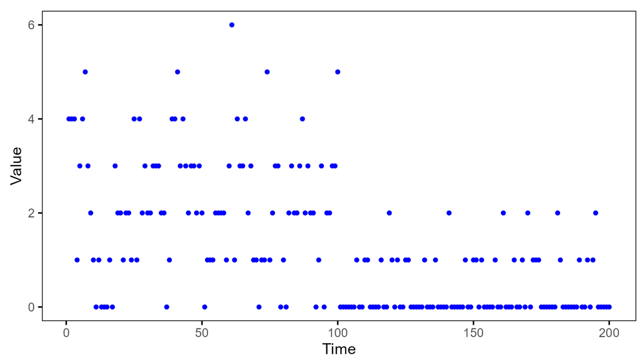

Figure 6.2: Sample path of linear Poisson-INARCH(1) with a change point where $\omega_0 = 1$, $\omega_0^* = 0.3$, $\alpha_0 = 0.5$, $\alpha_0^* = 0.15$ and $n = 200$.

6.3 Numerical results of the test with a linear intensity

As in Section 2, where the concept of change point testing was introduced for independent normally distributed data, we want to inspect the quality of the test on simulated data. Moreover, we want to familiarize ourselves with this test, since the test for gradual changes, which we will introduce later, is based on the same foundations. We start by generating two sample paths - one with and one without a change point. In Figure 6.1 we see a path with no change and in Figure 6.2 one with a change point. In contrast to the model from Section 2, not only the mean of the time series changes but also the variance and covariance change. That can be seen by (3.10) and (3.11), where both depend on the parameters ω and α. But besides that, there are no surprising observations. And when we calculate the CUSUM test statistic as well as the critical values, we get a familiar result. In Figure 6.3, displaying the weighted CUSUMs and critical value for no change point, the values fluctuate relatively close to 0. With peaks

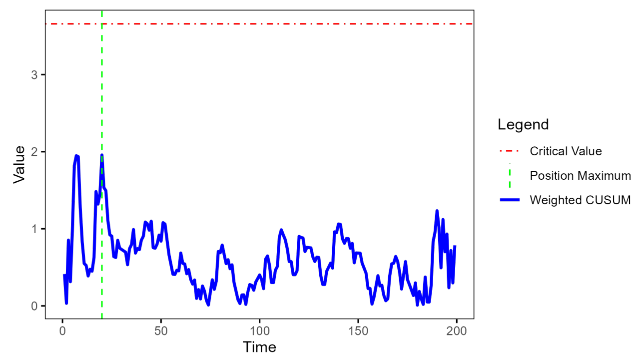

Figure 6.3: Weighted CUSUM, test statistic and critical value of linear Poisson-INARCH(1) without a change point where $\omega_0 = 1$, $\alpha_0 = 0.5$ and a significance level of 5%.

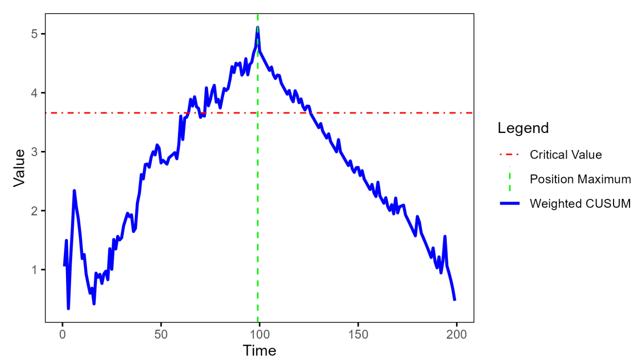

Figure 6.4: Weighted CUSUM, test statistic and critical value of linear Poisson-INARCH(1) with a change point where $\omega_0 = 1$, $\omega_0^* = 0.3$, $\alpha_0 = 0.5$, $\alpha_0^* = 0.15$ and a significance level of 5%.

being around half the critical value. The null hypothesis is accepted with the test statistic being at around 50% of the critical value. The weighted CUSUMs of the time series with a change point are observable in Figure 6.4. They also display similar behavior as in the case of independent normally distributed data. The curve rises and falls in a triangular shape from 0 to the maximum at 99 and reverts back to 0. In comparison to the normally distributed case, there is more noise in this curve, especially before the change point. At the beginning we can even see a small spike. As stated before, we now have changes in variance which could explain this observation. The CLS estimates are calculated based on the whole time series. And applying those estimates to the whole time series may tend to underestimate the real variance in the part with a higher variance, possibly ending up in higher noise of the cumulative sums. Overall, the test

Number of observations	100	200	500	1000
H_0	1.3%	1.87%	2.65%	3.04%
$H_1,\ \tau = 0.5$	22.43%	88.2%	100%	100%
$H_1,\ \tau = 0.75$	6.74%	40.54%	100%	100%

Table 6.1: Power of the test statistic for $\omega_0 = 1$, $\omega_0^* = 0.3$, $\alpha_0 = 0.5$, $\alpha_0^* = 0.15$, significance level of 5%, 10000 repetitions.

Number of observations	100	200	500	1000
$H_1,\ \tau = 0.5$	0.5176	0.4704	0.4882	0.497
$H_1,\ \tau = 0.75$	0.6839	0.6831	0.7347	0.7429

Table 6.2: Averaged estimated relative position of the change point for $\omega_0 = 1$, $\omega_0^* = 0.3$, $\alpha_0 = 0.5$, $\alpha_0^* = 0.15$, significance level of 5%, 10000 repetitions.

works in those two cases similar to the independent normally distributed case.

Next, we will inspect the empirical power and level of the test, based on an experiment of 10000 repetitions in the same setting as the two examples before. The results are presented in Table 6.1. Under the null hypothesis, the test again seems to be conservative, rejecting less than the target of 5%. Here, the convergence seems faster than in the independent normally distributed case. In both settings for τ most of the cases for $n = 100$ observations are accepted. For larger n, the rejection rate converges to 100%. So again, more observations yield results which are closer to the real value. And not only the absolute amount of observations, it is also important, that there are enough observations, both before and after the change point. Table 6.2 shows the estimated relative positions of the change point $\hat{\tau}_n$. The resulting values are already close to the real value for small sample sizes and effectively equal for large ones. If the test is able to identify the alternative correctly, the relative position of the change is easy to estimate.

If we change the value of ω, the effect is similar to changing the difference of the mean before and after the change in the normally distributed setting, see Figure 6.5. We get an S-shaped curve which increases from close to 0 to 100%. This may be due to the fact that ω is an additive term similar to the mean of a normally distribution. Regarding the parameter α, we get a whole new consequence on the test results when varying it. Firstly, α is restricted to be between 0 and 1. So in one experiment, we can cover the whole parameter space for a given ω_0. In particular, we can get close to the extreme case of the expectation and variance converging to infinity if α tends to 1. Secondly, it is a parameter not only influencing both expectation and variance, but growing much faster

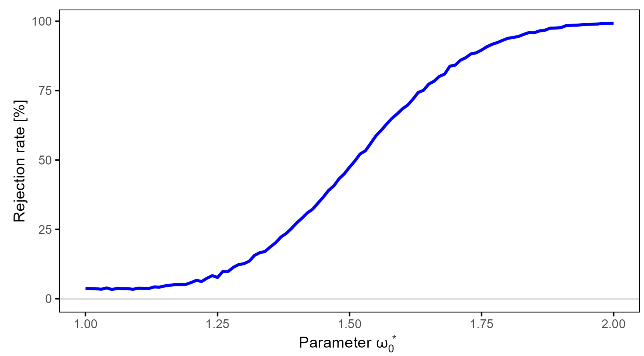

Figure 6.5: Power of the test for $\omega_0^* \in [1, 2]$ with $\omega_0 = 1$, $\alpha_0 = \alpha_0^* = 0.15$, $\tau = 0.5$, $n = 500$, significance level of 5% and 10000 repetitions.

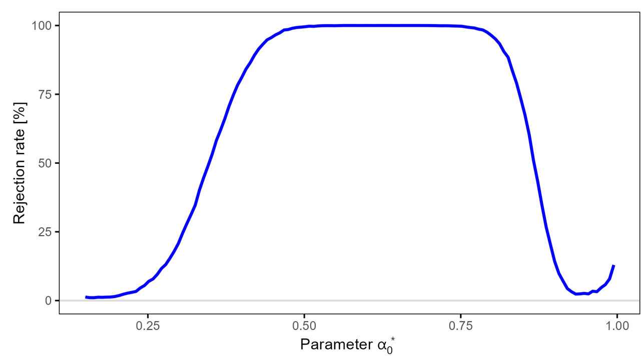

Figure 6.6: Power of the test for $\alpha_0^* \in [0.15, 1)$ with $\omega_0 = \omega_0^* = 2$, $\alpha_0 = 0.15$, $\tau = 0.5$, $n = 500$, significance level of 5% and 10000 repetitions.

regarding the variance than the expectation for increasing α_0, since

$$\mathrm{Var}(Y_0) = \frac{\mathrm{E}[Y_0]}{1 - \alpha_0^2}.$$

For small changes in α_0^*, the rejection rate under the alternative behaves as we would expect, see Figure 6.6: Increasing slowly at first, until α_0^* reaches approximately 0.25. Thereafter the rejection rate increases steeply and arrives at 100% around $\alpha_0^* = 0.5$. The rejection rate stays on this level until around $\alpha_0^* = 0.75$ at 100% but surprisingly falls again steeply after. We suspect this is a consequence of the two properties regarding α discussed before. As the value tends to 1, the expectation and variance get pretty large. For $\alpha_0 = 0.5$ the expectation is 4 and the variance 5.33. At $\alpha_0 = 0.8$ we get an expectation of 10 and variance of 27.78, which is nearly three times its expectation. The large variance after the change seems to have the effect, that a change which is large

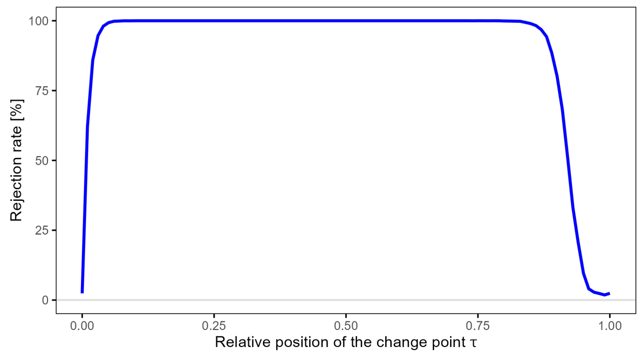

Figure 6.7: Power of the test for $\tau \in [0,1]$ with $\omega_0 = 1$, $\omega_0^* = 0.3$, $\alpha_0 = 0.5$, $\alpha_0^* = 0.15$, $n = 350$, significance level of 5% and 10000 repetitions.

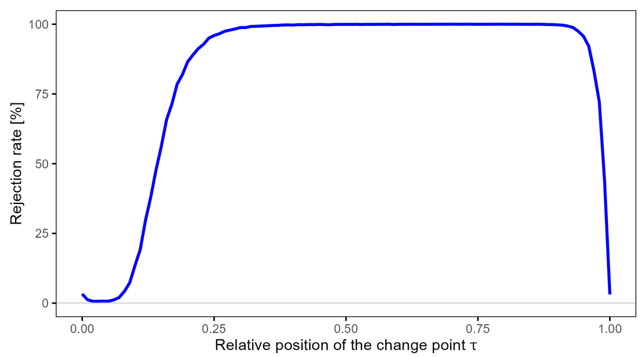

Figure 6.8: Power of the test for $\tau \in [0,1]$ with $\omega_0 = 0.3$, $\omega_0^* = 1$, $\alpha_0 = 0.15$, $\alpha_0^* = 0.5$, $n = 350$, significance level of 5% and 10000 repetitions.

in absolute terms results in a wrong classification by the test.

To conclude this numerical analysis, we apply the test on time series where the relative position of the change point varies from 0 to 1. The result for the same parameters as in Table 6.1, can be seen in Figure 6.7. In this case the curve is not symmetric with respect to a vertical line at $\tau = 0.5$. The rejection rate increases steeply in the beginning and decreases slower for a change point at the other end of the time series. This is surprising and led to further experiments. Since, from what we have seen for normally distributed observations, one would assume changing from one stationary distribution to a second one, is the same as in reverse order. Thus, we swapped the values of α_0 and α_0^* as well as ω_0 and ω_0^* and carried out the same experiment, see Figure 6.8 for the results. The resulting curve is a turned around version of Figure 6.7. So the test really behaves the same for change points in the beginning as those in the end. With the difference being, that a change from θ_0 to θ_0^* at τ is the same as a change from θ_0^* to θ_0 at $1-\tau$. So it does make a difference which distribution is the first one.

7 Modeling gradual changes: a logistic intensity function

In this thesis we are proposing two models for gradual changes. We look at the first proposal of a gradual change model for the Poisson-INARCH process in this section. The idea is to let the time series after the change point start with the last value of the time series before the change point. In difference to the abrupt change point model from Section 6 where we assumed that the time series after the change point already starts in the stationary distribution. Here, we include the transition from the stationary distribution before the change point to new stationary distribution after the change point. Resulting in transitions inspired by the growth rate in Figure 3.7. There we had $\alpha > 1$ for a linear Poisson-INARCH(1) process, resulting in an unbounded exponential increase. Here we want to recreate the exponential increase, but with bounded expectation. Then, we would have a gradual change, which begins to increase smoothly but does not directly jump to a new distribution as in the abrupt case. We will prove that the newly defined time series is still stationary and mixing while being able to model a gradual change as desired.

7.1 A logistic intensity function: Smoother transitions between stationary distributions

If we want to include the convergence to the stationary distribution after the change point, we need to reformulate the model from (6.2). We still have $(Y_i)_{i\in\mathbb{N}_0}$ and $(Y^*_{i,m})_{i\in\mathbb{N}_0}$, both being Poisson-INARCH(1) processes with intensity functions g_{θ_0} and $g_{\theta_0^*}$ respectively, for $\theta_0, \theta_0^* \in \Theta$ and $m \in \mathbb{N}_0$. With the difference being that we set the initial distribution of $(Y^*_{i,m})_{i\in\mathbb{N}_0}$ to the m-th random variable of $(Y_i)_{i\in\mathbb{N}_0}$, i.e., $Y^*_{0,m} = Y_m$. The observations $(X_i)_{i\leq n}$ are then given by

$$X_i = \begin{cases} Y_i, & 0 \leq i \leq m^* \\ Y^*_{i-m^*,m^*}, & m^* < i \leq n \end{cases} \tag{7.1}$$

for $0 \leq i \leq n$ and $0 \leq m^* \leq n$. This allows the time series after the change point to start with the last value of the time series before the change point. This last value then defines the starting point of the second time series, which we expect to converge to its stationary distribution.

Example 7.1. The linear model as proposed in Section 6.2 presents some difficulties if we use it with this approach. It is relatively inflexible, since we can only change two parameters. Changing $\omega_0 > 0$ to $\omega_0^* = \omega_0 + \Delta\omega$ for some $\Delta\omega > 0$ and $0 < \alpha_0 = \alpha_0^* < 1$ yields an expectation in the first time point after

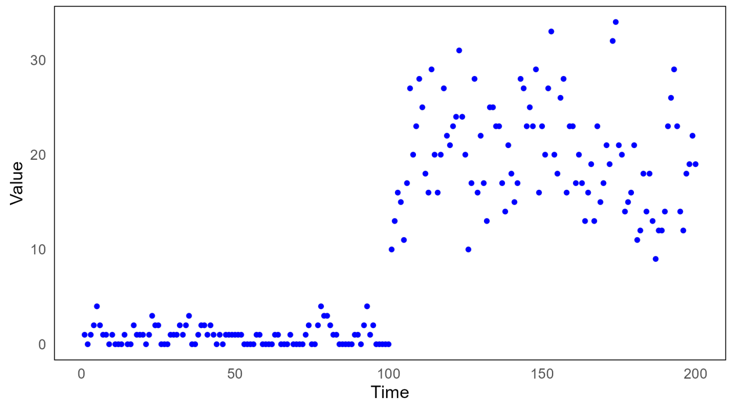

Figure 7.1: Gradual change point in linear Poisson-INARCH(1) process with $\alpha_0 = \alpha_0^* = 0.5$, $\omega_0 = 0.5$, $\omega_0^* = 10.5$, $n = 200$ and a change at $m = 100$.

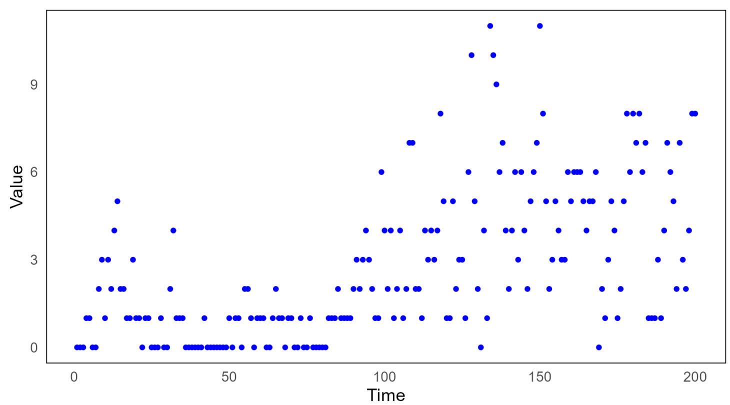

Figure 7.2: Gradual change point in linear Poisson-INARCH(1) process with $\alpha_0 = \alpha_0^* = 0.5$, $\omega_0 = 0.5$, $\omega_0^* = 2.5$, $n = 200$ and a change at $m = 100$.

the change of

$$\begin{aligned}\mathrm{E}[Y^*_{1,m}] &= \mathrm{E}[Y^*_{1,m}|Y_m] = \mathrm{E}[g_{\theta_0^*}(Y_m)] = \omega_0^* + \alpha_0^*\mathrm{E}[Y_m] \\ &= \omega_0 + \alpha_0\frac{\omega_0}{1-\alpha_0} + \Delta\omega = \frac{\omega_0}{1-\alpha_0} + \Delta\omega = \mathrm{E}[Y_m] + \Delta\omega, \end{aligned} \tag{7.2}$$

meaning the expected increase after one step is in expectation the value of the change in ω. As an effect, larger changes in ω can not result in a smooth transition but in expectation always start with a jump of $\Delta\omega$. This can be observed in Figure 7.1. Small changes in ω on the other hand yield new stationary distributions that are only slightly above the one before the change. In particular for count time series, gradual transitions would not be discernible from abrupt

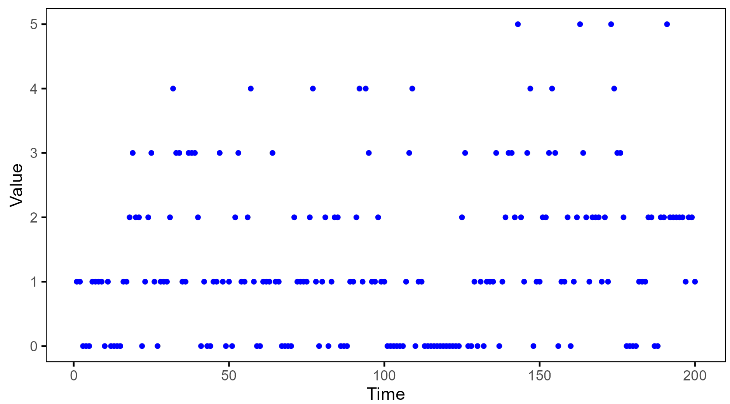

Figure 7.3: Gradual change point in linear Poisson-INARCH(1) process with $\alpha_0 = 0.5$, $\alpha_0^* = 0.6$, $\omega_0 = \omega_0^* = 0.5$, $n = 200$ and a change at $m = 100$.

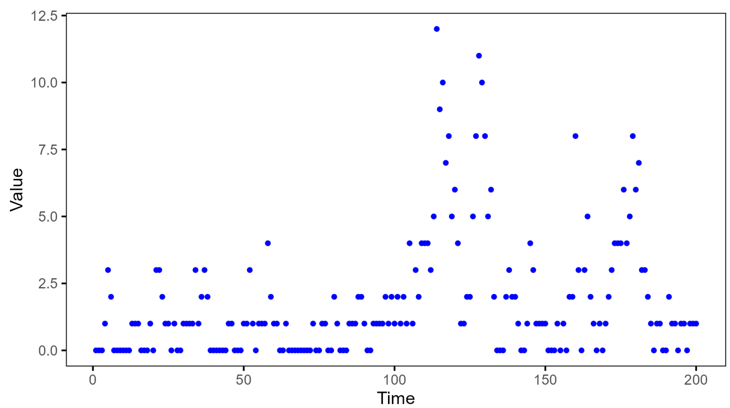

Figure 7.4: Gradual change point in linear Poisson-INARCH(1) process with $\alpha_0 = 0.5$, $\alpha_0^* = 0.8$, $\omega_0 = \omega_0^* = 0.5$, $n = 200$ and a change at $m = 100$.

changes, as can be observed in Figure 7.2. That is reinforced by the variance in the new stationary distribution increasing by $\frac{\Delta\omega}{(1-\alpha_0)(1-\alpha_0^2)}$, since

$$\mathrm{Var}(Y_{i,m}^*) = \frac{\omega_0^*}{(1-\alpha_0^*)(1-\alpha_0^{*2})} = \frac{\omega_0 + \Delta\omega}{(1-\alpha_0)(1-\alpha_0^2)} \tag{7.3}$$

where we assume $Y_{i,m}^*$ to be in its stationary distribution, i.e., $i >> 1$.
Changing the multiplicative parameter from α_0 to $\alpha_0^* = \alpha_0 + \Delta\alpha$ where $\Delta\alpha > -\alpha_0$ with $\alpha_0 + \Delta\alpha < 1$ also does not yield satisfying results. Small values of $\Delta\alpha$ have the same problems as in the case of small changes in ω, for which an example path is given in Figure 7.3. But larger changes in α are not necessarily accompanied by large changes in the expectation of the process directly after

the change. The growth is here not independent of the value of Y_m:

$$\mathrm{E}[Y^*_{1,m}] = \omega^*_0 + \alpha^*_0 \mathrm{E}[Y_m] = \omega_0 + (\alpha_0 + \Delta\alpha)\mathrm{E}[Y_m]. \tag{7.4}$$

But with the α^*_0 getting closer to 1, the variance tends to explode. The larger variance then also means that the difference between transition to the new stationary distribution and the stationary distribution itself vanishes, see Figure 7.4. One would hope that a combination of changes in both parameters could help. But there would still be the problem that small changes tend to blur the difference between a gradual and an abrupt change in the countable case. Or a large $\Delta\omega$ yielding jumps directly after the change point.

To overcome the issues described in the example above, we introduce a more adaptive intensity function which can be non-contractive for small values but is strictly contractive for large values. If the growth rate for small values is low we start with an exponential increase as in Figure 3.7. But for larger values such a function is contractive and should result in a stationary distribution. A function which fulfills this property is the logistic function

$$f(x) = \frac{s}{1 + e^{-\kappa(x - x_0)}}$$

for $\kappa, s, x_0, x \in \mathbb{R}$. Depending on the parameter s, the maximum growth rate can be larger than 1, but for $x \to \infty$ the derivative always tends to 0. For our purposes we set the location parameter x_0 to 0, since it governs the position of the maximum growth rate. With the location parameter set to 0, the maximum of the first derivative with respect to y is at $y = 0$. Since now $f(0) = s/2$, to prevent jumps depending on the choice of s, we set the function to 0 at $x = 0$ by subtracting $s/2$. Then, we add a constant parameter ω for the same reasons as in the linear case. The process would otherwise have an invariant distribution being equal to 0. This results in the following new type of intensity function.

Definition 7.2. A function $g_\theta : \mathbb{N}_0 \times \mathbb{R}^3_+ \to \mathbb{R}$ with

$$g_\theta(y) = \frac{s}{1 + e^{-\kappa y}} - \frac{s}{2} + \omega = \frac{s}{2} \tanh\left(\frac{\kappa y}{2}\right) + \omega \tag{7.5}$$

for $\theta = (\omega, s, \kappa)^\top$ and $\kappa, s, \omega > 0$, $y \in \mathbb{N}_0$ is called logistic intensity function. We call κ the shape parameter, s the scale parameter and ω additive parameter. A Poisson- INARCH(1) process $(Y_i)_{i \in \mathbb{N}_0}$ with a logistic intensity function is called logistic Poisson-INARCH(1) process.

For $y \in \mathbb{N}_0$ we have

$$g_\theta(y) \leq \frac{s}{1 + e^{-\kappa y}} - \frac{s}{2} + \omega < s - \frac{s}{2} + \omega = \frac{s}{2} + \omega,$$

hence the function is bounded by $\frac{s}{2} + \omega$. The possibility of the process diverging to infinity is already prevented by the boundedness of the logistic intensity. The

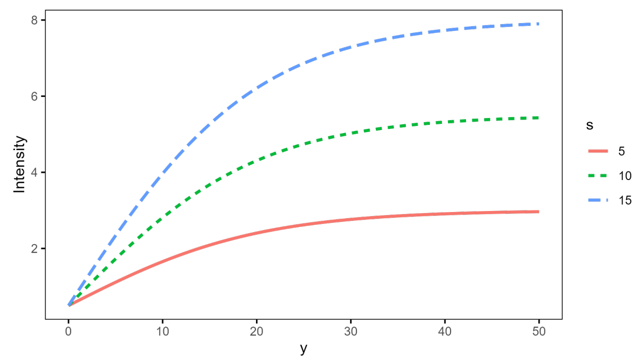

Figure 7.5: Logistic intensity function with $\kappa = 0.1$, $\omega = 0.5$ and $s \in \{5, 10, 15\}$ for values of y between 0 and 50.

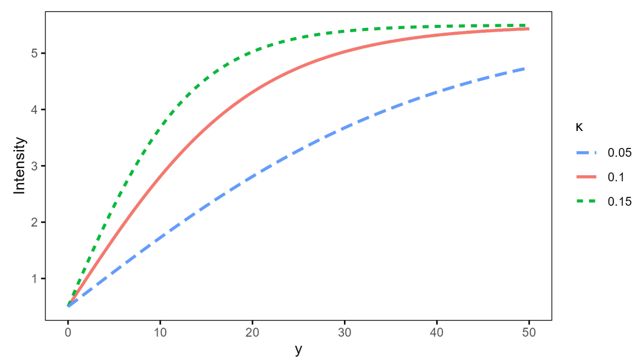

Figure 7.6: Logistic intensity function with $s = 10$, $\omega = 0.5$ and $\kappa \in \{0.05, 0.1, 0.15\}$ for values of y between 0 and 50.

expectation of the process can never exceed the bound of the intensity. The first and second derivatives with respect to y are

$$\frac{\partial}{\partial y} g_\theta(y) = \frac{\kappa s}{4}(1 - \tanh^2\left(\frac{\kappa y}{2}\right)) \tag{7.6}$$

$$\frac{\partial^2}{\partial^2 y} g_\theta(y) = -\frac{\kappa^2 s}{4} \tanh\left(\frac{\kappa y}{2}\right)(1 - \tanh^2\left(\frac{\kappa y}{2}\right)). \tag{7.7}$$

From this we can deduce that $g_\theta(y)$ is increasing in y, since the first derivative is strictly positive. The second derivative is strictly negative and hence $g_\theta(y)$ is concave for $y \in \mathbb{N}_0$ and the first derivative is decreasing. So the first derivative with respect to y takes in fact its maximum value at 0 with

$$\frac{\partial}{\partial y} g_\theta(0) = \frac{\kappa s}{4} \tag{7.8}$$

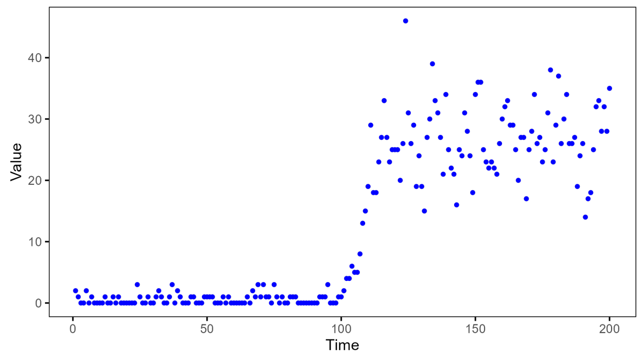

Figure 7.7: Gradual change point in logistic Poisson-INARCH(1) process with $\kappa_0 = 1$, $\kappa_0^* = 0.1$, $s_0 = 1$, $s_0^* = 60$, $\omega_0 = \omega_0^* = 0.5$, $n = 200$ and a change at $m = 100$.

which means that the intensity function is non-contractive if $\kappa s > 4$. Different parameter constellations and how they affect in particular the growth rate close to 0 can be observed in Figure 7.5 and Figure 7.6. The proposed intensity function has the structure which we desire. The path in Figure 7.7 displays no jump directly after the change point. Instead after a few time steps a smooth increase starts, which is bounded although the intensity is non-contractive. That is due to $\kappa_0^* s_0^* = 6 > 4$. Figure 7.8 on the other hand depicts two shortcomings of the logistic intensity function. Especially if the maximum growth rate is just above 1, it can take some time until the growth actually starts. For short sample paths, that could in the worst case lead to the increase not starting until the end of the sample. If the overall change is small, the problems are the same as in the linear model. It can be hard to discern between the new stationary distribution and the transition.

Note that although the distribution after the change has been called stationary, it is as of now not clear that stationarity will still be attained. Since g_θ can be non-contractive, the theory presented in Section 3.4 does not apply. We therefore need to prove these properties in the following chapter with results from Markov chain theory.

7.2 Properties of a logistic Poisson-INARCH(1) process

In this chapter we will use the results from Section 4 to show stationarity and the strong mixing property for logistic Poisson-INARCH(1) processes. Let $(Y_i)_{i \in \mathbb{N}_0}$ be a Poisson-INARCH(1) process with g_θ as in Definition 7.2, $\theta = (\omega, s, \kappa)^\top$, $\kappa, L, \omega > 0$. We consider it as a Markov chain with transition probabilities

$$P(k, \ell) = e^{-g_\theta(k)} \frac{g_\theta(k)^\ell}{\ell!}$$

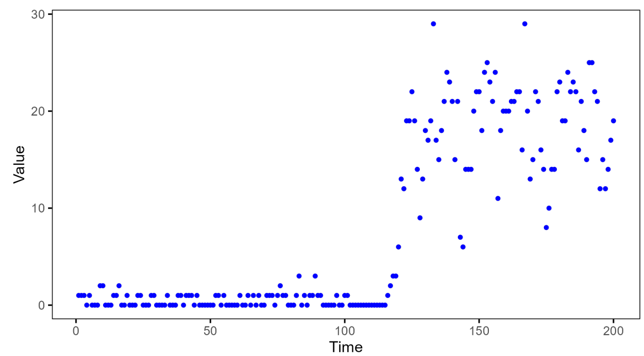

Figure 7.8: Gradual change point in logistic Poisson-INARCH(1) process with $\kappa_0 = 1$, $\kappa_0^* = 0.11$, $s_0 = 1$, $s_0^* = 40$, $\omega_0 = \omega_0^* = 0.5$, $n = 200$ and a change at $m = 100$.

where $k, \ell \in \mathbb{N}_0$, which is the probability mass function of a Poisson distributed random variable with intensity $g_\theta(k)$. To apply the theorems from Section 4, we need to gather the necessary properties of Markov chains for $(Y_i)_{i\in\mathbb{N}_0}$, starting with the smallness of the state space $\mathbb{N}_0$ as defined in Definition 4.3.

Lemma 7.3. *Let $(Y_i)_{i\in\mathbb{N}_0}$ be a logistic Poisson-INARCH(1) process. Then, the set $\mathbb{N}_0$ is small.*

Proof. Set $m = 1$. Since $P^1(k, A) = P(k, A) = e^{-g_\theta(k)} \sum_{\ell\in A} \frac{g_\theta(k)^\ell}{\ell!}$, $k \in \mathbb{N}$, $A \subseteq \mathbb{N}_0$, it suffices to show that there exists a measure ν with the property, that $\nu(k) \leq e^{-g_\theta(k)} \frac{g_\theta(k)^\ell}{\ell!}$, $\forall k, \ell \in \mathbb{N}_0$. Such a measure is:

$$\nu(A) = \sum_{\ell=0}^{N-1} \mathbb{1}_A(\ell) c_\ell + e^{-\omega/e} \sum_{\ell=N}^{\infty} \mathbb{1}_A(\ell) \frac{(\omega/e)^\ell}{\ell!}$$

where $A \subseteq \mathbb{N}_0$, $N = \lceil \frac{s}{2} + \omega(1 - \frac{1}{e}) \rceil$ and $c_\ell = \min_{\lambda\in[\omega, \frac{s}{2}+\omega]} e^{-\lambda} \frac{\lambda^\ell}{\ell!}$.

Then, we have $P(k, \ell) \geq \nu(\ell)$ for $k \in \mathbb{N}$, $A \in \mathcal{B}(\mathbb{N})$ and $\ell < N$, by the definition of ν. Now set $\ell \geq N$ and $\lambda \in [\omega, \frac{s}{2} + \omega]$, i.e., the range of the intensity function. If the inequality

$$e^{-\lambda} \lambda^\ell \geq e^{-\frac{\omega}{e}} \left(\frac{\omega}{e}\right)^\ell. \tag{7.9}$$

is fulfilled, then $P(n, \ell) \geq \nu(\ell)$. Rearranging the inequality yields

$$\left(\frac{\lambda e}{\omega}\right)^\ell \geq e^{\lambda - \frac{\omega}{e}}$$

and applying a logarithm with base $\frac{\lambda e}{\omega}$ to both sides and changing its base to e, results in

$$\ell \geq \frac{\lambda - \frac{\omega}{\ell}}{\log(e\frac{\lambda}{\omega})}.$$

Note that the maximum of the numerator is $\frac{s}{2} + \omega(1 - \frac{1}{e})$ and the minimum of the denominator is 1. Thus, the right hand side of the last inequality is smaller or equal than N, which means that is always fulfilled for $\ell \geq N$. Therefore, we have $P(k, \ell) \geq \nu(\ell)$ for all $k, \ell \in \mathbb{N}_0$ and for $A \subset \mathbb{N}_0$

$$P(k, A) = \sum_{\ell \in A} P(k, \ell) \geq \sum_{\ell \in A} \nu(\ell) = \nu(A)$$

and hence, $\mathbb{N}_0$ is ν-small. □

The next important property to achieve stationarity and strong mixing is Harris recurrence, see Definition 4.7(iii).

Lemma 7.4. *Let $(Y_i)_{i \in \mathbb{N}_0}$ be a logistic Poisson-INARCH(1) process. The Markov chain $(Y_i)_{i \in \mathbb{N}_0}$ is Harris recurrent.*

Proof. We start by showing that $(Y_i)_{i \in \mathbb{N}_0}$ is φ-irreducible (see Definition 4.6) for the measure $\varphi(A) = e^{-}\lambda \sum_{\ell \in A} \frac{\lambda^\ell}{\ell!}$ where $\ell \in \mathbb{N}_0$, $A \subset \mathbb{N}_0$ and $\lambda > 0$. Then, $\varphi(A) > 0$ for all $A \subseteq \mathbb{N}_0$. Hence we need to show $L(k, A) < \infty$ for all $A \subset \mathbb{N}_0$ and $k \in \mathbb{N}_0$. We have

$$\begin{aligned} L(k, A) :=& P\left(\min_{i \in \mathbb{N}}\{Y_i \in A | Y_0 = k\} < \infty\right) \geq P\left(\min_{i \in \mathbb{N}}\{Y_i \in A | Y_0 = k\} = 1\right) \\ \geq& P(Y_1 \in A | Y_0 = k) = e^{-g_\theta(k)} \sum_{\ell \in A} \frac{g_\theta(k)^\ell}{\ell!} > 0. \end{aligned}$$

To apply Proposition 4.9 we need a petite set $C \subset \mathbb{N}_0$ such that for all $k \in \mathbb{N}_0$, $L(k, C) = 1$. By Lemma 7.3, $\mathbb{N}_0$ is small and therefore by Proposition 4.5 also petite. Since the time series always hits the natural numbers including 0, we have that $L(k, \mathbb{N}_0) = 1$. So we can choose $C = \mathbb{N}_0$ and get with Proposition 4.9 that $(Y_i)_{i \in \mathbb{N}_0}$ is Harris recurrent. □

Lastly, we need the time series to be aperiodic as defined in Definition 4.12.

Lemma 7.5. *Let $(Y_i)_{i \in \mathbb{N}_0}$ be a logistic Poisson-INARCH(1) process. The Markov chain $(Y_i)_{i \in \mathbb{N}_0}$ is aperiodic.*

Proof. For each $k \in \mathbb{N}_0$, we have

$$P(k, k) = e^{-g_\theta(k)} \frac{g_\theta(k)^k}{k!} > 0$$

and hence $d(k) = 1$ for all $k \in \mathbb{N}_0$. □

These properties can now be used for the following theorem showing that all logistic Poisson-INARCH(1) processes attain a stationary distribution.

Theorem 7.6. *Let $(Y_i)_{i\in\mathbb{N}_0}$ be a logistic Poisson-INARCH(1) process. Then, a unique invariant probability measure π exists and for every initial condition $k \in \mathbb{N}_0$*

$$\sup_{A\subset\mathbb{N}_0} |P_n(k, A) - \pi(A)| \overset{n\to\infty}{\longrightarrow} 0. \tag{7.10}$$

Proof. Lemma 7.4 guarantees Harris recurrence of the chain and from Remark 4.8 we know that it is also recurrent. As in the proof of Lemma 7.3 we get that $\mathbb{N}_0$ is petite. Since each Y_i lies in $\mathbb{N}_0$, $\tau_{\mathbb{N}_0} = \min\{n \geq 1 : Y_n \in \mathbb{N}_0\} = 1$. Thus, also

$$\sup_{k\in\mathbb{N}_0} E_k[\tau_{\mathbb{N}_0}] = 1 < \infty.$$

From Theorem 4.11 we get that a unique invariant probability measure π exists. With this, Harris recurrence and aperiodicity from Lemma 7.5, we can apply Theorem 4.13 and get the second part of the theorem. Namely, the convergence of the n-step transition matrix to the invariant probability measure, independently of the initial value of the chain. □

Theorem 7.6 implies, besides the existence of the stationary distribution, also the convergence to the invariant distribution for all possible starting values. Consequently, the process after the change $(Y^*_{i,{}^*m})_{i\in\mathbb{N}_0}$ will attain its stationary distribution independently of Y_{m^*}, the last observation before the change. Additionally, by the fact that $\mathbb{N}_0$ is small, we can prove a strong mixing property for $(Y_i)_{i\in\mathbb{N}_0}$.

Theorem 7.7. *Let $Y = (Y_i)_{i\in\mathbb{N}_0}$ be a logistic Poisson-INARCH(1) process. The process is α-mixing with rate*

$$\alpha_Y(n) \leq R\rho^n(1 + \rho^n) \tag{7.11}$$

for some $0 < \rho < 1$ and $0 < R < \infty$.

Proof. By Lemma 7.3 we know that the state space is small. Theorem 4.15 implies then uniform ergodicity of the chain and Theorem 4.16 yields the geometric mixing property with exponentially decreasing rate. Lemma 4.17 then yields the strong mixing property with also exponentially decreasing rate. □

Going back to the test for abrupt changes in the Poisson-INARCH(1) model, we see that the contractive property is essentially used for the same: stationarity and strong mixing property. If we exchange Assumption(A4) by

(A4') $(Y_i)_{i\in\mathbb{N}_0}$ is a logistic Poisson-INARCH(1) process.

the test can be applied in the abrupt case in the same manner. We would only need to prove that the rest of the assumptions are still fulfilled.
Beforehand, we shortly inspect the expectation of the stationary process for the logistic Poisson-INARCH(1). Solving $\mathrm{E}[Y_i] = \mathrm{E}[g_\theta(Y_i)]$ for $\mathrm{E}[Y_i]$ is not as straightforward as in the linear case. We can still get a hint for a possible expectation. The expectation in the linear case is equal to the fixed point of the linear intensity function. So we test this also for the logistic case. But as we will see in the next proposition, it is only an upper bound for the expectation.

Proposition 7.8. *An upper bound for the expected value of a logistic Poisson-INARCH(1) process is the fixed point y_F of g_θ.*

Proof. Since $g_\theta(0) > 0$ and $g_\theta(y) \xrightarrow{y\to\infty} \frac{s}{2} + \omega$, we can find $g_\theta(y_F) = y_F$ by the intermediate value theorem for $f(y) = g_\theta(y) - y$. It is unique since the logistic function is strictly monotonically increasing. Assume that $\mathrm{E}[Y_i] \geq y_F$, i.e., $\mathrm{E}[Y_i] \geq g_\theta(\mathrm{E}[Y_i])$. Then:

$$\mathrm{E}[Y_i] \geq g_\theta(\mathrm{E}[Y_i]) \underset{concave}{\overset{g_\theta\ strictly}{>}} \mathrm{E}[g_\theta(Y_i)] = \mathrm{E}[\mathrm{E}[Y_i|\mathcal{F}_{i-1}]] = \mathrm{E}[Y_{i1}] \overset{stationarity}{=} \mathrm{E}[Y_i]$$

which is an contradiction. Therefore, $\mathrm{E}[Y_i] < y_F$. □

For completeness, we also present a lower bound for the expectation in the stationary distribution.

Proposition 7.9. *If $\kappa s > 4$, a lower bound for the expected value of a logistic Poisson-INARCH(1) process is given by $y_l = \max(\lfloor y_1 \rfloor, \omega)$, where $y_1 \geq 0$ with $\frac{\partial}{\partial y} g_\theta(y_1) = 1$.*

Proof. Since $g_\theta(y) > g_\theta(0) = \omega$ for all $y \in \mathbb{R}$, it holds:

$$\mathrm{E}[Y_i] = \mathrm{E}[\mathrm{E}[Y_i|Y_{i-1}]] = \mathrm{E}\left[\frac{s}{2}\tanh\left(\frac{\kappa Y_{i-1}}{2}\right) + \omega\right] > \omega.$$

So ω is a lower bound for the process. Moreover, if $\frac{\partial}{\partial y} g_\theta(0) \geq 1$ there exists $y_1 \geq 0$ with $\frac{\partial}{\partial y} g_\theta(y_1) = 1$. For $y \in \mathbb{N}_0$ with $y \leq y_1$, there exists a $c \geq 1$ with $g_\theta(y) = cy + \omega$, since $\frac{\partial}{\partial y} g_\theta$ is monotonically decreasing. Then:

$$\begin{aligned}\mathrm{E}[Y_i|Y_{i-1} = y] &= \mathrm{E}[\mathrm{E}[Y_i|Y_{i-1}]|Y_{i-1} = y] \\ &= \mathrm{E}[g_\theta(Y_{i-1})|Y_{i-1} = y] = g_\theta(y) = cy + \omega\end{aligned}$$

This means, that the conditional mean increases at least ω above the previous observation if this is below y_1. Therefore, the mean of the stationary distribution must be above $\lfloor y_1 \rfloor$ if $\frac{\partial}{\partial y} g_\theta(0) \geq 1$. If on the other hand $\frac{\partial}{\partial y} g_\theta(0) < 1$, then because of the concavity of g_θ we get $y_1 < 0$. The process is by definition larger or equal to 0 and thus, $\lfloor y_1 \rfloor$ is still a lower bound. Since the expectation of the process is always larger than both, ω and y_1, the expectation has to be larger than $y_l = \max(\lfloor y_1 \rfloor, \omega)$. □

7.3 Discussion

The logistic intensity function provides a way to implement a gradual change with an exponential growth, by using the transition between stationary distributions and a Lipschitz parameter larger than 1. Because the logistic function flattens to a nearly constant function, the resulting Poisson-INARCH(1) process is still stationary and strongly mixing. The fact that we were not able to compute the expectation analytically, gives a hint about the problems of this approach. Having a random variable inside a tangens hyperbolicus makes it hard to calculate not only the first moment, but also higher moments and covariances. We first tried to prove the assumptions of the abrupt change point test for this model, as in Section 6.2 for a linear intensity function. Most of the assumptions are straightforward, since an upper bound is easy to find, if the random variables are bounded, as is $\tanh\left(\frac{\kappa Y_i}{2}\right)$. To verify assumption (A3) on the other hand, one needs to explicitly find a unique minimum. This was beyond the scope of this work and we rest with the result, that there are stationary and mixing Poisson-INARCH(1) processes, which have non-contractive intensity functions. Thus, we move on to the next section, where we introduce a second approach to model gradual changes, for which we can show the results necessary for formulating a change point test.

8 Modeling gradual changes: adding a time dependent term

Here we introduce a second type of gradual change. This approach was already mentioned when introducing the Poisson-INARCH process in Section 3. After the change point we add a time-dependent but deterministic function to the process which has the desired form of the gradual change. This way, we can very closely control the behavior of the gradual change. By interpreting this new time series as a triangular array, we are able to prove consistency of the CLS estimates. We verify the results experimentally thereafter.

8.1 Adding a deterministic drift term to the intensity function

Adding a time-dependent part on the process itself is either restricted to only integer values or yields a change of the state space of the time series. Therefore, we add it on the intensity function instead. An intensity function with this feature is given by

$$\begin{aligned} \tilde{g}_\theta(\cdot|n) : \mathbb{N}_0 \times \mathbb{N}_0 &\to \mathbb{R} \\ (y, i) &\mapsto g_\theta(y) + f(i, n) \end{aligned} \tag{8.1}$$

where g_θ is a contractive intensity function with $\theta \in \mathbb{R}^d$, a deterministic function $f : \mathbb{N}^2 \to \mathbb{R}^+$ and $n \in \mathbb{N}$. Besides requiring (6.1), the contractive part

$$g_\theta(y) = \theta_1 + g_{\theta_2,\dots,\theta_d}(y) \tag{8.2}$$

must fulfill that,

$$g_{\theta_2,\dots,\theta_d}(y) \geq 0. \tag{8.3}$$

Moreover we have to assume $f(i, n) = 0$ for $i \leq m^*$, i.e., before the change point. This ensures that a process following this intensity function is a Poisson-INARCH(1) process with intensity g_θ until the change point. Thus, under the null hypothesis H_0, we have a time independent Poisson-INARCH(1) process with the same properties as in the abrupt case. Since we still want some kind of regularity for our intensity function, we restrict f to be of a certain class of function, given by

$$f(i, n) := \delta \begin{cases} \left(\frac{i-m}{n}\right)_+^\gamma & i \leq n \\ 1 & i > n \end{cases} \tag{8.4}$$

with $\delta > 0$, $\gamma > 1/2$ and $m, n \in \mathbb{N}$ and $(x)_+ := \max(0, x)$. The class of functions which we use here is based on a paper by Hušková and Steinebach [2000], where the authors introduce a similar kind of gradual change to independently distributed random variables. Their approach is based on i.i.d. distributed observations under the null hypothesis. Under the alternative, they add the same time-dependent function on the process itself. As argued before, this is not possible for the integer valued case, which is why we added the function on the

intensity. The function is equal to 0 for time points before the time of change m, i.e., $i - m < 0$. In particular, $m = n$ means, $f(i, n) = 0$ for all $0 \leq i \leq n$. So we can still define the null hypothesis as $m = n$ and we have $\tilde{g}_\theta(y, i|n) = g_\theta(y)$, which results in an overall stationary process. The parameter δ can be used to change the scale of the overall change and γ controls the shape of the gradual change. Since $f(i, n)$ is monotonically increasing for all $i \in \mathbb{N}_0$, the gradual change now presents itself as adding a larger value to the intensity function, compared to the time point before. The difference of these values between two time points is given by the shape of the time-dependent function. So we expect that the gradual change mirrors the shape of this function.
The intensity function in (8.1) with f given by (8.4) depends on both the time point i and number of observations n. Thus, a time series which is a Poisson-INGARCH(1) process with intensity as in (8.1) has two time components i and n where $i \geq n$. For constant values of n this is not an issue. But if we want to analyze asymptotic behavior related to such a model, we need to take that into account. We approach this by considering such a process as a triangular array $\{Y_{i,n} | i \in \mathbb{N}_0,\ n \in \mathbb{N},\ i \leq n\}$ with pairwise independent rows. Consequently, the filtration will also be a triangular array $\{\mathcal{F}_i^n | i \in \mathbb{N}_0,\ n \in \mathbb{N},\ i \leq n\}$, where $\mathcal{F}_i^n = \sigma(Y_{0,n}, ..., Y_{i,n})$.

Definition 8.1. A triangular array $\{Y_{i,n} | i \in \mathbb{N}_0,\ n \in \mathbb{N},\ i \leq n\}$ is called triangular Poisson-INARCH(1) process with intensity function $\tilde{g}_\theta$ as in (8.1), if for each $n \in \mathbb{N}$, the n-th row $(Y_{i,n})_{i \in \mathbb{N}_0}$ is a Poisson-INARCH(1) process with intensity function $\tilde{g}_\theta(\cdot|n)$. If g_θ is linear, we call it linear triangular Poisson-INARCH(1) process.

Example 8.2. If we set the parameter γ to $1/2$, we get a gradual change in the form of a square root function. Thus, after the change point, the time series starts with a steep increase which then flattens, see Figure 8.1. Although the growth is steepest directly after the change point, there is no behavior as in the previous approach, where changes in ω led to a jump. Setting $\gamma = 2$ leads to a change similar to a parabola, as displayed in Figure 8.2. Here, the change is in the beginning not even discernible from the process before the change. Only after a few steps the value of the quadratic function gets large enough to make a visible impact on the time series. Overall, theses examples seems to confirm our expectation about the behavior of the time series after the change.

With this approach we get a gradual change which is comparatively easier to control compared to the transition between stationary distributions, as we have seen in Section 7. By having γ as a parameter, we can decide the shape of the gradual change. While in Section 7, we only had an exponential growth rate.
Note, that we can not formulate a model for the change point problem in the same manner and assumptions as in Section 6. Although the time series under the null hypothesis H_0 and consequently all results concerning H_0 will stay the same, the alternative will deviate strongly. If we still want to use a CUSUM test statistic, we first need to comprehend the behavior of the CLS estimates for θ under the alternative. Hence, the rest of this subsection will be concerned

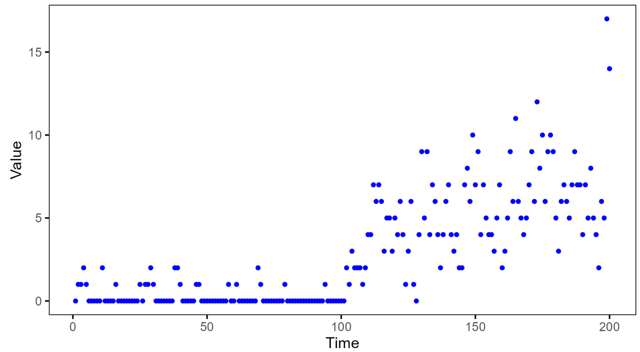

Figure 8.1: Gradual change point in linear Poisson-INARCH(1) process with $\alpha_0 = 0.5$, $\omega_0 = 0.5$, $\gamma = 1/2$. $\delta = 5$, $n = 200$ and a change at $m = 100$.

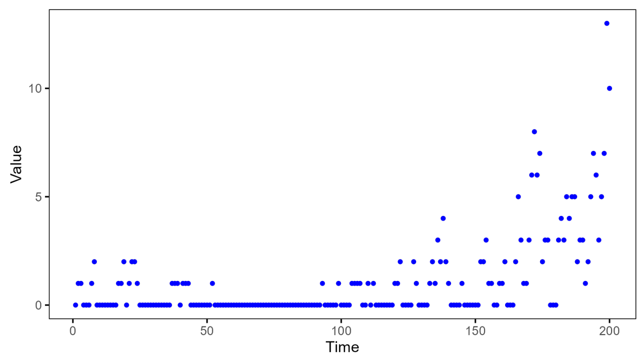

Figure 8.2: Gradual change point in linear Poisson-INARCH(1) process with $\alpha_0 = 0.5$, $\omega_0 = 0.5$, $\gamma = 2$. $\delta = 10$, $n = 200$ and a change at $m = 100$.

with the asymptotics of CLS estimates if a gradual change (8.1) is present. Thereafter, we prove that the CLS estimates converge if the contractive part g_θ of the time-dependent intensity function $\tilde{g}_\theta$ is linear.

8.2 Convergence of conditional least squares estimators for θ under the alternative

For change point tests as in Section 6, we need an estimate of the parameter θ. By construction, the test has the null hypothesis of no change. So we construct the CLS function as if we would know, that the process is stationary with intensity function g_{θ_0} for some $\theta_0 \in \Theta$, i.e.,

$$Q_n(\theta) := \frac{1}{n} \sum_{i=1}^{n} q(Y_{i,n}, Y_{i-1,n}, \theta) := \frac{1}{n} \sum_{i=1}^{n} (Y_i - g_\theta(Y_{i-1,n}))^2 \qquad (8.5)$$

where

$$q(Y_{i,n}, Y_{i-1,n}, \theta) := (Y_i - g_\theta(Y_{i-1,n}))^2 . \tag{8.6}$$

We get this formulation since for a stationary process $(Y_i)_{i\in\mathbb{N}_0}$ the conditional expectation of Y_i given the filtration at $i-1$ is equal to $g_\theta(Y_{i-1})$. The CLS estimate is given by the minimizers of this function. To know, how the test statistic based on this estimate behaves, we must know how the estimates themselves behave asymptotically. Thus, the goal of this subsection is to give conditions, under which the CLS estimates converge, for a model based on the gradual change from (8.1). As presented in Section 5, we start with proving a strong law of large numbers, which we generalize to a uniform law of large numbers. This, together with unique identifiability of the minimizers, yields asymptotic properties of the CLS estimate. To start, we check if each row of a triangular Poisson-INARCH(1) process $(Y_i)_{i\in\mathbb{N}_0}$ with an intensity function (8.1) fulfills a strong mixing condition. As we have seen in the sections before, it is a crucial property which ensures many useful other results.

Theorem 8.3. *Let $\{Y_{i,n} | i \in \mathbb{N}_0,\ n \in \mathbb{N},\ i \le n\}$ be a triangular Poisson-INARCH(1) process with intensity function $\tilde{g}_\theta$ as in (8.1). For $n \in \mathbb{N}$, the process $Y = (Y_i, n)_{i\in\mathbb{N}_0}$ is absolutely regular with*

$$\beta_Y(i) \le 2L_\theta^{i-1}\frac{g_\theta(0)+\delta}{1-L_\theta}, \tag{8.7}$$

where L_θ is the Lipschitz constant of the contractive part g_θ of the time-dependent intensity function $\tilde{g}_\theta$ and δ the scale parameter from (8.4).

Proof. We show this by verifying conditions (C1)-(C3) from Section 3.5 with $d(\lambda, \lambda') = |\lambda - \lambda'|$. We can drop the $(Z_i)_{i\in\mathbb{N}_0}$ variables from our coupling arguments, since we do not have this additional random variable.

(C1) The first condition is straightforward, since the intensity one step before is not an argument of the intensity function. The rest then cancels out. Let $\lambda, \lambda' > 0$, $i, n \in \mathbb{N}_0$ and $Y \sim Pois(\lambda)$. Then,

$$\mathrm{E}[d\left(\tilde{g}_\theta(Y, i|n), \tilde{g}_\theta(Y, i|n)\right)] = 0.$$

Thus, we can set $L_1 = 0$.

(C2) We construct a coupling as follows. Let $\lambda', \lambda > 0$, $Y \sim Pois(\lambda)$ and $Y' \sim Pois(\lambda')$. If $\lambda' \le \lambda$, then $Y = Y' + W$ where $W \sim Pois(\lambda - \lambda')$ is independent of Y'. Conversely, if $\lambda' > \lambda$, then $Y' = Y + W$ where $W \sim Pois(\lambda' - \lambda)$ is independent of Y'. This ensures that we have

$$\mathrm{E}\left[|Y - Y'|\right] = |\lambda - \lambda'|.$$

We also get that $Y' \le Y$ if $\lambda' \le \lambda$ and $Y' \ge Y$ if $\lambda' \ge \lambda$. Note that we dropped the index i. We can do this since

$$\tilde{g}_\theta(x, i+1|n) - \tilde{g}_\theta(y, i+1|n) = g_\theta(x) - g_\theta(y)$$

for all $x, y \in \mathbb{N}_0$. We then have

$$\mathrm{E}[|g_\theta(Y) - g_\theta(Y')|] \leq L_\theta \mathrm{E}[|Y - Y'|] = L_\theta |\lambda - \lambda'|$$

and $L_2 = L_\theta$.

(C3) As in the proof of Corollary 3.1 in Doukhan et al. [2022], we do not prove (C3) itself. Instead we show the property itself, which is usually ensured by (C3). That is for all $i \in \mathbb{N}_0$

$$\tilde{E}\left[|\tilde{\lambda}_i - \tilde{\lambda}'_i|\right] < \infty$$

for the processes $((\tilde{Y}_i, \tilde{\lambda}_i))_{i \in \mathbb{N}_0}$ and $((\tilde{Y}'_i, \tilde{\lambda}'_i))_{i \in \mathbb{N}_0}$ which are independent copies of the original process $((Y_i, \lambda_i))_{i \in \mathbb{N}_0}$ on a probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$. We get that by

$$\begin{aligned}
\tilde{E}\left[|\tilde{\lambda}_{i+1} - \tilde{\lambda}'_{i+1}|\right]/2 &< \tilde{E}\left[\tilde{\lambda}_{i+1}\right] = \tilde{E}\left[g_\theta(\tilde{Y}_i, i+1|n)\right] \\
&\leq L_\theta \tilde{E}\left[\tilde{Y}_i\right] + g_\theta(0) + f(i+1, n) = L_\theta \tilde{E}\left[\tilde{\lambda}_i\right] + g_\theta(0) + \delta \\
&\leq L_\theta^2 \tilde{E}\left[\tilde{\lambda}_{i-1}\right] + (g_\theta(0) + \delta)(1 + L_\theta) \\
&< \ldots < L_\theta^{i+1} \mathrm{E}[\tilde{\lambda}_0] + (g_\theta(0) + \delta) \sum_{j=0}^{i} L_\theta^j \\
&< \mathrm{E}[\tilde{\lambda}_0] + \frac{g_\theta(0) + \delta}{1 - L_\theta} < \infty
\end{aligned}$$

With (2.8) from Doukhan et al. [2022], i.e.,

$$\beta_Y(i, k) := \beta(\sigma(Y_0, \ldots, Y_k), \sigma(Y_{k+i}, Y_{k+i+1}, \ldots)) \leq L_\theta^{i-1} \tilde{E}\left[|\tilde{\lambda}_{k+1} - \tilde{\lambda}'_{k+1}|\right]$$

and taking the supremum over k on both sides of the inequality, we get a β-mixing rate of

$$\beta_Y(i) \leq L_\theta^{i-1} \tilde{E} \sup_{k \in \mathbb{N}} \left[|\tilde{\lambda}_k - \tilde{\lambda}'_k|\right] < 2 L_\theta^{i-1} \frac{g_\theta(0) + \delta}{1 - L_\theta}$$

by setting $\lambda_0 = 0$ since the process in the given formulation is independent of λ_0.

□

In particular we get from this result, that the mixing rate is independent of the function $f(i, n)$.

With the triangular representation of the time series, we can apply our results for triangular arrays, namely the strong law of large numbers given in Theorem 5.10. This theorem applies only to mixingale arrays. To prove the first condition of mixingale arrays, we make use of Lemma 3.7. A part of the upper bound in that lemma is given by the r-th moment for some $r > 1$ of the random variable we want to bound from above. As we will see later, it is useful to have these moments be uniformly bounded. The next lemma provides us with such an uniform upper bound in the case that the respective moment exists for the process generated by the contractive part of the gradual intensity function.

Lemma 8.4. *Let $\{Y_{i,n} | i \in \mathbb{N}_0,\ n \in \mathbb{N},\ i \leq n\}$ be a triangular Poisson-INARCH(1) process with intensity function $\tilde{g}_\theta$. Assume that for $k > 0$, the k-th moment Poisson-INARCH(1) generated by g_θ exists. Then, the k-th moment of $Y_{i,n}$ exists for all $i \in \mathbb{N}_0$, $n \in \mathbb{N}$, $i \leq n$ and is bounded by a constant C, $0 < C < \infty$.*

Proof. Let $(Y_i)_{i\in\mathbb{N}_0}$ be a stationary Poisson-INARCH(1) process with contractive intensity function g_θ. Note that for $i \leq m$ the time series is stationary with $\mathrm{E}[Y_{i,n}^k] = \mathrm{E}[Y_0^k] < \infty$. We have with Lemma A.2(iii) for $i = m+j$, where $j \in \mathbb{N}$, with $j < (n - m + 1)$

$$
\begin{aligned}
E[Y_{i,n}^k] &\leq \mathrm{E}\left[\exp\left(\frac{k^2}{2\tilde{g}_\theta(Y_{i-1,n}, i|n)}\right)\tilde{g}_\theta^k(Y_{i-1,n}, i|n)\right] \\
&\leq \exp\left(\frac{k^2}{2g_\theta(0)}\right)\mathrm{E}\left[\left(g_\theta(Y_{i-1,n}) + f(i,n)\right)^k\right] \\
&= \exp\left(\frac{k^2}{2g_\theta(0)}\right)\mathrm{E}\left[g_\theta^k(Y_{i-1,n})\left(1 + \frac{f(i,n)}{g_\theta(Y_{i-1,n})}\right)^k\right] \\
&\leq \exp\left(\frac{k^2}{2g_\theta(0)}\right)\left(1 + \frac{\delta}{g_\theta(0)}\right)\mathrm{E}[g_\theta^k(Y_{i-1,n})] \\
&\leq L_\theta^k \underbrace{\exp\left(\frac{k^2}{2g_\theta(0)}\right)\left(1 + \frac{\delta}{g_\theta(0)}\right)}_{=:c_1}\mathrm{E}\left[\left(Y_{i-1,n} + g_\theta(0)\right)^k\right] \\
&= L_\theta^k c_1\left(g_\theta(0)^k P(Y_{i-1,n} = 0) + \mathrm{E}\left[Y_{i-1,n}^k\left(1 + \frac{g_\theta(0)}{Y_{i-1,n}}\right)^k \mathbb{1}(Y_{i-1,n} > 0)\right]\right) \\
&\leq L_\theta^k c_1\left(g_\theta(0)^k + (1 + g_\theta(0))^k\,\mathrm{E}[Y_{i-1,n}^k]\right) \\
&\leq L_\theta^k\left(c_2 + c_3\mathrm{E}[Y_{i-1,n}^k]\right) \\
&\leq \ldots \leq L_\theta^{kj} c_3\mathrm{E}[Y_0^k] + c_2\sum_{\ell=0}^{m-1} L_\theta^{k\ell} \leq c_3\mathrm{E}[Y_0^k] + \frac{c_2}{1 - L_\theta^k} =: C < \infty
\end{aligned}
$$

□

Now we turn back to the CLS function (8.5). A triangular process, which represents row-wise each summand of the CLS function, can be defined by

$$Q_{i,n} := q(Y_{i,n}, Y_{i-1,n}, \theta) \tag{8.8}$$

for $\theta \in \Theta$ and $i \in \mathbb{N}_0$, $n \in \mathbb{N}$, $i \leq n$. This is the triangular array that we are interested in and want to apply a strong law of large numbers on. So the next step will be to verify, that it is indeed a mixingale array.

Lemma 8.5. *Let $\{Y_{i,n} | i \in \mathbb{N}_0,\ n \in \mathbb{N},\ i \leq n\}$ be a triangular Poisson-INARCH(1) process with intensity function $\tilde{g}_{\theta_0}(y, i|n)$ for some $\theta_0 \in \Theta$. Assume the $(2+\nu)$-th moments exist for the Poisson-INARCH(1) generated by g_{θ_0} and some $\nu > 0$. Then, $X = \{X_{i,n} | i \in \mathbb{N}_0,\ n \in \mathbb{N},\ i \leq n\}$ given by*

$$X_{i,n} := q(Y_{i,n}, Y_{i-1,n}, \theta) - \mathrm{E}[q(Y_{i,n}, Y_{i-1,n}, \theta)]$$

for $i \in \mathbb{N}_0$, $n \in \mathbb{N}$, $i \leq n$ and $\theta \in \Theta$ is a triangular L_1-mixingale array with the constants $c_{i,n} = c$ for some constant $c < \infty$.

Proof. Set $\varepsilon = \nu/2$ and $Y_{\cdot,n} = (Y_{i,n})_{i \in \mathbb{N}_0}$. By Lemma 8.4 we know that the second moment of $Y_{i,n}$ is uniformly bounded for all $i \in \mathbb{N}_0$, $n \in \mathbb{N}$, $i \leq n$. Theorem 8.3 then implies, that $Y_{\cdot,n}$ is strongly mixing for all $n \in \mathbb{N}$ with the mixing rate $\beta_{Y_{\cdot,n}}(j) \leq L_\theta^{j-1} \frac{g_\theta(0)+\delta}{1-L_\theta}$. With Lemma 3.10, we get the row-wise mixing rate of X which is $\beta_{X_{\cdot,n}}(j) \leq L_\theta^{j-2} \frac{g_\theta(0)+\delta}{1-L_\theta}$, since $X_{i,n}$ is a continuous transformation of $Y_{i,n}$ and $Y_{i-1,n}$. Applying Lemma 3.7 and having $\mathrm{E}[Y_{i,n}^{2+\nu}] \leq C < \infty$ as in Lemma 8.4, we get

$$\begin{aligned}
&\left\| \mathrm{E}[X_{i,n} | \mathcal{F}_{i-j}^n] \right\|_1 \leq 6\alpha(\mathcal{F}_{i-j}^n, \mathcal{F}_i^n)^{1-1/(1+\varepsilon)} \left\| X_{i,n} \right\|_{1+\varepsilon} \\
&\leq 6\beta_{X_{\cdot,n}}(j)^{1-1/(1+\varepsilon)} \left\| q(Y_{i,n}, Y_{i-1,n}, \theta) - \mathrm{E}[q(Y_{i,n}, Y_{i-1,n}, \theta) \right\|_{1+\varepsilon} \\
&\leq 6\beta_{X_{\cdot,n}}(j)^{1-1/(1+\varepsilon)} \left(\left\| q(Y_{i,n}, Y_{i-1,n}, \theta) \right\|_{1+\varepsilon} + \left\| \mathrm{E}[q(Y_{i,n}, Y_{i-1,n}, \theta) \right\|_{1+\varepsilon} \right) \\
&\leq 12 \left(2L_\theta^{j-2} \frac{g_\theta(0)+\delta}{1-L_\theta} \right)^{\varepsilon/(1+\varepsilon)} \mathrm{E}\left[\left| Y_{i,n} - g_\theta(Y_{i-1,n}) \right|^{2+\nu} \right]^{1/(1+\varepsilon)} \\
&\leq 12 \left(2L_\theta^{j-2} \frac{g_\theta(0)+\delta}{1-L_\theta} \right)^{\varepsilon/(1+\varepsilon)} \left\| Y_{i,n} - g_\theta(Y_{i-1,n}) \right\|_{2+\nu}^{(2+\nu)/(1+\varepsilon)} \\
&\leq 12 \left(2L_\theta^{j-2} \frac{g_\theta(0)+\delta}{1-L_\theta} \right)^{\varepsilon/(1+\varepsilon)} \left((1 + L_\theta^{2+\nu}) \left\| Y_{i,n} \right\|_{2+\nu} + \left\| g_\theta(0) \right\|_{2+\nu} \right)^2 \\
&\leq \underbrace{2L_\theta^{j\varepsilon/(1-\varepsilon)}}_{=:\psi(j)} \underbrace{12 \left(L_\theta^{-2} \frac{g_\theta(0)+\delta}{1-L_\theta} \right)^{\varepsilon/(1+\varepsilon)} \left((1 + L_\theta^{2+\nu}) C^{1/(2+\nu)} + g_\theta(0) \right)^2}_{=:c} \\
&= c\psi(j).
\end{aligned}$$

Since $L_\theta < 1$, we get that $\psi(j) \overset{j \to \infty}{\longrightarrow} 0$, because $L_\theta^{j\varepsilon/(1-\varepsilon)}$ converges to 0 for all $\varepsilon > 0$. The constants $c_{i,n}$ are all set to the same value $c < \infty$. Thus, we get the first condition for a triangular L_1-mixingale array. Since the filtration $(\mathcal{F}_i^n)_{i \in \mathbb{N}_0}$ is adapted to $(Y_{\cdot,n})_{i \in \mathbb{N}_0}$, we get that $X_{i,n}$ is measurable with respect to $\mathcal{F}_{i+j}^n$ for all $j \in \mathbb{N}_0$. Remark 5.8 then yields the second condition for a triangular L_1-mixingale array. □

This lemma reveals the first new assumption for the process under the alternative to get converging CLS estimates: some moments of the triangular Poisson-INARCH(1) process need to exist. From Lemma 8.5, the $(2+\nu)$-th moments would be necessary. Recall the discussion of Theorem 5.10 regarding the strong assumptions in that theorem. Taking these assumptions into account, we require the triangular Poisson-INARCH(1) process to have finite $(6+\nu)$-th moments, yielding the assumption.

(A10) The $(6+\nu)$-th moments exist for the Poisson-INARCH(1) generated by $g_{\theta_0}(y)$ for some $\nu > 0$.

Having verified that the triangular array (8.8) from which we subtracted its expectation is a mixingale array, we can go on with the strong law of large numbers.

Theorem 8.6. *Let $\{Y_{i,n} | i \in \mathbb{N}_0,\ n \in \mathbb{N},\ i \leq n\}$ be a triangular Poisson-INARCH(1) process with with intensity function $\tilde{g}_{\theta_0}(y, i|n)$ for some $\theta_0 \in \Theta$. Assume that the Poisson-INARCH(1) process generated by g_{θ_0} fulfills (A1)-(A8) and (A10). Then,*

$$\frac{1}{n}\sum_{i=1}^{n} Q_{i,n}(\theta) \overset{n\to\infty}{\longrightarrow} \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}[Q_{i,n}(\theta)] \tag{8.9}$$

almost surely for all $\theta \in \Theta$.

Proof. By Lemma 8.5, X is a triangular L_1-mixingale array. With the same Lemma we get $\psi(n) = L_\theta^{n\varepsilon/(1-\varepsilon)}$ for $\varepsilon < \nu/2$ and some $c = c_{i,n} < \infty$. Set $B_{i,n} = n^\eta$ for $\eta > 0$ with $1/(2+\varepsilon) < \eta < 1/2$ and $a_n = \lfloor n^\varphi \rfloor$ for $0 < \varphi < 1 - 2\eta$, $i \in \mathbb{N}_0$, $n \in \mathbb{N}$ and $i \leq n$. We verify the assumptions of Theorem 5.10:

(i) We show the third condition of (i). With the same arguments as in the proof of Lemma 8.5, where we showed that $\|X_{i,n}\|_{1+\varepsilon}$ is bounded we get, that $\|X_{i,n}\|_{3+\varepsilon} < K$ for some $0 < K < \infty$. We set $r = 3+\varepsilon$ and get

$$\sum_{n=1}^{\infty} \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}[|X_{i,n}|^{3+\varepsilon}] B_{i,n}^{-(2+\varepsilon)} \leq K \sum_{n=1}^{\infty} n^{-(2+\varepsilon)\eta} < \infty.$$

(ii) With $\psi(n)$ being exponentially decreasing in n, we get that for some constant $0 < M < \infty$

$$\psi(n) \leq M n^{-(1+1/\varphi)}$$

for all $n \in \mathbb{N}$. Applying this to assumption (ii) of Theorem 5.10 yields

$$\sum_{n=1}^{\infty} \frac{1}{n} \sum_{i=1}^{n} c_{i,n} \psi(a_n) \leq cM \sum_{n=1}^{\infty} a_n^{-(1+1/\varphi)} = cM \sum_{n=1}^{\infty} (\lfloor n^{\varphi} \rfloor)^{-(1+1/\varphi)}$$

$$\leq cM \left(1 + \sum_{n=2}^{\infty} (n^{\varphi} - 1)^{-(1+1/\varphi)} \right)$$

$$= cM \left(1 + \sum_{n=2}^{\infty} n^{-(1+\varphi)} \left(1 - \frac{1}{n^{\varphi}} \right)^{-(1+1/\varphi)} \right)$$

$$\leq cM \left(1 + \left(1 - \frac{1}{2^{\varphi}} \right)^{-(1+1/\varphi)} \sum_{n=2}^{\infty} n^{-(1+\varphi)} \right) < \infty.$$

(iii) Let $\beta > 0$ and as in (ii), there exists a $0 < M < \infty$ with

$$\exp(-\beta^2 n) \leq M n^{-2/(1-2\eta-2\varphi)}$$

for all $n \in \mathbb{N}$. By using this for assumption (iii), we get

$$\sum_{n=1}^{\infty} a_n \exp\left(-\beta^2 n^2 a_n^{-2} (\sum_{i=1}^{n} B_{i,n}^2)^{-1} \right)$$

$$= \sum_{n=1}^{\infty} \lfloor n^{\varphi} \rfloor \exp\left(-\beta^2 n^2 \lfloor n^{\varphi} \rfloor^{-2} \left(\sum_{i=1}^{n} n^{2\eta} \right)^{-1} \right)$$

$$\leq \sum_{n=1}^{\infty} n^{\varphi} \exp(-\beta^2 n^2 n^{-2\varphi} n^{-(1+2\eta)})$$

$$= \sum_{n=1}^{\infty} n^{\varphi} \exp(-\beta^2 n^{1-2\eta-2\varphi})$$

$$\leq M \sum_{n=1}^{\infty} n^{\varphi-2} < \infty.$$

□

To generalize this strong law of large numbers to a ULLN, we know from Theorem 5.16 that both $Q_n(\theta)$ as in (8.5) and

$$\overline{Q}_n(\theta) := \mathrm{E}[Q_n(\theta)] \tag{8.10}$$

must be asymptotically uniformly equicontinuous, which which we prove in the following lemma.

Lemma 8.7. *Let $\{Y_{i,n}|i \in \mathbb{N}_0,\ n \in \mathbb{N},\ i \leq n\}$ be a triangular Poisson-INARCH(1) process with with intensity function $\tilde{g}_{\theta_0}(y, i|n)$ for some $\theta_0 \in \Theta$ fulfilling the same assumptions as in Theorem 8.6. Moreover, let the $(3+\nu)$-th central moments of $Y_{i,n} \|\nabla g_{\theta_0}(Y_{i-1,n})\|$ be uniformly bounded for some $\nu > 0$. Then $Q_n(\theta)$ and $\overline{Q}_n(\theta) = \mathrm{E}[Q_n(\theta)]$ are asymptotically uniformly equicontinuous.*

Proof. Let $X = \{X_{i,n}|i \in \mathbb{N}_0,\ n \in \mathbb{N},\ i \leq n\}$ be a triangular array given by

$$X_{i,n} := h(Y_{i,n}, Y_{i+k,n}) - \mathrm{E}[h(Y_{i,n}, Y_{i+k,n})]$$

where $k \in \{0, 1\}$ and $h : \mathbb{R}^2 \mapsto \mathbb{R}$ is defined as

$$h(x, y) := y \sup_{\theta\in\Theta} \|\nabla g_\theta(x)\|\,.$$

Since Θ is compact, the function h is continuous. By Lemma 3.10 the triangular array X is absolutely regular with $\beta_{X_{\cdot,n}}(j) \leq L_\theta^{j-1-k} \frac{g_\theta(0)+\delta}{1-L_\theta}$. Note that for $k = 1$, $\{\mathrm{E}[|X_{i,n}|^{1+\varepsilon}]|i \in \mathbb{N}_0,\ n \in \mathbb{N},\ i \leq n\}$ is uniformly bounded since we have

$$\begin{aligned}
E[|X_{i,n}|^{1+\varepsilon}]^{1/(1+\varepsilon)} - \mathrm{E}[h(Y_{i,n}, Y_{i+1,n})] &= \|\mathrm{E}[X_{i,n}]\|_{1+\varepsilon} - \mathrm{E}[h(Y_{i,n}, Y_{i+1,n})] \\
&\leq \|h(Y_{i,n}, Y_{i+1,n})\|_{1+\varepsilon} = \left\| Y_{i+1,n} \sup_{\theta\in\Theta} \|\nabla g_\theta(Y_{i,n})\| \right\|_{1+\varepsilon} \\
&= \mathrm{E}\left[(Y_{i+1,n} \sup_{\theta\in\Theta} \|\nabla g_\theta(Y_{i,n})\|)^{1+\varepsilon} \right]^{1/(1+\varepsilon)} \\
&= \mathrm{E}\left[\mathrm{E}[(Y_{i+1,n} \sup_{\theta\in\Theta} \|\nabla g_\theta(Y_{i,n})\|)^{1+\varepsilon} | \mathcal{F}_i^n] \right]^{1/(1+\varepsilon)} \\
&= \mathrm{E}\left[\sup_{\theta\in\Theta} \|\nabla g_\theta(Y_{i,n})\|^{1+\varepsilon}\, \mathrm{E}[Y_{i+1,n}^{1+\varepsilon} | \mathcal{F}_i^n] \right]^{1/(1+\varepsilon)}
\end{aligned}$$

and applying Lemma A.2(iii) then yields

$$\begin{aligned}
&\leq \mathrm{E}\left[\sup_{\theta\in\Theta} \|\nabla g_\theta(Y_{i,n})\|^{1+\varepsilon}\, \tilde{g}_{\theta_0}(Y_{i,n}, i+1|n)^{1+\varepsilon} \exp\left(\frac{(1+\varepsilon^2)}{2\tilde{a}_{\theta_0}(Y_{i,n}, i+1|n)} \right) \right]^{1/(1+\varepsilon)} \\
&\leq \exp\left(\frac{(1+\varepsilon^2)}{2\tilde{g}_{\theta_0}(0, i+1|n)} \right) \mathrm{E}\left[(\sup_{\theta\in\Theta} \|\nabla g_\theta(Y_{i,n})\| \,(L_{\theta_0} Y_{i,n} + \tilde{g}_{\theta_0}(0, i+1|n)))^{1+\varepsilon} \right]^{1/(1+\varepsilon)} \\
&\leq \exp\left(\frac{(1+\varepsilon^2)}{2\tilde{g}_{\theta_0}(0, i+1|n)} \right) \Bigg(\tilde{g}_{\theta_0}(0, i+1|n)^{1+\varepsilon} P(Y_{i,n} = 0) \sup_{\theta\in\Theta} \|\nabla g_\theta(0)\| \\
&\qquad + (L_{\theta_0} + \tilde{g}_{\theta_0}(0, i+1|n)^{1+\varepsilon}) \mathrm{E}\left[(Y_{i,n} \sup_{\theta\in\Theta} \|\nabla g_\theta(Y_{i,n})\|)^{1+\varepsilon} \right]^{1/(1+\varepsilon)} \Bigg)
\end{aligned}$$

which is bounded by assumption. With Lemma 3.7, we get

$$\begin{aligned}
&\left\|\mathrm{E}[X_{i,n}|\mathcal{F}_{i-j}^n]\right\|_1 \leq 6\alpha(\mathcal{F}_{i-j}^n, \mathcal{F}_i^n)^{1-1/(1+\varepsilon)} \left\|X_{i,n}\right\|_{1+\varepsilon} \\
&\leq 6\beta_{X_{\cdot,n}}(j)^{1-1/(1+\varepsilon)} \sup_{i\in\mathbb{N}_0, n\in\mathbb{N}_0, i\leq n} \left\|X_{i,n}\right\|_{1+\varepsilon} \\
&\leq \underbrace{L_\theta^{j\varepsilon/(1-\varepsilon)}}_{=:\psi(j)} \underbrace{6\left(L_\theta^{-2}\frac{g_\theta(0)+\delta}{1-L_\theta}\right)^{\varepsilon/(1+\varepsilon)} \sup_{i\in\mathbb{N}_0, n\in\mathbb{N}_0, i\leq n} \left\|X_{i,n}\right\|_{1+\varepsilon}}_{=:c_{i,n}} \\
&= c_{i,n}\psi(j),
\end{aligned}$$

where the $c_{i,n} = c < \infty$ and $\psi(j) \overset{j\to\infty}{\longrightarrow} 0$, similar to the proof of Lemma 8.5. This fulfills the first condition of an L_1-mixingale array. By continuity of h, $X_{i,n}$ is measurable with respect to $\mathcal{F}_i^n$. Therefore, $\mathrm{E}[X_{i,n}|\mathcal{F}_{i+j}^n] = X_{i,n}$ for $j \in \mathbb{N}_0$, which shows the second condition and thus X is an L_1-mixingale array. In Theorem 8.6, we have essentially proven, that an L_1-mixingale array that has constant $c_{i,n}$, exponentially decaying $\psi(j)$ and existing $(3+\nu)$-th moments for some $\nu > 0$ fulfills the assumptions of Theorem 5.10. Thus we have with the same arguments as in the former theorem

$$\frac{1}{n}\sum_{i=1}^n X_{i,n} \overset{n\to\infty}{\longrightarrow} 0.$$

We now prove the AUEC condition for $Q(n)$ by applying Lemma 5.15. Let $\theta, \theta' \in \Theta$ and define $\theta_{i,n}, \theta_{i,n}^*$ such that $g_\theta(Y_{i,n}) - g_{\theta'}(Y_{i,n}) = \nabla g_{\theta_{i,n}}(Y_{i,n})(\theta - \theta')$

and $g^2_\theta(Y_{i,n}) - g^2_{\theta'}(Y_{i,n}) = 2g_{\theta^*_{i,n}}(Y_{i,n})\nabla g_{\theta^*_{i,n}}(Y_{i,n})(\theta-\theta')$.

$$
\begin{aligned}
&|Q_n(\theta) - Q_n(\theta')| = \frac{1}{n}\left|\sum_{i=1}^{n} q(Y_{i,n}, Y_{i-1,n}, \theta) - q(Y_{i,n}, \theta')\right| \\
&= \frac{1}{n}\left|\sum_{i=1}^{n} (Y_{i,n} - g_\theta(Y_{i-1,n}))^2 - (Y_{i,n} - g_{\theta'}(Y_{i-1,n}))^2\right| \\
&= \frac{1}{n}\left|\sum_{i=0}^{n-1} g^2_\theta(Y_{i,n}) - g^2_{\theta'}(Y_{i,n}) + 2Y_{i+1,n}\left(g_\theta(Y_{i-1,n}) - g_{\theta'}(Y_{i-1,n})\right)\right| \\
&= \frac{1}{n}\left|\sum_{i=0}^{n-1} 2g_{\theta^*_{i,n}}(Y_{i,n})\nabla g_{\theta^*_{i,n}}(Y_{i,n})(\theta-\theta') + 2Y_{i+1,n}\nabla g_{\theta^*_{i,n}}(Y_{i,n})(\theta-\theta')\right| \\
&\le \frac{2\|\theta-\theta'\|}{n}\sum_{i=0}^{n-1}\left\|g_{\theta^*_{i,n}}(Y_{i,n})\nabla g_{\theta^*_{i,n}}(Y_{i,n})\right\| + Y_{i+1,n}\left\|\nabla g_{\theta^*_{i,n}}(Y_{i,n})\right\| \\
&\le \frac{1}{\|\theta-\theta'\|}\sum_{i=0}^{n-1}(L_{\theta^*_{i,n}}Y_{i,n} + g_{\theta^*_{i,n}}(0))\left\|\nabla g_{\theta^*_{i,n}}(Y_{i,n})\right\| + Y_{i+1,n}\left\|\nabla g_{\theta^*_{i,n}}(Y_{i,n})\right\| \\
&\le \frac{2\|\theta-\theta'\|}{n}\sum_{i=0}^{n-1}\left((\sup_{\theta\in\Theta} g_\theta(0)+1)Y_{i,n} + Y_{i+1,n}\right)\sup_{\theta\in\Theta}\|\nabla g_\theta(Y_{i,n})\| \\
&\quad + \sup_{\theta\in\Theta}\|\nabla g_\theta(0)\|\,\mathbb{1}(Y_{i,n}=0) \\
&\le 2\|\theta-\theta'\|\left(\sup_{\theta\in\Theta}\|\nabla g_\theta(0)\| + \frac{1}{n}\sum_{i=0}^{n-1}(\sup_{\theta\in\Theta} g_\theta(0)+1)h(Y_{i,n}, Y_{i,n}) + h(Y_{i,n}, Y_{i+1,n})\right) \\
&=: 2\|\theta-\theta'\| B_n
\end{aligned}
$$

We have a Lipschitz bound B_n which is independent of θ. If we prove that $\limsup_{n\to\infty} B_n < \infty$, then Q_n is AUEC by Lemma 5.15. Note that by compactness of Θ and continuity of g_θ and ∇g_θ we get $\sup_{\theta\in\Theta} g_\theta(0) < \infty$ and $\sup_{\theta\in\Theta}\|\nabla g_\theta(0)\| < \infty$. Moreover, with the strong law of large numbers for mixingale arrays from the first part of the proof we get almost surely

$$
\begin{aligned}
&\limsup_{n\to\infty} B_n = \lim_{n\to\infty} B_n \\
&= \sup_{\theta\in\Theta}\|\nabla g_\theta(0)\| + \frac{\sup_{\theta\in\Theta} g_\theta(0)+1}{n}\sum_{i=0}^{n-1}\mathrm{E}[h(Y_{i,n}, Y_{i,n})] \\
&\quad + \frac{2}{n}\sum_{i=0}^{n-1}\mathrm{E}[h(Y_{i,n}, Y_{i+1,n})] \\
&\le \sup_{\theta\in\Theta}\|\nabla g_\theta(0)\| + (\sup_{\theta\in\Theta} g_\theta(0)+3)\sup_{i\in\mathbb{N}_0, n\in\mathbb{N}, i\le n}\mathrm{E}[h(Y_{i,n}, Y_{i,n})]. \qquad (8.11)
\end{aligned}
$$

Since the first moments of $h(Y_{i,n}, Y_{i,n})$ and $h(Y_{i,n}, Y_{i+1,n})$ are uniformly bounded, we have $\limsup_{n\to\infty} B_n < \infty$. Moreover, since the absolute value is convex we get with Jensen's inequality

$$|\overline{Q}_n(\theta) - \overline{Q}_n(\theta')| = \left|\mathrm{E}[\overline{Q}_n(\theta) - \overline{Q}_n(\theta')]\right| \leq \mathrm{E}[|Q_n(\theta) - Q_n(\theta')|] \leq \|\theta - \theta'\| \, \mathrm{E}[B_n]$$

where the last term is equal to the last line of (8.11) and thus has a bounded limes superior and $\overline{Q}_n$ is also AUEC. □

From the prerequisites of Lemma 8.7, we get another new assumption regarding the regularity of the gradient of g_θ, again influenced by the strong assumptions in Theorem 5.10.

(A11) The $(3+\nu)$-th central moments of $Y_{i,n} \, \|\nabla g_{\theta_0}(Y_{i-1,n})\|$ are uniformly bounded for some $\nu > 0$.

We have now all the prerequisites to generalize the law of large numbers to a uniform case.

Theorem 8.8. *Let $\{Y_{i,n} | i \in \mathbb{N}_0,\ n \in \mathbb{N},\ i \leq n\}$ be a triangular Poisson-INARCH(1) process with the same assumptions as in Lemma 8.7. Then, we have almost surely*

$$\sup_{\theta \in \Theta} \left| Q_n(\theta) - \overline{Q}_n(\theta) \right| \stackrel{n\to\infty}{\longrightarrow} 0 \quad . \tag{8.12}$$

Proof. By Theorem 8.6, we have that $Q_n(\theta)$ fulfills a strong law of large numbers for all $\theta \in \Theta$. Lemma 8.7 ensures that both Q_n and $\overline{Q}_n$ are AUEC. Since Θ is compact, we can apply Theorem 5.16 and $Q_n(\theta)$ fulfills the uniform law of large numbers (8.12). □

The convergence of the minimizer now follows Theorem 5.3 if we have a sequence of identifiable unique minimizers. In the abrupt case that is guaranteed by assumption (A3). The case of the alternative in the gradual setting is not covered by this assumption, which requires to generalize the assumption to

(A3') For all $n \in \mathbb{N}$ there exists a unique minimizer $\overline{Q}_n(\theta)$ which converges to the unique minimizer $\tilde{\theta}_0$ of $Q(\theta) := \lim_{n\to\infty} Q_n(\theta)$.

Under the null hypothesis, i.e., $m = n$, this reduces to the same condition as (A3) the abrupt case. We can now go on to the main result of this subsection. We will see, that the CLS estimates converge almost surely under the new assumptions.

Theorem 8.9. *Let $\{Y_{i,n} | i \in \mathbb{N}_0,\ n \in \mathbb{N},\ i \leq n\}$ be a triangular Poisson-INARCH(1) process with with intensity function $\tilde{g}_{\theta_0}(y, i|n)$ for some $\theta_0 \in \Theta$ fulfilling the same assumptions as in Lemma 8.7 and (A3'). Moreover, let $\tilde{\theta}_n$ be the unique minimizer of $\overline{Q}_n$, $\overline{Q}_n(\theta)$ converge to $\overline{Q}(\theta)$ for all $\theta \in \Theta$ and $\tilde{\theta}_0 = \lim_{n\to\infty} \tilde{\theta}_n$ be the unique minimizer of $\overline{Q}$. Then, almost surely*

$$\left\| \hat{\theta}_n - \tilde{\theta}_0 \right\| \stackrel{n\to\infty}{\longrightarrow} 0 \tag{8.13}$$

for the CLS estimates $\hat{\theta}_n$ of $Q_n(\theta)$.

Proof. We want to apply Theorem 5.3 for this result. For this, Q_n and $\tilde{\theta}_n$ need to fulfill two requirements. First, Q_n needs to fulfill an uniform law of large numbers, which we get by Theorem 8.8. The second requirement is that the sequence $\tilde{\theta}_n$ is identifiably unique. To get this, we assume that the sequence is not identifiably unique. Then, there exists a $\nu > 0$ and sequence θ_n with $\left\|\theta_n - \tilde{\theta}_n\right\| \geq \nu$ for all $n \in \mathbb{N}$ with

$$\liminf_{n\to\infty} \left[\overline{Q}_n(\theta_n) - \overline{Q}_n(\tilde{\theta}_n)\right] = 0. \tag{8.14}$$

We have

$$\overline{Q}_n(\theta_n) - \overline{Q}_n(\tilde{\theta}_n) = \overline{Q}_n(\theta_n) - \overline{Q}(\tilde{\theta}_0) + \overline{Q}(\tilde{\theta}_0) - \overline{Q}_n(\tilde{\theta}_0) + \overline{Q}_n(\tilde{\theta}_0) - \overline{Q}_n(\tilde{\theta}_n),$$

where the second difference of the latter term converges to 0 by the ULLN and the third difference by The AUEC condition. Thus, if (8.14) holds true, then the first difference converges to 0 for a subsequence. This implies the existence of a subsequence of $\overline{Q}_n(\theta_n)$ which converges to $\overline{Q}(\tilde{\theta}_0)$. Since $\Theta_\nu := \{\theta \in \Theta : \left\|\theta - \tilde{\theta}_0\right\| \geq \nu\}$ is closed, the image space $\overline{Q}(\Theta_\nu)$ is also closed. Thus, there exists a $\theta^* \in \{\theta \in \Theta : \left\|\theta - \tilde{\theta}_0\right\| \geq \nu\}$ to which the subsequence converges. This implies $\overline{Q}(\theta^*) = \overline{Q}(\tilde{\theta}_0)$, violating the assumption that $\tilde{\theta}_0$ is the unique minimizer of $\overline{Q}$. Therefore, $\tilde{\theta}_n$ is identifiably unique. The assumptions of Theorem 5.3 are therefore fulfilled and we have almost surely

$$\left\|\hat{\theta}_n - \tilde{\theta}_0\right\| \leq \left\|\hat{\theta}_n - \tilde{\theta}_n\right\| + \left\|\tilde{\theta}_n - \tilde{\theta}_0\right\| \overset{n\to\infty}{\longrightarrow} 0.$$

□

The foremost implication of this theorem is that we now know that the CLS estimates $\hat{\theta}_n$ converge to some vector $\tilde{\theta}_0 \in \Theta$. Besides, this theorem provides us with a toolbox to prove convergence results for the estimators of the variance of the residuals under the alternative.

Corollary 8.10. *Under the assumptions of Theorem 8.9 let*

$$\hat{\varepsilon}_{i,n} = Y_{i,n} - g_{\hat{\theta}_n}(Y_{i-1,n}) \tag{8.15}$$

for all $i, n \in \mathbb{N}$ *with* $i \leq n$*. Then, almost surely*

$$\hat{\sigma}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n} \hat{\varepsilon}_{i,n}^2 \overset{n\to\infty}{\longrightarrow} \overline{Q}(\tilde{\theta}_0). \tag{8.16}$$

Proof. We have

$$\hat{\sigma}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n} \hat{\varepsilon}_{i,n}^2 = \frac{n}{n-1}\left(\overline{Q}_n(\hat{\theta}_n) - \overline{Q}_n(\tilde{\theta}_0) + \overline{Q}_n(\tilde{\theta}_0) - \overline{Q}(\tilde{\theta}_0) + \overline{Q}(\tilde{\theta}_0)\right),$$

where Theorem 8.9 and the AUEC condition guarantee convergence to 0 of $\overline{Q}_n(\hat{\theta}_n) - \overline{Q}_n(\tilde{\theta}_0)$. The ULLN guarantees, that $\overline{Q}_n(\tilde{\theta}_0) - \overline{Q}(\tilde{\theta}_0)$ converges to 0. This results in (8.16). □

The properties proven above are also useful for analyzing the residuals of the process, i.e.,

$$\varepsilon_{i,n} = Y_{i,n} - g_\theta(Y_{i-1,n}) \tag{8.17}$$

for $i \in \mathbb{N}_0$, $n \in \mathbb{N}$, $i \leq n$ and $\theta \in \Theta$. We can use the same approach as for the results above to prove the following.

Proposition 8.11. *Let $\{Y_{i,n} | i \in \mathbb{N}_0,\ n \in \mathbb{N},\ i \leq n\}$ be a triangular Poisson-INARCH(1) process with with intensity function $\tilde{g}_{\theta_0}(y, i|n)$ for some $\theta_0 \in \Theta$ fulfilling the same assumptions as in Lemma 8.5. Then the arithmetic mean of the residuals (8.17) and its expectation are AUEC and the residuals itself fulfill an ULLN with respect to Θ.*

Proof. The $\varepsilon_{i,n}$ are continuous transformations of the original process and retain the exponentially decreasing mixing rate of the original triangular array, in the same manner as the $q(Y_{i,n}, Y_{i-1,n}, \theta)$. The uniform boundedness of the $(6+\nu)$-th moments for some $\nu > 0$ of $Y_{i,n}$ gives us the mixingale property as well as the SLLN as in the proofs of Lemma 8.5 and Theorem 8.6. We get the AUEC condition for $n^{-1} \sum_{i=1}^{n} \varepsilon_{i,n}$ and $n^{-1} \sum_{i=1}^{n} \mathrm{E}[\varepsilon_{i,n}]$ by assumption (A11) as in the proof of Lemma 8.7. Finally, both the AUEC conditions and the SLLN grant us the ULLN for the residuals by Theorem 5.16. □

Finally, with assumptions (A3'), (A10) and (A11), we have new conditions which ensures desirable convergence of the CLS estimates under the gradual intensity function (8.1). Moreover, we know how the estimates of the variances of residuals behave under the alternative. In the next section, we want to examine, if these new assumptions are weak enough in the sense that they are still feasible for at least one example. Thus, we check the assumptions for a linear triangular Poisson-INARCH(1) process.

8.3 Convergence of the CLS estimates for a linear triangular Poisson-INARCH(1) process

We check the assumptions on a linear triangular Poisson-INARCH(1) process, i.e.,

$$\tilde{g}_\theta(y, i|n) = g_\theta(y) + \delta \left(\frac{i+1-m}{n} \right)_+^\gamma = \alpha y + \omega + \delta \left(\frac{i+1-m}{n} \right)_+^\gamma \tag{8.18}$$

with parameters $\theta = (\omega, \alpha)^\top \in \Theta \subset (0,1) \times \mathbb{R}_+$, $\delta, \gamma > 0$ and position of the change point $0 \leq m < n$. To check if the new assumptions are fulfilled, we start by setting up the expected CLS function

$$
\begin{aligned}
\overline{Q}_n(\theta) &= \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\left[(Y_{i,n} - g_\theta(Y_{i-1,n}))^2\right] \\
&= \frac{1}{n}\sum_{i=1}^{n} \big(\mathrm{E}[Y_{i,n}^2] - 2\alpha \mathrm{E}[Y_{i,n}Y_{i-1,n}] - 2\omega \mathrm{E}[Y_{i,n}] + \alpha^2 \mathrm{E}[Y_{i-1,n}^2] \\
&\qquad +2\alpha\omega \mathrm{E}[Y_{i-1,n}] + \omega^2\big) . \qquad (8.19)
\end{aligned}
$$

With the first derivative with respect to θ

$$
\nabla \overline{Q}_n(\theta) = \begin{pmatrix} 2\left(\omega + \frac{1}{n}\sum\limits_{i=1}^{n} \alpha \mathrm{E}[Y_{i-1,n}] - \mathrm{E}[Y_{i,n}]\right) \\ \frac{2}{n}\left(\sum\limits_{i=1}^{n} \alpha \mathrm{E}[Y_{i-1,n}^2] - 2\mathrm{E}[Y_{i,n}Y_{i-1,n}] + \omega \mathrm{E}[Y_{i-1,n}]\right) \end{pmatrix},
$$

we can for each $n \in \mathbb{N}$ find the critical point $\tilde{\theta}_n = (\tilde{\omega}_n, \tilde{\alpha}_n)^\top$

$$
\tilde{\omega}_n = \frac{1}{n}\sum_{i=1}^{n} (\mathrm{E}[Y_{i,n}] - \tilde{\alpha}_n \mathrm{E}[Y_{i-1,n}]) \qquad (8.20)
$$

$$
\tilde{\alpha}_n = \frac{\sum\limits_{i=1}^{n} \mathrm{E}[Y_{i,n}Y_{i-1,n}] - \frac{1}{n}\sum\limits_{i=1}^{n} \mathrm{E}[Y_{i,n}] \sum\limits_{i=1}^{n} \mathrm{E}[Y_{i-1,n}]}{\sum\limits_{i=1}^{n} \mathrm{E}[Y_{i-1,n}^2] - \frac{1}{n}\left(\sum\limits_{i=1}^{n} \mathrm{E}[Y_{i-1,n}]\right)^2}. \qquad (8.21)
$$

To check if this is in fact a unique minimizer of $\overline{Q}_n$, we need to check the Hesse matrix for positive definiteness. The Hesse matrix is

$$
\nabla^2 \overline{Q}_n(\theta) = \begin{pmatrix} 2 & \frac{2}{n}\sum\limits_{i=1}^{n} \mathrm{E}[Y_{i-1,n}] \\ \frac{2}{n}\sum\limits_{i=1}^{n} \mathrm{E}[Y_{i-1,n}] & \frac{2}{n}\sum\limits_{i=1}^{n} \mathrm{E}[Y_{i-1,n}^2] \end{pmatrix} \qquad (8.22)
$$

which is positive definite, if

$$
\det\left(\nabla^2 \overline{Q}_n(\theta)\right) = \frac{4}{n}\sum_{i=1}^{n} \mathrm{E}[Y_{i,n}^2] - \frac{4}{n^2}\left(\sum_{i=1}^{n} \mathrm{E}[Y_{i,n}]\right)^2 > 0.
$$

With the fact that $\mathrm{E}[Y_{i,n}^2] = \mathrm{Var}(Y_{i,n}) + \mathrm{E}[Y_{i,n}]^2$ and Jensen's inequality we get

$$
\frac{4}{n}\sum_{i=1}^{n} \mathrm{E}[Y_{i,n}^2] = \frac{4}{n}\sum_{i=1}^{n} \mathrm{Var}(Y_{i,n}) + \mathrm{E}[Y_{i,n}]^2 \geq \frac{4}{n}\sum_{i=1}^{n} \mathrm{Var}(Y_{i,n}) + \frac{4}{n^2}\left(\sum_{i=1}^{n} \mathrm{E}[Y_{i,n}]\right)^2.
$$

Since the variance is larger than 0, the determinant of the Hesse matrix is positive definite for all $\theta \in \Theta$. Thus, the critical point is a unique minimizer for all $n \in \mathbb{N}$.
Next, we are interested in the asymptotic behavior of the components of (8.19). If the arithmetic means of all of those expected values converge, then the CLS function in expectation converges. The limit is then a quadratic form with respect to θ and we can compute the minimizer in a similar manner. To determine the convergence, we need to develop closed form formulas for the terms which $\overline{Q}_n$ consists of.

Lemma 8.12. *Let $\{Y_{i,n} | i \in \mathbb{N}_0,\, n \in \mathbb{N},\, i \le n\}$ be a linear Poisson-INARCH(1) process following the intensity function from (8.18) and $(Y_i)_{i\in\mathbb{N}_0}$ a Poisson-INARCH(1) process generated by g_θ. Then the following equations hold:*

(i) $\mathrm{E}[Y_{m+i,n}] = \mathrm{E}[Y_0] + \delta \sum_{j=1}^{i} \alpha^{i-j} \left(\frac{j}{n}\right)^\gamma$,

(ii) $\mathrm{E}[\overline{Y}_n] = \mathrm{E}[Y_0] + \frac{\delta}{n^{\gamma+1}} \sum_{i=1}^{n-m} i^\gamma \frac{1-\alpha^{n-m-i+1}}{1-\alpha}$,

(iii)
$$\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}[Y_{i,n}]^2 = \mathrm{E}[Y_0]^2 + \frac{2\delta \mathrm{E}[Y_0]}{n^{\gamma+1}} \sum_{i=1}^{n-m} i^\gamma \frac{1-\alpha^{n-m-i+1}}{1-\alpha} + \frac{\delta^2}{n^{2\gamma+1}} \sum_{i=1}^{n-m} \left(\sum_{j=1}^{i} \alpha^{i-j} j^\gamma \right)^2,$$

(iv) $\mathrm{Var}(Y_{m+i,n}) = \alpha^{2i}\mathrm{Var}(Y_0) + \sum_{j=1}^{i} \alpha^{2(i-j)} \mathrm{E}[Y_{m+j,n}]$,

(v) $\frac{1}{n}\sum_{i=1}^{n} \mathrm{Var}(Y_{i,n}) = \mathrm{Var}(Y_0) + \frac{\delta}{n^{\gamma+1}} \sum_{i=1}^{n-m} \sum_{j=1}^{i} \alpha^{2(i-j)} \sum_{k=1}^{j} k^\gamma \alpha^{j-k}$,

(vi)
$$\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}[Y_{i,n} Y_{i-1,n}] = \alpha \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}[Y_{i-1,n}^2] + \frac{\mathrm{E}[\omega]}{n} \sum_{i=1}^{n} \mathrm{E}[Y_{i-1,n}] + \frac{\delta \mathrm{E}[Y_0]}{n^\gamma} \sum_{i=1}^{n-m} i^\gamma + \frac{\delta^2}{n^{2\gamma+1}} \sum_{i=1}^{n-m} i^\gamma \sum_{j=1}^{i-1} j^\gamma \alpha^{i-j}.$$

Proof. As the proof consist of lengthy straightforward calculations, we shifted the proof to the Appendix D. □

Note that since $\mathrm{E}[Y_{i,n}^2] = \mathrm{Var}(Y_{i,n}) + \mathrm{E}[Y_{i,n}]^2$, we already have a solution for the arithmetic mean of $\mathrm{E}[Y_{0,n}^2], ..., \mathrm{E}[Y_{n,n}^2]$. To find out if those arithmetic means actually converge, we analyze the convergence of the sequences we got in the lemma above.

Lemma 8.13. *For $0 < \alpha < 1$, $\omega > 0$, $\gamma > 0$ and $m_n = \lfloor \tau n \rfloor$ for $0 < \tau < 1$, the following sequences converge, i.e.,*

(i) $\frac{1}{n^\gamma} \sum_{i=1}^{n} \alpha^{n-i} i^\gamma \overset{n\to\infty}{\longrightarrow} \frac{1}{(1-\alpha)(\gamma+1)}$,

(ii) $\frac{1}{n^{\gamma+1}} \sum_{i=1}^{n-m} i^\gamma \left(1 - \alpha^{n-m-i+1}\right) \overset{n\to\infty}{\longrightarrow} \frac{(1-\tau)^{\gamma+1}}{1+\gamma}$,

(iii) $\frac{1}{n^{2\gamma+1}} \sum_{i=1}^{n-m} \left(\sum_{j=1}^{i} \alpha^{i-j} j^\gamma \right)^2 \overset{n\to\infty}{\longrightarrow} \left(\frac{1}{1-\alpha}\right)^2 \frac{(1-\tau)^{2\gamma+1}}{2\gamma+1}$,

(iv) $\frac{1}{n^{\gamma+1}} \sum_{i=1}^{n-m} \sum_{j=1}^{i} \alpha^{2(i-j)} \sum_{k=1}^{j} k^\gamma \alpha^{j-k} \overset{n\to\infty}{\longrightarrow} \frac{1}{(1-\alpha)(1-\alpha^2)} \frac{(1-\tau)^{\gamma+1}}{(\gamma+1)}$,

(v) $\frac{1}{n^{2\gamma+1}} \sum_{i=1}^{n-m} i^\gamma \sum_{j=1}^{i-1} \alpha^{i-j} j^\gamma \overset{n\to\infty}{\longrightarrow} \frac{1}{(1-\alpha)} \frac{(1-\tau)^{2\gamma+1}}{2\gamma+1}$.

Proof. We prove the convergences one after another.

(i) First, we observe that the upper limit of the sequence is $(1-\alpha)^{-1}$, by bounding $(i/n)^\gamma$ from above by one, i.e.,

$$\frac{1}{n^\gamma} \sum_{i=1}^{n} \alpha^{n-i} i^\gamma < \sum_{i=1}^{n} \alpha^{n-i} = \frac{1-\alpha^n}{1-\alpha} < \frac{1}{1-\alpha}.$$

Moreover, we get from

$$\frac{1}{(n+1)^\gamma} \sum_{i=1}^{n+1} \alpha^{n-i+1} i^\gamma = \alpha \left(\frac{n}{n+1}\right)^\gamma \frac{1}{n^\gamma} \sum_{i=1}^{n} \alpha^{n-i} i^\gamma + 1$$

that we can represent this sequence by the recursive sequence $(a_n)_{n\in\mathbb{N}}$ given by $a_{n+1} = \alpha \left(\frac{n}{n+1}\right)^\gamma a_n + 1$ for all $n > 1$ and $a_1 = 1$. Let $n, m \in \mathbb{N}$, we have

$$\begin{aligned} a_{n+m} &= \alpha \left(\frac{n+m-1}{n+m}\right)^\gamma a_{n+m-1} + 1 > \alpha \left(\frac{n}{n+1}\right)^\gamma a_{n+m} + 1 \\ &= \alpha \left(\frac{n}{n+1}\right)^\gamma \left(\alpha \left(\frac{n+m-2}{n+m-1}\right)^\gamma a_{n+m-2} + 1 \right) + 1 \end{aligned}$$

$$> \alpha^2 \left(\frac{n}{n+1}\right)^{2\gamma} a_{n+m-2} + \alpha \left(\frac{n}{n+1}\right)^{\gamma} + 1$$

$$> ... > \sum_{i=0}^{m} \alpha^i \left(\frac{n}{n+1}\right)^{i\gamma} = \frac{1-\alpha^{m+1}\left(\frac{n}{n+1}\right)^{(m+1)\gamma}}{1-\alpha\left(\frac{n}{n+1}\right)^{\gamma}}$$

$$=: b_m^n \overset{m\to\infty}{\longrightarrow} \frac{1}{1-\alpha\left(\frac{n}{n+1}\right)^{\gamma}}. \tag{8.23}$$

Now let $\varepsilon > 0$ and choose $N \in \mathbb{N}$ such that

$$0 < \frac{1}{1-\alpha} - \frac{1}{1-\alpha\left(\frac{N}{N+1}\right)^{\gamma}} < \frac{\varepsilon}{2}. \tag{8.24}$$

Since b_m^N converges, we can choose $M \in \mathbb{N}$ such that for all $m \geq M$,

$$0 < \frac{1}{1-\alpha\left(\frac{N}{N+1}\right)^{\gamma}} - \frac{1-\alpha^{m+1}\left(\frac{N}{N+1}\right)^{(m+1)\gamma}}{1-\alpha\left(\frac{N}{N+1}\right)^{\gamma}} < \frac{\varepsilon}{2}. \tag{8.25}$$

We get with (8.23)-(8.25) for all $n \geq M+N$,

$$0 < \frac{1}{1-\alpha} - a_m < \frac{1}{1-\alpha} - \frac{1}{1-\alpha\left(\frac{N}{N+1}\right)^{\gamma}} + \frac{1}{1-\alpha\left(\frac{N}{N+1}\right)^{\gamma}} - b_m < \varepsilon.$$

(ii) We start by showing that $\frac{1}{n^{1+\gamma}} \sum\limits_{i=1}^{n-m_n} i^\gamma$ converges. Consider the upper and lower bounds

$$\frac{1}{n^{1+\gamma}} \sum_{i=1}^{n-m_n} i^\gamma = \frac{1}{n^{1+\gamma}} \sum_{i=1}^{n-m_n} \inf_{x\in[i,i+1]} x^\gamma \leq \frac{(n-m_n+1)^{1+\gamma}}{n^{1+\gamma}(1+\gamma)} \leq \frac{1}{1+\gamma}$$

and

$$\frac{1}{n^{1+\gamma}} \sum_{i=1}^{n-m_n} i^\gamma = \frac{1}{n^{1+\gamma}} \sum_{i=0}^{n-m_n-1} \sup_{x\in[i,i+1]} x^\gamma \geq \frac{(n-m_n)^{1+\gamma}}{n^{1+\gamma}(1+\gamma)},$$

which we get by using the antiderivate of a function $h(x) = x^\gamma$. Subtracting the lower from the upper bound yields

$$\Rightarrow \frac{(n-m_n+1)^{1+\gamma}}{n^{1+\gamma}(1+\gamma)-1} - \frac{(n-m_n)^{1+\gamma}}{n^{1+\gamma}(1+\gamma)} \leq \frac{(n-m_n+1)^{1+\gamma}-(n-m_n)^{1+\gamma}}{n^{1+\gamma}(1+\gamma)}$$

$$= \frac{\tilde{n}^\gamma}{n^{1+\gamma}} \overset{n\to\infty}{\longrightarrow} 0$$

where $\tilde{n} \in [n - m_n, (n - m_n + 1)]$ with

$$(1+\gamma)\tilde{n}^\gamma = (n - m_n + 1)^{1+\gamma} - (n - m_n)^{1+\gamma}$$

by the mean value theorem. Since the difference of the upper and lower bound of the sequence converge to 0 and the sequence being bounded, the sequence itself converges. Moreover, then the limit is the same as the shared limit of the bounds, being

$$\frac{(n - m_n + 1)^{1+\gamma}}{n^{1+\gamma}(1+\gamma)} = \left(\frac{n - \lfloor \tau n \rfloor}{n}\right)^{\gamma+1} \frac{1}{\gamma} \overset{n\to\infty}{\longrightarrow} \frac{(1-\tau)^{\gamma+1}}{1+\gamma}.$$

For the second part of the sum it holds

$$\begin{aligned} &\frac{1}{n^{1+\gamma}} \sum_{i=1}^{n-m_n} i^\gamma \alpha^{n-m_n-i+1} = \frac{(n-m_n)^\gamma}{n^{\gamma+1}} a_{n-m_n} \\ &\leq \frac{(n-m_n)^\gamma}{n^{1+\gamma}} \frac{1}{1-\alpha} \overset{n\to\infty}{\to} 0, \end{aligned}$$

where $(a_n)_{n\in\mathbb{N}}$ is the sequence from part (i) of the proof. Since both sequences converge, the sum of both converges to $\frac{(1-\tau)^{\gamma+1}}{1+\gamma}$.

(iii) We can find the upper bound

$$\begin{aligned} &\frac{1}{n^{2\gamma+1}} \sum_{i=1}^{n-m_n} \left(\sum_{j=1}^{i} \alpha^{i-j} j^\gamma \right)^2 < \frac{1}{n^{2\gamma+1}} \sum_{i=1}^{n-m_n} i^\gamma \left(\sum_{j=1}^{i} \alpha^{i-j} \right)^2 \\ &< \frac{1}{1-\alpha} \frac{(n-m_n)^{2\gamma+1}}{(2\gamma+1)n^{2\gamma+1}} \overset{n\to\infty}{\longrightarrow} \left(\frac{1}{1-\alpha}\right)^2 \frac{(1-\tau)^{2\gamma+1}}{2\gamma+1}. \end{aligned} \tag{8.26}$$

Let $\varepsilon > 0$ and $N \in \mathbb{N}$ such that $0 < (1-\alpha)^{-1} - a_n < \varepsilon$ for all $n \geq N$. We have for all $n \geq N$ with $n - m_n > N$

$$\begin{aligned} &\frac{1}{n^{2\gamma+1}} \sum_{i=1}^{n-m_n} \left(\sum_{j=1}^{i} \alpha^{i-j} j^\gamma \right)^2 > \frac{1}{n^{2\gamma+1}} \sum_{i=N}^{n-m_n} \left(\sum_{j=1}^{i} \alpha^{i-j} j^\gamma \right)^2 \\ &= \frac{1}{n^{2\gamma+1}} \sum_{i=N}^{n-m_n} i^{2\gamma} a_i^2 > \frac{1}{n^{2\gamma+1}} \sum_{i=N}^{n-m_n} i^{2\gamma} \left(\frac{1}{1-\alpha} - \varepsilon \right)^2 \\ &\geq \left(\frac{1}{1-\alpha} - \varepsilon \right)^2 \frac{n - m_m^{2\gamma+1} - N^{2\gamma+1}}{(2\gamma+1)n^{2\gamma+1}} \overset{n\to\infty}{\longrightarrow} \left(\frac{1}{1-\alpha} - \varepsilon \right)^2 \frac{(1-\tau)^{2\gamma+1}}{2\gamma+1}. \end{aligned} \tag{8.27}$$

Thus, for all $\varepsilon > 0$ the limes inferior of the sequence is larger than the limit (8.27). Together with the upper bound we get the existence of the limit and that is equal to the upper bound (8.26).

(iv) In the same manner as before we have the upper bound

$$\frac{1}{n^{\gamma+1}}\sum_{i=1}^{n-m_n}\sum_{j=1}^{i}\alpha^{2(i-j)}\sum_{k=1}^{j}k^\gamma\alpha^{j-k} < \frac{1}{(1-\alpha)(1-\alpha^2)}\sum_{i=1}^{n-m_n} i^\gamma$$
$$\leq \frac{1}{(1-\alpha)(1-\alpha^2)}\frac{(n-m_n)^{\gamma+1}}{(\gamma+1)n^{\gamma+1}} \overset{n\to\infty}{\longrightarrow} \frac{1}{(1-\alpha)(1-\alpha^2)}\frac{(1-\tau)^{\gamma+1}}{\gamma+1}. \tag{8.28}$$

Define the sequence $(\tilde{a}_n)_{n\in\mathbb{N}}$ by $\tilde{a}_{n+1} = \alpha^2\,(n/(n+1))^\gamma\,\tilde{a}_n + 1$ and $\tilde{a}_1 = 1$. It converges with the same arguments as in (i) to $(1-\alpha^2)^{-1}$.
Let $\varepsilon > 0$ and $N \in \mathbb{N}$ as in (iii) and also, such that for all $n \geq N$ we have $0 < (1-\alpha^2)^{-1} - \tilde{a}_n < \varepsilon$, then

$$\frac{1}{n^{\gamma+1}}\sum_{i=1}^{n-m_n}\sum_{j=1}^{i}\alpha^{2(i-j)}\sum_{k=1}^{j}k^\gamma\alpha^{j-k} > \frac{1}{n^{\gamma+1}}\sum_{i=N}^{n-m_n}\sum_{j=N}^{i}\alpha^{2(i-j)}j^\gamma a_j$$
$$> \left(\frac{1}{1-\alpha}-\varepsilon\right)\frac{1}{n^{\gamma+1}}\left(\sum_{i=N}^{n-m_n} i^\gamma\tilde{a}_i - \sum_{i=N}^{n-m_n}\sum_{j=1}^{N}\alpha^{2(i-j)}j^\gamma\right)$$
$$> \left(\frac{1}{1-\alpha}-\varepsilon\right)\frac{1}{n^{\gamma+1}}\left(\left(\frac{1}{1-\alpha^2}-\varepsilon\right)\sum_{i=N}^{n-m_n} i^\gamma - N^\gamma\sum_{i=N}^{n-m_n}\sum_{j=1}^{N}\alpha^{2(i-j)}\right)$$
$$> \left(\frac{1}{1-\alpha}-\varepsilon\right)\left(\left(\frac{1}{1-\alpha^2}-\varepsilon\right)\frac{(n-m_n)^{\gamma+1}-N^{\gamma+1}}{(\gamma+1)n^{\gamma+1}} - \frac{nN^\gamma}{(1-\alpha^2)n^{\gamma+1}}\right)$$
$$\overset{n\to\infty}{\longrightarrow} \left(\frac{1}{1-\alpha}-\varepsilon\right)\left(\frac{1}{1-\alpha^2}-\varepsilon\right)\frac{(1-\tau)^{\gamma+1}}{\gamma+1}.$$

With the same arguments as before, we get that the sequence converges to the upper bound (8.28).

(v) As before we have an upper bound

$$\frac{1}{n^{2\gamma+1}}\sum_{i=1}^{n-m} i^\gamma\sum_{j=1}^{i-1}\alpha^{i-j}j^\gamma < \frac{1}{1-\alpha}\frac{1}{n^{2\gamma+1}}\sum_{i=1}^{n-m_n} i^{2\gamma} \tag{8.29}$$
$$< \frac{1}{1-\alpha}\frac{(n-m_n+1)^{2\gamma+1}}{n^{2\gamma+1}} \overset{n\to\infty}{\longrightarrow} \frac{1}{(1-\alpha)}\frac{(1-\tau)^{2\gamma+1}}{2\gamma+1}. \tag{8.30}$$

Let $\varepsilon > 0$ and $N \in \mathbb{N}$ as in (iii), then

$$\frac{1}{n^{2\gamma+1}}\sum_{i=1}^{n-m} i^\gamma\sum_{j=1}^{i-1}\alpha^{i-j}j^\gamma > \frac{1}{n^{2\gamma+1}}\sum_{i=N}^{n-m_n} i^{2\gamma}a_i$$
$$> \left(\frac{1}{1-\alpha}-\varepsilon\right)\frac{1}{n^{2\gamma+1}}\sum_{i=N}^{n-m_n} i^{2\gamma} > \left(\frac{1}{1-\alpha}-\varepsilon\right)\frac{(n-m_n-1)^{2\gamma+1}-N^{2\gamma+1}}{(2\gamma+1)n^{2\gamma+1}}$$
$$\overset{n\to\infty}{\longrightarrow} \left(\frac{1}{1-\alpha}-\varepsilon\right)\frac{(1-\tau)^{2\gamma+1}}{2\gamma+1}$$

Again, as in (iii) we get that the sequence converges to the upper limit (8.30).

□

The aim is to find the limit of the minimizer. Therefore, we need some explicit limits for the sums defining $\tilde{\theta}_n$.

Lemma 8.14. *Let $\{Y_{i,n}|i \in \mathbb{N}_0, n \in \mathbb{N}, i \leq n\}$ and $(Y_i)_{i\in\mathbb{N}_0}$ be as in Lemma 8.12 and $m_n = \lfloor \tau n \rfloor$ for $0 < \tau < 1$. Then,*

(i) $\mathrm{E}[\overline{Y}_n] \overset{n\to\infty}{\longrightarrow} \mathrm{E}[Y_0] + \frac{\delta(1-\tau)^{\gamma+1}}{(1-\alpha)(\gamma+1)}$

(ii) $\displaystyle \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}[Y_{i,n}]^2 \overset{n\to\infty}{\longrightarrow} \mathrm{E}[Y_0]^2 + \frac{2\delta\mathrm{E}[Y_0](1-\tau)^{\gamma+1}}{(1-\alpha)(\gamma+1)} + \frac{\delta^2(1-\tau)^{2\gamma+1}}{(2\gamma+1)(1-\alpha)^2}$

(iii) $\frac{1}{n}\sum_{i=1}^{n} \mathrm{Var}(Y_{i,n}) \overset{n\to\infty}{\longrightarrow} \mathrm{Var}(Y_0) + \frac{\delta}{(1-\alpha)(1-\alpha^2)} \frac{(1-\tau)^{\gamma+1}}{\gamma+1}$

Proof. For each convergence, we apply the results of Lemma 8.13 to the terms in Lemma 8.12.

(i) $$\mathrm{E}[\overline{Y}_n] = \mathrm{E}[Y_0] + \frac{\delta}{n^{\gamma+1}} \sum_{i=1}^{n-m_n} i^\gamma \frac{1-\alpha^{n-m-i+1}}{1-\alpha} \overset{n\to\infty}{\longrightarrow} \mathrm{E}[Y_0] + \frac{\delta(1-\tau)^{\gamma+1}}{(1-\alpha)(\gamma+1)}$$

(ii) $$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}[Y_{i,n}]^2 &= \mathrm{E}[Y_0]^2 + \frac{2\delta\mathrm{E}[Y_0]}{n^{\gamma+1}} \sum_{i=1}^{n-m_n} i^\gamma \frac{1-\alpha^{n-m-i+1}}{1-\alpha} \\ &\quad + \frac{\delta^2}{n^{2\gamma+1}} \sum_{i=1}^{n-m_n} \left(\sum_{j=1}^{i} \alpha^{i-j} j^\gamma \right)^2 \\ &\overset{n\to\infty}{\longrightarrow} \mathrm{E}[Y_0]^2 + \frac{2\delta\mathrm{E}[Y_0](1-\tau)^{\gamma+1}}{(1-\alpha)(\gamma+1)} + \frac{\delta^2(1-\tau)^{2\gamma+1}}{(2\gamma+1)(1-\alpha)^2} \end{aligned}$$

(iii) $$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} \mathrm{Var}(Y_{i,n}) &= \mathrm{Var}(Y_0) + \frac{\delta}{n^{\gamma+1}} \sum_{i=1}^{n-m_n} \sum_{j=1}^{i} \alpha^{2(i-j)} \sum_{k=1}^{j} k^\gamma \alpha^{j-k} \\ &\overset{n\to\infty}{\longrightarrow} \mathrm{Var}(Y_0) + \frac{\delta}{(1-\alpha)(1-\alpha^2)} \frac{(1-\tau)^{\gamma+1}}{\gamma+1} \end{aligned}$$

□

With these results we are able to prove that the new assumptions are all fulfilled by the linear model.

Theorem 8.15. *Let $\{Y_{i,n}|i \in \mathbb{N}_0,\ n \in \mathbb{N},\ i \leq n\}$ and $(Y_i)_{i\in\mathbb{N}_0}$ be a linear triangular Poisson-INARCH(1) process as in Lemma 8.12 for some $\theta_0 \in \Theta$ and $m_n = \lfloor \tau n \rfloor$ for $0 < \tau < 1$. Then, (A3'), (A10) and (A11) are fulfilled.*

Proof. We prove the assumptions one after another.

(A3') By the computations from before Lemma 8.12, we know that there exists an unique minimizer of $\overline{Q}_n$ for all $n \in \mathbb{N}$. By Lemmas 8.12 and 8.13, we get that the arithmetic means, which $\overline{Q}_n$ consists of, all converge. Thus, $\overline{Q}_n$ converges to some function $\overline{Q}$. The first and second derivative are as (8.20) and (8.22) with $\lim\limits_{n\to\infty}$ in front of them. Meaning we still have a positive determinant, since $\lim\limits_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \mathrm{Var}(Y_{i,n}) > \mathrm{Var}(Y_0) > 0$. Therefore, a critical point is a unique minimizer of $\overline{Q}$. We have

$$\frac{1}{n}\sum_{i=1}^n \mathrm{E}[Y_{i-1,n}]\sum_{i=1}^n \mathrm{E}[Y_{i,n}] = \frac{1}{n}\sum_{i=1}^n \mathrm{E}[Y_{i-1,n}]\sum_{i=1}^n \mathrm{E}[g_{\theta_0}(Y_{i-1,n}, i|n)]$$
$$= \frac{\alpha_0}{n}\left(\sum_{i=1}^n \mathrm{E}[Y_{i-1,n}]\right)^2 + \left(\omega_0 + \frac{\delta}{n^{\gamma+1}}\sum_{i=1}^n (i-m_n)_+^\gamma\right)\sum_{i=1}^n \mathrm{E}[Y_{i-1,n}]$$

and then,

$$\sum_{i=1}^n \mathrm{E}[Y_{i,n}Y_{i-1,n}] - \frac{1}{n}\sum_{i=1}^n \mathrm{E}[Y_{i,n}]\sum_{i=1}^n \mathrm{E}[Y_{i-1,n}]$$
$$= \alpha_0\sum_{i=1}^n \mathrm{E}[Y_{i-1,n}^2] + \omega_0\sum_{i=1}^n \mathrm{E}[Y_{i-1,n}] + \frac{\delta \mathrm{E}[Y_0]}{n^\gamma}\sum_{i=1}^{n-m_n} i^\gamma$$
$$+ \frac{\delta^2}{n^{2\gamma+1}}\sum_{i=1}^{n-m_n} i^\gamma \sum_{j=1}^{i-1} j^\gamma \alpha^{i-j}$$
$$- \alpha_0\left(\sum_{i=1}^n \mathrm{E}[Y_{i-1,n}]\right)^2 - \left(\omega_0 + \frac{\delta}{n^{\gamma+1}}\sum_{i=1}^n (i-m_n)_+^\gamma\right)\sum_{i=1}^n \mathrm{E}[Y_{i-1,n}]$$
$$= \alpha_0\left(\sum_{i=1}^n \mathrm{E}[Y_{i-1,n}^2] - \left(\sum_{i=1}^n \mathrm{E}[Y_{i-1,n}]\right)^2\right) + \frac{\delta \mathrm{E}[Y_0]}{n^\gamma}\sum_{i=1}^{n-m_n} i^\gamma$$
$$+ \frac{\delta^2}{n^{2\gamma}}\sum_{i=1}^{n-m_n} i^\gamma \sum_{j=1}^{i-1} j^\gamma \alpha^{i-j} - \frac{\delta}{n^{\gamma+1}}\sum_{i=1}^{n-m_n} i^\gamma \sum_{i=1}^n \mathrm{E}[Y_{i-1,n}].$$

Putting the last equation in the critical point $\tilde{\alpha}_n$, we get

$$\tilde{\alpha}_n = \frac{\sum\limits_{i=1}^n \mathrm{E}[Y_{i,n}Y_{i-1,n}] - \frac{1}{n}\sum\limits_{i=1}^n \mathrm{E}[Y_{i,n}]\sum\limits_{i=1}^n \mathrm{E}[Y_{i-1,n}]}{\sum\limits_{i=1}^n \mathrm{E}[Y_{i-1,n}^2] - \frac{1}{n}\left(\sum\limits_{i=1}^n \mathrm{E}[Y_{i-1,n}]\right)^2} \tag{8.31}$$

$$
\begin{aligned}
&= \frac{\alpha_0 \sum\limits_{i=1}^{n} \mathrm{E}[Y_{i-1,n}^2] - \alpha_0 \frac{1}{n}\left(\sum\limits_{i=1}^{n} \mathrm{E}[Y_{i-1,n}]\right)^2}{\sum\limits_{i=1}^{n} \mathrm{E}[Y_{i-1,n}^2] - \frac{1}{n}\left(\sum\limits_{i=1}^{n} \mathrm{E}[Y_{i-1,n}]\right)^2} \\
&\quad + \frac{\mathrm{E}[Y_0]\frac{\delta}{n^{\gamma}} \sum\limits_{i=1}^{n-m_n} i^{\gamma} + \frac{\delta^2}{n^{2\gamma+1}} \sum\limits_{i=1}^{n-m_n} i^{\gamma} \sum\limits_{j=1}^{i-1} j^{\gamma}\alpha^{i-j} - \frac{\delta}{n^{\gamma+1}} \sum\limits_{i=1}^{n-m_n} i^{\gamma} \sum\limits_{i=1}^{n} \mathrm{E}[Y_{i-1,n}]}{\sum\limits_{i=1}^{n} \mathrm{E}[Y_{i-1,n}^2] - \frac{1}{n}\left(\sum\limits_{i=1}^{n} \mathrm{E}[Y_{i-1,n}]\right)^2} \\
&= \alpha_0 + \frac{\mathrm{E}[Y_0]\frac{\delta}{n^{\gamma+1}} \sum\limits_{i=1}^{n-m_n} i^{\gamma} + \frac{\delta^2}{n^{2\gamma+1}} \sum\limits_{i=1}^{n-m_n} i^{\gamma} \sum\limits_{j=1}^{i-1} j^{\gamma}\alpha^{i-j} - \frac{\delta}{n^{\gamma+2}} \sum\limits_{i=1}^{n-m_n} i^{\gamma} \sum\limits_{i=1}^{n} \mathrm{E}[Y_{i-1,n}]}{\frac{1}{n} \sum\limits_{i=1}^{n} (\mathrm{Var}(Y_{i-1,n}) + \mathrm{E}[Y_{i-1,n}]^2) - \frac{1}{n^2}\left(\sum\limits_{i=1}^{n} \mathrm{E}[Y_{i-1,n}]\right)^2} \\
&\overset{n\to\infty}{\longrightarrow} \alpha_0 + \frac{\left(\frac{\delta \mathrm{E}[Y_0](1-\tau)^{\gamma+1}}{\gamma+1} + \frac{\delta^2(1-\tau)^{2\gamma+1}}{(1-\alpha_0)(2\gamma+1)} - \frac{\delta(1-\tau)^{\gamma+1}}{\gamma+1}\left(\mathrm{E}[Y_0] + \frac{\delta(1-\tau)^{\gamma+1}}{(\gamma+1)(1-\alpha_0)}\right)\right)}{\mathrm{Var}(Y_0) + \frac{\delta}{(1-\alpha_0)(1-\alpha_0^2)}\frac{(1-\tau)^{\gamma+1}}{\gamma+1} + \frac{\delta^2(1-\tau)^{2\gamma+1}(1-\alpha_0)}{(1-\alpha_0)^2}\left(\frac{1}{2\gamma+1} - \frac{1-\tau}{(\gamma+1)^2}\right)} \\
&= \alpha_0 + \frac{\delta(1-\tau)^{2\gamma+1}(1-\alpha_0)\left(\frac{1}{(2\gamma+1)} - \frac{(1-\tau)}{(\gamma+1)^2}\right)}{\frac{\omega_0}{(1+\alpha_0)\delta} + \frac{(1-\tau)^{\gamma+1}}{(1+\alpha_0)(\gamma+1)} + \delta(1-\tau)^{2\gamma+1}\left(\frac{1}{2\gamma+1} - \frac{1-\tau}{(\gamma+1)^2}\right)} \\
&=: \tilde{\alpha}_0
\end{aligned}
\tag{8.32}
$$

with Lemmas 8.13 and 8.14 and $\mathrm{Var}(Y_0) = \frac{\omega_0}{(1-\alpha_0)(1-\alpha_0^2)}$. For $\tilde{\omega}_n$ we get

$$
\begin{aligned}
\tilde{\omega}_n &= \frac{1}{n}\sum_{i=1}^{n} (\mathrm{E}[Y_{i,n}] - \tilde{\alpha}_n \mathrm{E}[Y_{i-1,n}]) \\
&\overset{n\to\infty}{\longrightarrow} (1-\tilde{\alpha}_0)\left(\mathrm{E}[Y_0] + \frac{\delta(1-\tau)^{\gamma+1}}{(1-\alpha_0)(\gamma+1)}\right) := \tilde{\omega}_0.
\end{aligned}
\tag{8.33}
$$

(A10) Since by Theorem 3.25 all moments of a linear contractive Poisson-INARCH(1) process exist, the $(6+\nu)$-th moment exist for all $\nu > 0$.

(A11) We have

$$
\begin{aligned}
E[Y_{i,n}^4 \, \|\nabla g_{\theta_0}(Y_{i-1,n})\|^4] &= \mathrm{E}[Y_{i,n}^4 \left(1 + Y_{i-1,n}^4\right)] \\
&= \mathrm{E}[Y_{i,n}^4 + Y_{i,n}^8]
\end{aligned}
$$

and by Theorem 3.25 all moments of a linear contractive Poisson-INARCH(1) process exist. Then, together with Lemma 8.4, we get the uniform boundedness of the eighth moments of the whole time series. Thus, the assumption is fulfilled

□

Remark 8.16. In the linear case, the expectation of the process is strictly monotonously increasing. In fact for $i \in \mathbb{N}_0$

$$\begin{aligned}
&E[Y_{m+i+1,n}] - \mathrm{E}[Y_{m+i,n}] \\
&= \mathrm{E}[Y_0] + \frac{\delta}{n^\gamma} \sum j = 1^{i+1} j^\gamma \alpha_0^{i-j+1} - \mathrm{E}[Y_0] - \frac{\delta}{n^\gamma} \sum j = 1^i j^\gamma \alpha_0^{i-j} \\
&= \frac{\delta}{n^\gamma} \left((i+1)^\gamma - (1-\alpha_0) \sum j = 1^i j^\gamma \alpha_0^{i-j} \right) \\
&\geq \frac{\delta}{n\gamma} \left((i+1)^\gamma - (1-\alpha_0) i^\gamma \sum j = 1^i \alpha_0^{i-j} \right) \\
&> \frac{\delta}{n^\gamma} ((i+1)^\gamma) - i^\gamma) > 0.
\end{aligned}$$

Moreover, the covariances fulfill

$$\begin{aligned}
\mathrm{Cov}(Y_{i,n}, Y_{i-1,n}) &= \mathrm{E}\left[(Y_{i,n} - \mathrm{E}[Y_{i,n}])(Y_{i-1,n} - \mathrm{E}[Y_{i-1,n}])\right] \\
&= \mathrm{E}\left[(\tilde{g}_{\theta_0}(Y_{i-1,n}) - \mathrm{E}[\tilde{g}_{\theta_0}(Y_{i-1,n})])(Y_{i-1,n} - \mathrm{E}[Y_{i-1,n}])\right] \\
&= \alpha_0 \mathrm{E}\left[(Y_{i-1,n} - \mathrm{E}[Y_{i-1,n}])^2\right] = \alpha_0 \mathrm{Var}(Y_{i-1,n}).
\end{aligned}$$

Applying these two results to (8.31) yields

$$\begin{aligned}
\tilde{\alpha}_n &= \frac{\sum_{i=1}^n \mathrm{E}[Y_{i,n}]\mathrm{E}[Y_{i-1,n}] + \alpha_0 \sum_{i=1}^n \mathrm{Var}(Y_{i-1,n}) - \frac{1}{n} \sum_{i=1}^n \mathrm{E}[Y_{i,n}] \sum_{i=1}^n \mathrm{E}[Y_{i-1,n}]}{\sum_{i=1}^n \mathrm{E}[Y_{i-1,n}]^2 + \sum_{i=1}^n \mathrm{Var}(Y_{i-1,n}) - \frac{1}{n} \left(\sum_{i=1}^n \mathrm{E}[Y_{i-1,n}] \right)^2} \\
&< \frac{\sum_{i=1}^n \mathrm{E}[Y_{i,n}]^2 + \sum_{i=1}^n \mathrm{Var}(Y_{i-1,n}) - \frac{1}{n} \sum_{i=1}^n \mathrm{E}[Y_{i,n}] \sum_{i=1}^n \mathrm{E}[Y_{i-1,n}]}{\sum_{i=1}^n \mathrm{E}[Y_{i-1,n}]^2 + \sum_{i=1}^n \mathrm{Var}(Y_{i-1,n}) - \frac{1}{n} \left(\sum_{i=1}^n \mathrm{E}[Y_{i-1,n}] \right)^2} \\
&\overset{n \to \infty}{\longrightarrow} 1.
\end{aligned}$$

So we have in particular that $\tilde{\alpha}_0$ is smaller than 1.

Example 8.17. To double check if those limits hold, we calculated in a simulation the mean absolute errors of the CLS estimate $\hat{\theta}_n$ against the calculated value of $\tilde{\theta}_0$ over 1000 repetitions. This was repeated for observation lengths n between 100 and 10000. We did this for a base parameter set with $\omega_0 = 0.3$, $\alpha_0 = 0.5$, $\gamma = 2$, $\delta = 1$ and $\tau = 0.5$. Moreover, we construct variations of the base parameter set, by changing one of the parameters. The results are displayed in Figures 8.3 and 8.4. In the simulations, the consistency of the CLS estimates shows up, as the error decreases to 0 for large observation lengths. While for smaller observation lengths, the estimator for $\tilde{\alpha}_0$ tends to underestimate the real value. Consequently, a higher value of α tends to result in higher values of $\hat{\omega}_n$ compared to $\tilde{\alpha}_0$, as $\tilde{\alpha}_0$ is included as $(1 - \tilde{\alpha}_0)$ in the solution to $\tilde{\alpha}_0$. The other factor in calculating $\hat{\omega}_n$ is the arithmetic mean of the expectations. The

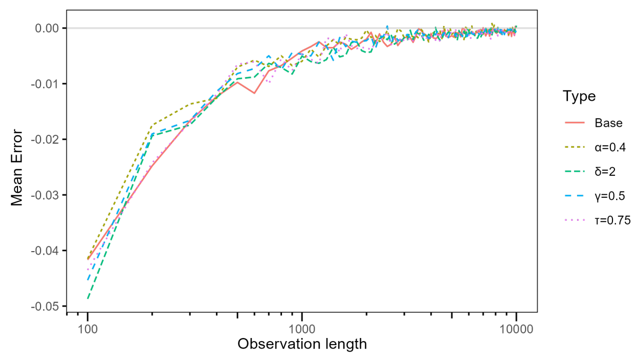

Figure 8.3: Mean error of the CLS estimates $\hat{\alpha}_n$ with respect to $\tilde{\alpha}_0$, each computed with base parameters or a variation given in the legend.

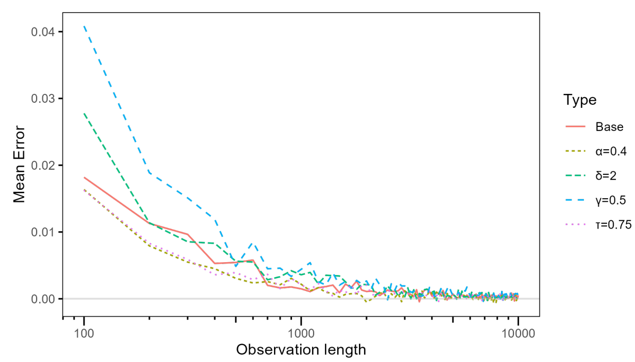

Figure 8.4: Mean error of the CLS estimates $\hat{\omega}_n$ with respect to $\tilde{\alpha}_0$, each computed with base parameters or a variation given in the legend.

limit is given by $\mathrm{E}[Y_0] + \frac{\delta(1-\tau)^{\gamma+1}}{(1-\alpha)(\gamma+1)}$, and the errors with respect to this limit are given in Figure 8.5. Those errors are for small observation lengths already about a tenth of the errors regarding $\hat{\alpha}_n$. So the error regarding $\hat{\omega}_n$ is mainly arising from the underestimation of $\tilde{\alpha}_0$.

8.4 Discussion

The gradual change model with a time-dependent intensity function is a versatile model by including different polynomial growth rates. And if desired, similar approaches can give similar results for other types of time-dependent functions. With two new and one modified assumptions compared to the abrupt change point model, the convergence of the CLS estimates can be guaranteed. The main parts of proving this have been the SLLN and AUEC condition which we achieved by having the moments of the process bounded and the mixing rate.

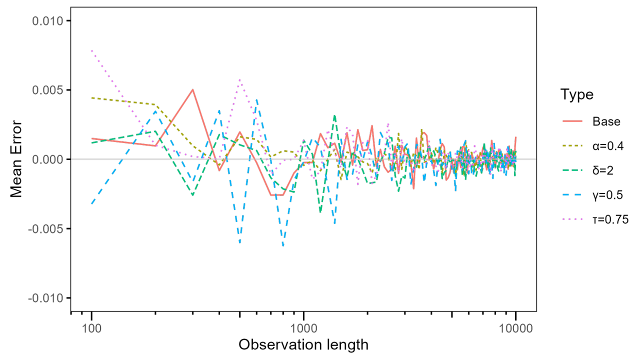

Figure 8.5: Mean error of the arithmetic with respect to its limit, each computed with base parameters or a variation given in the legend.

If the contractive part of the intensity function is linear there exists at least one model fulfilling these assumptions. Moreover, the limits of the CLS estimates can be calculated explicitly and provide a way to test these results numerically. Altogether, this is a gradual change model ready to be included in a change point test to detect gradual change points in a count time series.

9 Two test statistics for detecting a gradual change

In this section we will finally introduce the gradual change point test based on the model from Section 8. Besides the already introduced CUSUM test statistic, which is used analogously to the abrupt change model, a second test statistic is introduced. This new test statistic uses weighting factors based on the shape of the gradual change. We will see, that it also converges to a Gumbel distribution under the null hypothesis. Moreover, both test statistics will, under the alternative, converge to infinity in probability, given their respective conditions on the change itself. We conclude this section with a numerical analysis of both test statistics and a comparison of their strengths and weaknesses.

9.1 The definition of the null hypothesis and alternative

We start by setting up the model of the observations of the change point test for gradual changes. Whilst, in the change point models before, the observations were either equal to one stationary time series under the null hypothesis or they switch to a time series with a different stationary distribution. Here, for $n \in \mathbb{N}$, the observations $X_0, ..., X_n$ are given by

$$X_i = Y_{i,n} \tag{9.1}$$

for a triangular Poisson-INARCH(1) process with intensity function as in (8.1). If the change point m is equal to n, then each row of the triangular process is part of a stationary Poisson-INARCH(1) process with intensity g_θ. If $0 \le m < n$, then the gradual change starts at this point and there is a change. This leads to the following definition of a change point.

Definition 9.1. Let $(X_i)_{0 \le i \le n}$ be count time series defined as in (9.1). Then, the change point problem (for at most one change) is to test the null hypothesis

$$H_0 : m^* = n$$

against the alternative

$$H_1 : 1 \le m^* < n.$$

Under the null hypothesis, the time series is a stationary Poisson-INARCH(1) process with intensity g_θ. Hence, under the null hypothesis, the model is the same as in the test for an abrupt change point. Thus, Theorems 6.2 and 6.3 as well as Corollary 6.5 hold true under the null hypothesis. Moreover, if we apply the CUSUM test statistic (6.5) to this case, i.e.,

$$T_n^C = \max_{1 \le k < n} \sqrt{\frac{n}{k(n-k)}} |\hat{S}_n(k)|, \tag{9.2}$$

where

$$\hat{S}_n(k) = \sum_{i=1}^{k} \Big(Y_{i,n} - g_{\hat{\theta}_n}(Y_{i-1,n}) \Big) = \sum_{i=1}^{k} \hat{\varepsilon}_{i,n}, \tag{9.3}$$

and the sample residuals $\hat{\varepsilon}_{i,n}$ from (8.15), we get convergence of the test statistic to a Gumbel distribution, as in Theorem 6.4. Consequentially, when applying the CUSUM test statistic to this model, we only need to prove new convergence results under the alternative. Equivalents of Theorem 6.2 and Corollary 6.5, guaranteeing the convergence of the CLS estimate, are already given with Theorem 8.9 and Corollary 8.10. It remains to introduce a new assumption similar to (A9) which ensures convergence of the test statistic to infinity. Before doing this, we present in the following section a second test statistic, specifically designed for gradual changes of this type.

9.2 A test statistic of weighted partial sums

In Hušková and Steinebach [2000], the authors not only introduce a way of modeling gradual changes, which inspired the time dependent gradual change model in this work. They also give a test statistic particularly defined for this case, given as

$$T_n^W = \max_{k=1,\ldots,n} \frac{\left|\sum_{i=2}^n (i-k)_+^\gamma \hat{\varepsilon}_{i,n}\right|}{(\sum_{i=1}^n (i-k)_+^{2\gamma})^{1/2}}, \tag{9.4}$$

where γ is the shape parameter from the intensity function $\tilde{g}_\theta$ and the sample residuals $\hat{\varepsilon}_{i,n}$ from (8.15). The difference to the CUSUM test statistic are the weighting factors. While in the CUSUM case, the sum of the residuals gets weighted, in this case each residual gets weighted before adding them up. The order of the weighting factor for a given k in the CUSUM case is $O(\sqrt{n}^{-1})$. If we assume the residuals are equal to 1, we get order $O(n^{\gamma+1})$ of the numerator and $O(n^{\gamma+1/2})$ of the denominator in the new test statistic. Thus, the weighting factor is similar to the CUSUM case with an overall order of also $O(\sqrt{n}^{-1})$. The intuition behind this approach is, that if under H_0 the residuals converge fast enough to 0, the numerator is of order $O_P(n^{\gamma+1/2})$ or less. Then, the fraction would not grow to infinity. Under the alternative we could imagine the residuals being ideally $\varepsilon_{k+i,n} \simeq \delta \left(\frac{i}{n}\right)^\gamma$ for some $k > 0$. Which would lead to an order of $O_P(n^{\gamma+1})$ for the numerator and $O_P(\sqrt{n})$ overall, meaning that the test statistic is consistent under the alternative. In other words, the test statistic converges in probability to infinity. The rest of the current subsection will treat the asymptotics of the test statistic under the null hypothesis. Thereafter, we will state results for the convergence of both test statistics under the alternative.

Hušková and Steinebach [2000] use a sequence of results to get from an invariance principle for a weighted sum of i.i.d. random variables to the desired convergence of the gradual test statistic to a Gumbel distribution. We will do the same for general strongly mixing processes. The proofs are for the most part identical. Because of the length of those proofs we refer to the paper for detailed proofs and present only the parts which differ. The first result is that, if a weight function fulfills some regularity conditions, then a weighted i.i.d. time series converges to a weighted Wiener process. These regularity conditions are:

(B1) The function $h = h(y)$, $y \in \mathbb{R}$, is nondecreasing, square integrable and absolutely continuous on $[0,1]$ with derivative h', $h(y) = 0$ for $y \leq 0$ and $h(y) > 0$ if $y > 0$. Moreover,

$$\limsup_{t \downarrow 0} \frac{th^2(t)}{\int_0^t h^2(y)dy} < \infty.$$

The proof in the original paper starts initially by using an invariance principle and thereafter only properties of the Wiener process and the weight function. Thus, we can rewrite this result to stationary and strongly mixing time series, since they also fulfill an invariance principle.

Theorem 9.2. *Let $(Z_i)_{i \in \mathbb{N}_0}$ be a stationary and strongly mixing time series, centered at expectation and having $(2+\nu)$-th moments exist with $0 < \nu \leq 1$. Let the weight function $h = h(y)$ fulfill assumption (B1). Then, there exists a Wiener process $W_n(t)$, $t \in [0,1], n = 1, 2, ...$ such that as $n \to \infty$*

$$n^{\beta} \sup_{1/n \leq t \leq 1} t^{\beta} \left| \frac{\sum_{i=1}^{n-1} h(([nt]-i)/n)Z_i}{\sigma(\sum_{i=1}^{n-1} h^2(([nt]-i)/n))^{1/2}} - \frac{\int_0^t h'(y)W_n(t-y)dy}{(\int_0^t h^2(y)dy)^{1/2}} \right| = O_P(1) \tag{9.5}$$

for $\beta > 0$ and

$$\sigma^2 = \mathrm{E}[Z_0^2] + 2\sum_{i=1}^{\infty} \mathrm{E}[Z_0 Z_i] < \infty \tag{9.6}$$

Proof. The proof follows the arguments from Theorem 2.1 in Hušková and Steinebach [2000]. Only the invariance principle is interchanged with that from Theorem 3.11, which leads to

$$\max_{1 \leq k \leq n} \left(\frac{1}{k}\right)^{1/2-\beta} \left| \sum_{i=1}^{k} \frac{Z_i}{\sigma} - W(k) \right| = O_P(1)$$

for a standard Brownian motion W, a constant $\beta > 0$ and σ^2 as in (3.6). The rest of the calculations derive from this fact, by proving convergence is maintained while transforming it to (9.5). The final bound of the order of convergence is $O\left(\frac{1}{2} - \frac{1}{2} + \beta\right) = O(\beta)$. □

Remark 9.3. If the gradual change is of the form (8.4), we get a weight function $h(y) = \delta y_+^{\gamma}$, which is absolutely continuous on $[0,1]$, differentiable with $h'(y) = \delta\gamma y_+^{\gamma-1}$. Moreover $h(y) = 0$ for $y \leq 0$ and $h(y) > 0$ for $y > 0$. Lastly,

$$\limsup_{t \downarrow 0} \frac{th^2(t)}{\int_0^t h^2(y)dy} = \limsup_{t \downarrow 0} \frac{\delta^2 t^{2\gamma+1}(2\gamma+1)}{\delta^2 t^{2\gamma+1}} = (2\gamma+1)$$

meaning, the gradual increase function defined in (8.4) fulfills (B1).

The asymptotics under the null hypothesis are derived from Theorem 9.2, which is used in the form of the next corollary.

Corollary 9.4. *Under the assumptions of Theorem 9.2, as $n \to \infty$,*

$$\left| \max_{\log(n) \le k \le n} \frac{\left|\sum_{i=1}^{n-1} h((k-i)/n) Z_i\right|}{\sigma(\sum_{i=1}^{n-1} h^2((k-i)/n))^{1/2}} - \sup_{\log(n)/n \le t \le 1} \frac{\left|\int_0^t h'(y) W_n(t-y) dy\right|}{(\int_0^t h^2(y) dy)^{1/2}} \right|$$
$$= o_P((\log\log n)^{-1/2}) \tag{9.7}$$

for σ as in (9.6).

Proof. The proof works as in the respective proof for Corollary 2.1 in Hušková and Steinebach [2000], since we still have

$$\left| \max_{\log(n) \le k \le n} \frac{\left|\sum_{i=1}^{n-1} h((k-i)/n) Z_i\right|}{\sigma(\sum_{i=1}^{n-1} h^2((k-i)/n))^{1/2}} - \sup_{\log(n)/n \le t \le 1} \frac{\left|\int_0^t h'(y) W_n(t-y) dy\right|}{(\int_0^t h^2(y) dy)^{1/2}} \right|$$
$$\le \sup_{\log(n)/n \le t \le 1} \frac{t^\beta}{t^\beta} \left| \frac{\sum_{i=1}^{n-1} h(([nt]-i)/n) Z_i}{\sigma(\sum_{i=1}^{n-1} h^2(([nt]-i)/n))^{1/2}} - \frac{\int_0^t h'(y) W_n(t-y) dy}{(\int_0^t h^2(y) dy)^{1/2}} \right|$$
$$\le \frac{1}{\log(n)^\beta} \sup_{\log(n)/n \le t \le 1} t^\beta \left| \frac{\sum_{i=1}^{n-1} h(([nt]-i)/n) Z_i}{\sigma(\sum_{i=1}^{n-1} h^2(([nt]-i)/n))^{1/2}} - \frac{\int_0^t h'(y) W_n(t-y) dy}{(\int_0^t h^2(y) dy)^{1/2}} \right|$$
$$= O_P((n \log(n))^{-\beta}) = o_P(\log(\log(n))^{-1/2}).$$

□

Using this corollary, we are able to prove the convergence of weighted strongly mixing time series to a Gumbel distribution. Note that we do not have any more assumptions on the process itself, aside from the mixing rate and stationarity. So the results including the next theorem are applicable to gradual change models for general strongly mixing time series.

Theorem 9.5. *Let $(Z_i)_{i \in \mathbb{N}_0}$ be a stationary and strongly mixing time series, centered at expectation and having $(2+\nu)$-th moments exist with $0 < nu \le 1$. Then*

$$\lim_{n\to\infty} P\left(a_n \max_{1 \le k \le n} \frac{|\sum_{i=1}^{k-1} (k-i)^\gamma Z_i|}{\sigma(\sum_{i=1}^{k-1} i^{2\gamma})^{1/2}} \le x + b_n(\gamma) \right) = exp(-2e^{-x})$$

for all x, σ as in (9.6), where

$$a_n = (\log\log n)^{1/2}$$

and in cases

(i) $\gamma > \frac{1}{2}$:
$b_n(\gamma) = 2\log\log n + \log\left(\frac{1}{4\pi}\left(\frac{2\gamma+1}{2\gamma-1}\right)^{1/2}\right)$,

(ii) $\gamma = \frac{1}{2}$:
$b_n(\gamma) = 2\log\log n + \frac{1}{2}\log\log\log\log n - \log(4\pi)$.

Proof. The proof again follows the argument in Theorem 3.1 from Hušková and Steinebach [2000]. The first difference is that we need to use the law of iterated logarithm from Theorem 3.12 to achieve

$$\begin{aligned}\max_{1\leq k\leq \log n}\frac{|\sum_{i=1}^{k-1}(k-i)^{\gamma}Z_i|}{\sigma(\sum_{i=1}^{k-1}i^{2\gamma})^{1/2}} &= \max_{1\leq k\leq \log n}\frac{|\sum_{i=1}^{k-1}(i^{\gamma}-(i-1)^{\gamma})\sum_{j=1}^{k-i}Z_j|}{\sigma(\sum_{i=1}^{k-1}i^{2\gamma})^{1/2}}\\ &= O_P\left(\max_{1\leq k\leq \log n}\frac{k^{\gamma}(k\log\log n)^{1/2}}{k^{\gamma+1/2}}\right) = O_P((\log\log\log n)^{1/2})\\ &= o_P((\log\log n)^{1/2}).\end{aligned}$$

The second and final difference is that we thereafter can apply Corollary 9.4 instead of the analogous corollary of the original paper. This reduces the asymptotics of the weighted strongly mixing time series to those of the weighted Wiener process. Then, the rest of the proof is as in the proof of Theorem 3.12 in Hušková and Steinebach [2000]. □

We can now apply this result to our case of Poisson-INARCH(1) processes under the condition that θ_0 and σ are known. This is achieved by rearranging the weighted sums.

Theorem 9.6. *Let $(Y_i)_{i\in\mathbb{N}_0}$ be a Poisson INARCH(1)-process with a contractive intensity function and having $(2+nu)$-th moments exist with $0 < nu \leq 1$. Then, under H_0*

$$\lim_{n\to\infty} P\left(a_n \max_{k=1,\dots,n-1}\frac{|\sum_{i=1}^{n}(i-k)_+^{\gamma}\varepsilon_i|}{\sigma(\sum_{i=1}^{n}(i-k)_+^{2\gamma})^{1/2}} \leq x + b_n(\gamma)\right) = exp(-2e^{-x})$$

for all x, where $\varepsilon_i = Y_i - g_{\theta_0}(Y_{i-1})$, $\hat{\sigma}$ as in (8.16) and a_n, $b_n(\gamma)$ are as in Theorem 9.5 according to the cases $\alpha > \frac{1}{2}$ and $\alpha = \frac{1}{2}$.

Proof. We get

$$(Z_j)_{j=1,..,n} := (\varepsilon_{n+1-j}/\sigma)_{j=1,..,n} \stackrel{\mathcal{D}}{=} (\varepsilon_j/\sigma)_{j=1,..,n},$$

by replacing $i = n+1-l$, $k = n+1-l$. Note that for all $i \in \mathbb{N}$, we have

$$\begin{aligned}\mathrm{E}[\varepsilon_1\varepsilon_{i+1}] &= \mathrm{E}[Y_1Y_{i+1}] - \mathrm{E}[Y_1g_\theta(Y_i)] - \mathrm{E}[g_\theta(Y_0)Y_{i+1}] + \mathrm{E}[g_\theta(Y_0)g_\theta(Y_i)]\\ &= \mathrm{E}[Y_1g_\theta(Y_i)] - \mathrm{E}[Y_1g_\theta(Y_i)] - \mathrm{E}[g_\theta(Y_0)g_\theta(Y_i)] + \mathrm{E}[g_\theta(Y_0)g_\theta(Y_i)] = 0\end{aligned}$$

and thus with (9.6), we get $\sigma = \mathrm{E}[\varepsilon_1^2] = \mathrm{Var}(\varepsilon_1)$. This time series fulfills the assumptions of Theorem 9.5 and we get the desired convergence as in the proof of Theorem 3.2 in Hušková and Steinebach [2000]. □

Finally, we need to verify if the asymptotics are retained with the CLS estimates instead of the real parameters.

Theorem 9.7. *Let $(Y_i)_{i\in\mathbb{N}_0}$ be a Poisson INARCH(1)-process with a contractive intensity function fulfilling (A1) to (A8) and having $(2+nu)$-th moments exist with $0 < nu \leq 1$. Then, under H_0*

$$\lim_{n\to\infty} P\left(a_n \frac{T_n^W}{\hat{\sigma}_n} \leq x + b_n\right) = exp(-2e^{-x})$$

for all x, where $\hat{\sigma}_n^2 = \mathrm{Var}(\hat{\varepsilon}_i)$ and a_n, $b_n = b_n(\gamma)$ are as in Theorem 9.5 according to the cases $\gamma > \frac{1}{2}$ and $\gamma = \frac{1}{2}$.

Proof. With assumptions (A1) to (A8), Theorem 6.2 and Corollary 6.5 we get for CLS estimates $\hat{\theta}_n$ and $\hat{\sigma}_n$ as in (6.8)

$$\sqrt{\frac{n}{\log(\log(n))}}\left(\hat{\theta}_n - \theta_0\right) = o_P(1) \tag{9.8}$$

and

$$\hat{\sigma}_n = \sigma + O_P(\frac{1}{\sqrt{n}}) \tag{9.9}$$

where σ is the variance of the residuals ε_i.

Set $i = n+1-j$; $k = n+1-\ell$; $\varepsilon_{n+1-j} = Z_j$, analogous to the proof of Theorem 9.6. With this we will consider

$$\max_{\ell=1,\dots,n} \frac{\sum_{j=1}^{\ell-1}(\ell-j)^\gamma \varepsilon_j}{\sigma(\sum_{j=1}^{\ell-1} j^{2\gamma})^{1/2}}.$$

Now define

$$S_\ell = \sum_{j=1}^{\ell-1}(\ell-j)^\gamma \varepsilon_j,$$
$$\hat{S}_\ell = \sum_{j=1}^{\ell-1}(\ell-j)^\gamma \hat{\varepsilon}_j.$$

If we can show that

$$\max_{\ell=1,\dots,n} \frac{\hat{S}_\ell}{\mathrm{Var}(\hat{S}_\ell)^{1/2}} = \max_{\ell=1,\dots,n} \frac{S_\ell}{\sigma(\sum_{j=1}^{\ell-1} j^{2\gamma})^{1/2}} + o_P(1), \tag{9.10}$$

then Theorem 9.7 follows from Theorem 9.6, since then it is ensured that both sides of (9.10) converge to the same distribution. To prove this, we apply a similar transformation as in Hušková and Steinebach [2000] to the term on the left hand side of (9.10)

$$\frac{\hat{S}_\ell}{\hat{\sigma}_n(\sum_{j=1}^{\ell-1} j^{2\gamma})^{1/2}} = \frac{S_\ell}{\sigma(\sum_{j=1}^{\ell-1} j^{2\gamma})^{1/2}} \frac{\sigma}{\hat{\sigma}_n} + \frac{\sum_{j=2}^{\ell-1}(\ell-j)^\gamma \left(g_{\theta_0}(Y_{j-1}) - g_{\hat{\theta}_n}(Y_{j-1})\right)}{\hat{\sigma}_n(\sum_{j=1}^{\ell-1} j^{2\gamma})^{1/2}}. \tag{9.11}$$

Our aim is to show the convergence of the form ratio of variances converges to 1 and the latter additive part converges to 0. The convergence of the ratio of the variances is guaranteed by (9.9). To get the convergence of the latter part, note that interpreting $\sum_{j=1}^{\ell-1}(l-j)^\gamma$ as a lower Riemann sum of a function $f(x) = x^\gamma$, we get an upper bound by its integral, i.e.,

$$\sum_{j=1}^{\ell-1}(l-j)^\gamma \leq \frac{\ell^{\gamma+1}}{\gamma+1}.$$

Note that the time series $(X_i)_{i\in\mathbb{N}_0}$ defined by

$$X_i = \sup_{\theta\in\Theta} \|\nabla_\theta g_\theta(Y_i)\| - \mathrm{E}[\sup_{\theta\in\Theta} \|\nabla_\theta g_\theta(Y_i)\|]$$

is a continuous transformation of $(Y_i)_{i\in\mathbb{N}_0}$ and therefore absolutely regular with exponentially decaying rate. Moreover, by assumption the $(2+\nu)$-th moments exist and $(X_i)_{i\in\mathbb{N}_0}$ fulfills Theorem 3.12. Thus, we get

$$\sum_{j=1}^{\ell-1} \sup_{\theta\in\Theta} \|\nabla_\theta g_\theta(Y_{j-1})\| = O_P(\sqrt{\ell \log(\log(\ell))}). \tag{9.12}$$

We can now apply (9.8), (9.9) and (9.12) to the second part in (9.11) and get

$$\begin{aligned}
&\left|\frac{\sum_{j=1}^{\ell-1}(\ell-j)^\gamma \left(g_{\theta_0}(Y_{j-1}) - g_{\hat{\theta}_n}(Y_{j-1})\right)}{\hat{\sigma}_n(\sum_{j=1}^{\ell-1} j^{2\gamma})^{1/2}}\right| \\
&\leq \frac{\sum_{j=1}^{\ell-1}(\ell-j)^\gamma \sup_{\theta\in\Theta} \|\nabla_\theta g_\theta(Y_{j-1})\| \left\|\hat{\theta}_n - \theta_0\right\|}{\hat{\sigma}_n(\sum_{j=1}^{\ell-1} j^{2\gamma})^{1/2}} \\
&\leq \left\|\hat{\theta}_n - \theta_0\right\| \frac{\ell^\gamma \sum_{j=1}^{\ell-1} \sup_{\theta\in\Theta} \|\nabla_\theta g_\theta(Y_{j-1})\|}{(\sigma^2 + O_P(1/\sqrt{n}))O(\sqrt{\ell})}
\end{aligned}$$

$$= o_P\left(\sqrt{\frac{\log(\log(n))}{n}}\right)\frac{O_P(\sqrt{\ell\log(\log(\ell))})}{(\sigma^2+O_P(1/\sqrt{n}))O(\ell^{1/2})}$$

$$= o_P\left(\sqrt{\frac{\log(\log(n))}{n}}\right)\frac{O_P(\sqrt{\log(\log(\ell))})}{(\sigma^2+O_P(1/\sqrt{n}))O(1)} \overset{n\to\infty}{\to} 0 \tag{9.13}$$

which holds uniformly over all $\ell < n$. Hence, equation (9.10) holds. □

This theorem guarantees that the weighted test statistic converges to a Gumbel distribution. Moreover, the CUSUM test statistic is known to converge to a Gumbel distribution by Franke et al. [2012], since the null hypothesis is the same as for the abrupt change point. This means, we have two test statistics, which yield asymptotic level α tests. It remains to show that these test statistics are also consistent under the alternative. This is the goal of the next subsection.

9.3 Both test statistics under the alternative

As in the test for an abrupt change, we need some condition ensuring that the change is recognizable, see (A9). For the gradual change and the CUSUM test statistic, the assumption is similar. We only drop the absolute value, since we know in which direction the change will occur.

(A9') $\mathrm{E}[g_{\tilde{\theta}_0}(Y_{i-1})] - \mathrm{E}[Y_i] = C > 0$ for $i \in \mathbb{N}$.

Example 9.8. Let $\{Y_{i,n} | i \in \mathbb{N}_0,\ n \in \mathbb{N},\ i \leq n\}$ be a linear triangular Poisson-INARCH(1) process. With $\tilde{\theta}_0 = (\tilde{\omega}_0, \tilde{\alpha}_0)^\top$ with $\tilde{\omega}_0$ as in (8.33) and $\tilde{\alpha}_0$ as in (8.32), we have

$$\mathrm{E}[g_{\tilde{\theta}_0}(Y_{i-1})] - \mathrm{E}[Y_i] = \tilde{\alpha}_0\mathrm{E}[Y_0] + \tilde{\omega}_0 - \mathrm{E}[Y_0] = \tilde{\omega}_0 - (1-\tilde{\alpha}_0)\mathrm{E}[Y_0]$$
$$= (1-\tilde{\alpha}_0)\left(\mathrm{E}[Y_0] + \frac{\delta(1-\tau)^{\gamma+1}}{(1-\alpha_0)(\gamma+1)}\right) - (1-\tilde{\alpha}_0)\mathrm{E}[Y_0]$$
$$= \frac{(1-\tilde{\alpha}_0)\delta(1-\tau)^{\gamma+1}}{(1-\alpha_0)(\gamma+1)} > 0$$

and thus (A9') is always fulfilled for a linear triangular Poisson-INARCH(1) process.

Before going on with the asymptotics of the test statistic under the alternative, we introduce a naming convention:

Definition 9.9. Let $\{Y_{i,n} | i \in \mathbb{N}_0,\ n \in \mathbb{N},\ i \leq n\}$ be a triangular Poisson-INARCH(1) process with intensity function $\tilde{g}_\theta$ as in (8.1). We say it fulfills assumptions (A1) to (A11), if the Poisson-INARCH(1) process with intensity function g_θ fulfills assumptions (A1), (A2) and (A4) to (A8) and the triangular process fulfills (A3'), (A9'), (A10) and (A11).

Since the assumption on the CLS estimates under the alternative is only concerned with the process before the change point, we need some understanding on what happens after the change point. In contrast to the process before the change point, the difference in (A9') after the change point does not equal one single value for each i. We would even expect it to start with a difference close to C directly after the change point, decreasing below 0 and a negative final value. So instead of the difference at one single i, we are concerned with the sum of all those differences after the change point. The asymptotic behavior is given by the next lemma.

Lemma 9.10. *Let $\{Y_{i,n} | i \in \mathbb{N}_0,\ n \in \mathbb{N},\ i \leq n\}$ be a triangular Poisson-INARCH(1) process fulfilling assumptions (A1) to (A11). Then, under H_1*

$$\sum_{i=m_n+1}^{n} \mathrm{E}[Y_{i,n}] - \mathrm{E}[g_{\tilde{\theta}_0}(Y_{i-1,n})] = m_n C + o_P(n).$$

Proof. The AUEC condition for the expectation of the residuals from Proposition 8.11 gives us for some constant $K > 0$

$$\begin{aligned} &\left| \sum_{i=m_n+1}^{n} \mathrm{E}[g_{\tilde{\theta}_0}(Y_{i-1,n})] - \mathrm{E}[g_{\tilde{\theta}_n}(Y_{i-1,n})] \right| \\ &= \left| \sum_{i=1}^{n} Y_{i,n} - \mathrm{E}[g_{\tilde{\theta}_0}(Y_{i-1,n})] - \big(Y_{i,n} - \mathrm{E}[g_{\tilde{\theta}_n}(Y_{i-1,n})] \big) \right| \\ &\leq nK \left\| \tilde{\theta}_0 - \tilde{\theta}_n \right\| = o_P(n). \end{aligned} \tag{9.14}$$

With the contractive part of the intensity function having the structure (6.1) and $\tilde{\theta}_n$ being the critical point of the CLS function in expectation, we get

$$\sum_{i=1}^{n} \mathrm{E}[Y_{i,n}] - \mathrm{E}[g_{\tilde{\theta}_n}(Y_{i-1,n})] = 0$$

and with (A9') there exists a $C > 0$ for which we have

$$\sum_{i=m_n+1}^{n} \mathrm{E}[Y_{i,n}] - \mathrm{E}[g_{\tilde{\theta}_n}(Y_{i-1,n})] = \sum_{i=1}^{m_n} \mathrm{E}[g_{\tilde{\theta}_n}(Y_{i-1,n})] - \mathrm{E}[Y_{i,n}] = m_n C. \tag{9.15}$$

If we expand (9.15) and apply (9.14), we get

$$\begin{aligned} m_n C &= \sum_{i=m_n+1}^{n} \mathrm{E}[Y_{i,n}] - \mathrm{E}[g_{\tilde{\theta}_0}(Y_{i-1,n})] + \mathrm{E}[g_{\tilde{\theta}_0}(Y_{i-1,n})] - \mathrm{E}[g_{\tilde{\theta}_n}(Y_{i-1,n})] \\ &= \sum_{i=m_n+1}^{n} \mathrm{E}[Y_{i,n}] - \mathrm{E}[g_{\tilde{\theta}_0}(Y_{i-1,n})] + o_P(n), \end{aligned}$$

which is equivalent ot the statement of the lemma. □

The next step is to check how the test statistic itself behaves asymptotically under the alternative.

Theorem 9.11. *Let $\{Y_{i,n} | i \in \mathbb{N}_0,\ n \in \mathbb{N},\ i \leq n\}$ be a triangular Poisson-INARCH(1) process fulfilling assumptions (A1) to (A11). Then, under H_1 for the CUSUM test statistic (9.2) it holds*

$$P\left(a(\log(n))\frac{T_n^C}{\hat{\sigma}_n} - b(\log(n)) \geq c\right) \overset{n\to\infty}{\longrightarrow} 1 \tag{9.16}$$

for all $c > 0$ and with $a(\log(n))$ and $b(\log(n))$ as in Theorem 6.4.

Proof. To proof this theorem, we start by showing that the cumulative sums $\hat{S}_n(m_n)$ at the position of the change point grow linearly in n for $n \longrightarrow \infty$. This results in growth rate for T_n^C of at least $\sqrt{n}$ for $n \longrightarrow \infty$. Comparing this growth rate to those of $a(\log(n))$ and $b(\log(n))$ results in the statement of this theorem. We begin similar to the proof of Lemma 9.10, the residuals with respect to the CLS estimates $\hat{\theta}_n$ sum up to 0, resulting in

$$\hat{S}_n(m_n) = \sum_{i=1}^{m_n} Y_{i,n} - g_{\hat{\theta}_n}(Y_{i-1,n}) = -\sum_{i=m_n+1}^{n} Y_{i,n} - g_{\hat{\theta}_n}(Y_{i-1,n}).$$

Regarding the sum of the residuals up to the change point, we have with the ULLN and AUEC conditions for the residuals from Proposition 8.11 and assumption (A9')

$$\begin{aligned}
\hat{S}_n(m_n) &= \sum_{i=1}^{m_n} Y_{i,n} - g_{\hat{\theta}_n}(Y_{i-1,n}) \\
&= \sum_{i=1}^{m_n} Y_{i,n} - g_{\hat{\theta}_n}(Y_{i-1,n}) - \mathrm{E}[Y_{i,n}] + \mathrm{E}[Y_{i,n}] - \mathrm{E}[g_{\tilde{\theta}_0}(Y_{i-1,n})] + \mathrm{E}[g_{\tilde{\theta}_0}(Y_{i-1,n})] \\
&= \sum_{i=1}^{m_n} \Big[Y_{i,n} - g_{\hat{\theta}_n}(Y_{i-1,n}) - \Big(\mathrm{E}[Y_{i,n}] - \mathrm{E}[g_{\hat{\theta}_0}(Y_{i-1,n})]\Big) + \mathrm{E}[Y_{i,n}] - \mathrm{E}[g_{\hat{\theta}_0}(Y_{i-1,n})] \\
&\qquad - \Big(\mathrm{E}[Y_{i,n}] - \mathrm{E}[g_{\tilde{\theta}_0}(Y_{i-1,n})]\Big) + \mathrm{E}[Y_{i,n}] - \mathrm{E}[g_{\tilde{\theta}_0}(Y_{i-1,n})]\Big] \\
&= \sum_{i=1}^{m_n} \hat{\varepsilon}_{i,n} - \mathrm{E}[\hat{\varepsilon}_{i,n}] + \sum_{i=1}^{m_n} \mathrm{E}[\tilde{\varepsilon}_{i,n}] - \mathrm{E}[\hat{\varepsilon}_{i,n}] + m_n C \\
&= m_n C + o_P(n).
\end{aligned} \tag{9.17}$$

For the part after the change point, we additionally apply Lemma 9.10 to get

$$\hat{S}_n(m_n) = -\sum_{i=m_n+1}^{n} Y_{i,n} - g_{\hat{\theta}_n}(Y_{i-1,n})$$

$$
\begin{aligned}
&= \sum_{i=m_n+1}^{n} \mathrm{E}[g_{\tilde{\theta}_0}(Y_{i-1,n})] - \mathrm{E}[Y_{i,n}] + \sum_{i=m_n+1}^{n} \hat{\varepsilon}_{i,n} - \mathrm{E}[\hat{\varepsilon}_{i,n}] \\
&\qquad + \sum_{i=m_n+1}^{n} \mathrm{E}[\tilde{\varepsilon}_{i,n}] - \mathrm{E}[\hat{\varepsilon}_{i,n}] \\
&= -m_n C + o_P(n).
\end{aligned}
$$

Thus, we have $\left|\hat{S}_n(m_n)\right| = m_n C + o_P(n)$ which yields as an upper bound of the CUSUM test statistic

$$
\begin{aligned}
T_n^C &\geq \sqrt{\frac{n}{m_n(n-m_n)}} |\hat{S}_n(m_n)| = \sqrt{\frac{n m_n}{n-m_n}} C + o_P(\sqrt{n}) \\
&= \sqrt{\frac{m_n}{1-m_n/n}} C + o_P(\sqrt{n}).
\end{aligned}
$$

We have by the assumptions of the theorem that $a(\log(n)) = O(\sqrt{\log(\log(n))})$ and $b(\log(n)) = O(\log(\log(n)))$. Assumption (A10) ensures the uniform boundedness of the second moments of the whole process $Y_{i,n}$. Thus, also the limit of $\hat{\sigma}_n$ is bounded from above. Altogether, we get

$$
\begin{aligned}
&a(\log(n)) \frac{T_n^C}{\hat{\sigma}_n} - b(\log(n)) \\
&= O(\sqrt{\log(\log(n))}) + \left(\sqrt{\frac{m_n}{1-m_n/n}} C + o_P(\sqrt{n}) \right) O_P(1) - O(\log(\log(n))) \\
&= O_P(\sqrt{n \log(\log(n))})
\end{aligned}
$$

which implies the assertion of the theorem. □

This concludes the analysis of the CUSUM test statistic under the alternative. Next we are interested in the asymptotics of the gradually weighted test statistic. As the weighting is done differently, we need another approach and another assumption to make sure the change is recognizable. As described regarding the intuition with respect to this test statistic, the numerator has to grow with at least a rate of $O_P(n^{\gamma+1})$. This can be ensured by the following assumption:

(A9*) $\lim\limits_{n\to\infty} \frac{1}{n^{\gamma+1}} \sum\limits_{i=m_n+1}^{n} (i-m_n)^{\gamma} \left(\mathrm{E}[Y_{i,n}] - \mathrm{E}[g_{\tilde{\theta}_n}(Y_{i-1,n})] \right) \geq C > 0.$

Example 9.12. Let $\{Y_{i,n} | i \in \mathbb{N}_0,\ n \in \mathbb{N},\ i \leq n\}$ be a linear triangular Poisson-INARCH(1) process. With $\tilde{\theta}_0$ as in Theorem 8.15, we have with Lemmas 8.12

and 8.13

$$
\begin{aligned}
&\frac{1}{n^{\gamma+1}} \sum_{i=m_n+1}^{n} (i-m_n)^{\gamma} \left(\mathrm{E}[Y_{i,n}] - \mathrm{E}[g_{\tilde{\theta}_n}(Y_{i-1,n})]\right) \\
&= \frac{1}{n^{\gamma+1}} \sum_{i=1}^{n-m_n} i^{\gamma} \left(\mathrm{E}[\tilde{g}_{\theta_0}(Y_{m_n+i-1,n})] - \mathrm{E}[g_{\tilde{\theta}_n}(Y_{m_n+i-1,n})]\right) \\
&= \frac{1}{n^{\gamma+1}} \sum_{i=1}^{n-m_n} i^{\gamma} \left((\alpha_0-\tilde{\alpha}_0)\mathrm{E}[(Y_{m_n+i-1,n})] + (\omega_0-\tilde{\omega}_0) + \delta\left(\frac{i}{n}\right)^{\gamma}\right) \\
&= \frac{1}{n^{\gamma+1}} \sum_{i=1}^{n-m_n} i^{\gamma} \left((\alpha_0-\tilde{\alpha}_0)\left(\mathrm{E}[Y_0] + \frac{\delta}{n^{\gamma}}\sum_{j=1}^{i} j^{\gamma}\alpha^{i-j}\right)\right. \\
&\qquad \left. +(1-\alpha_0)\mathrm{E}[Y_0] - (1-\tilde{\alpha}_0)\left(\mathrm{E}[Y_0] + \frac{\delta(1-\tau)^{\gamma+1}}{(1-\alpha_0)(\gamma+1)}\right) + \delta\left(\frac{i}{n}\right)^{\gamma}\right) \\
&= \frac{(\alpha_0-\tilde{\alpha}_0)\delta}{n^{2\gamma+1}} \sum_{i=1}^{n-m_n} i^{\gamma} \sum_{j=1}^{i} j^{\gamma}\alpha^{i-j} - \frac{(1-\tilde{\alpha}_0)\delta(1-\tau)^{\gamma+1}}{n^{\gamma+1}(1-\alpha_0)(\gamma+1)} \sum_{i=1}^{n-m_n} i^{\gamma} \\
&\qquad + \frac{\delta}{n^{2\gamma+1}} \sum_{i=1}^{n-m_n} i^{2\gamma} \\
&\xrightarrow{n\to\infty} \frac{(\alpha_0-\tilde{\alpha}_0)\delta(1-\tau)^{2\gamma+1}}{(1-\alpha_0)(2\gamma+1)} - \frac{(1-\tilde{\alpha}_0)\delta(1-\tau)^{2\gamma+2}}{(1-\alpha_0)(\gamma+1)^2} + \frac{\delta(1-\tau)^{2\gamma+1}}{2\gamma+1} \\
&= \frac{\delta(1-\tau)^{2\gamma+1}}{1-\alpha_0}\left(\frac{(\alpha_0-\tilde{\alpha}_0)}{2\gamma+1} - \frac{(1-\tilde{\alpha}_0)(1-\tau)}{(\gamma+1)^2} + \frac{(1-\alpha_0)}{2\gamma+1}\right) \\
&= \frac{(1-\tilde{\alpha}_0)\delta(1-\tau)^{2\gamma+1}}{1-\alpha_0}\left(\frac{1}{2\gamma+1} - \frac{1-\tau}{(\gamma+1)^2}\right)
\end{aligned}
$$

and thus, (A9*) is always fulfilled for a linear triangular Poisson-INARCH(1) process.

Having shown that this new assumption is fulfilled in the linear case we go on with the asymptotics of the test statistic. The next lemma will help us to handle the weighted residuals.

Lemma 9.13. *Let $\{Y_{i,n} | i \in \mathbb{N}_0,\ n \in \mathbb{N},\ i \leq n\}$ be a triangular Poisson-INARCH(1) process fulfilling assumptions (A1) to (A11) with (A9*) instead of (A9'). Then, under H_1 the triangular random array $\{X_{i,n} | i, n \in \mathbb{N}, i \leq n - m_n\}$ given by*

$$X_{i,n} = \left(\frac{i}{n}\right)^{\gamma} \varepsilon_{i+m_n,n}$$

fulfills an ULLN with respect to $\theta \in \Theta$.

Proof. From Proposition 8.11, we know that the residuals form an L_1-mixingale arrays fulfilling a ULLN. As a continuous transformation of the $\varepsilon_{i,n}$, the new triangular array retains the exponentially decaying mixing rate. Since all the factors are smaller than 1, the uniform upper bound of the $(3+\nu)$-th moments for some $\nu > 0$ is also retained. We get the SLLN with the same arguments as in that Proposition. Let $\theta, \theta' \in \Theta$ with $\varepsilon_{i,n}$ and $\varepsilon'_{i,n}$ be the residuals with respect to θ and θ' respectively. Then, since the arithmetic mean of the residuals itself is AUEC,

$$\left| \frac{1}{n^{\gamma+1}} \sum_{i=1}^{n-m_n} i^{\gamma} \left(\varepsilon_{i+m_n,n} - \varepsilon'_{i+m_n,n} \right) \right| \leq \frac{1}{n^{\gamma+1}} \sum_{i=1}^{n-m_n} i^{\gamma} \left| \varepsilon_{i+m_n,n} - \varepsilon'_{i+m_n,n} \right|$$
$$\leq \frac{1}{n} \sum_{i=1}^{n-m_n} \left| \varepsilon_{i+m_n,n} - \varepsilon'_{i+m_n,n} \right| \leq c \left\| \theta - \theta' \right\|$$

for some $c > 0$, which means that the $X_{i,n}$ are AUEC. With the same approach regarding the $\mathrm{E}[X_{i,n}]$, they are also AUEC and the ULLN holds with respect to $\theta \in \Theta$ by Theorem 5.16. □

With this lemma we are able to prove the last theorem of this thesis. As we will see, with assumption (A9*), the gradually weighted test statistic does in fact converge in probability to infinity.

Theorem 9.14. *Let $\{Y_{i,n} | i \in \mathbb{N}_0,\ n \in \mathbb{N},\ i \leq n\}$ be a triangular Poisson-INARCH(1) process fulfilling assumptions (A1) to (A11) with (A9*) instead of (A9'). Then, under H_1*

$$P\left(a_n \frac{T_n^W}{\hat{\sigma}_n} - b_n(\gamma) \geq c \right) \overset{n\to\infty}{\longrightarrow} 1 \tag{9.18}$$

for all $c > 0$ and with a_n and $b_n(\gamma)$ as in Theorem 9.5.

Proof. Similar as in (9.17) in the proof of Theorem 9.11, we have with (A9*) and the ULLN from Lemma 9.13

$$\frac{1}{n^{\gamma+1}} \sum_{i=1}^{n} (i - m_n)_+^{\gamma} \left(Y_{i,n} - \mathrm{E}[g_{\hat{\theta}_n}(Y_{i-1,n})] \right)$$
$$= \sum_{i=1}^{n} (i - m_n)_+^{\gamma} \left(\mathrm{E}[Y_{i,n}] - \mathrm{E}[g_{\tilde{\theta}_0}(Y_{i-1,n})] + \hat{\varepsilon}_{i,n} - \mathrm{E}[\hat{\varepsilon}_{i,n}] \right)$$
$$\geq c + o_P(1).$$

Putting this in the test statistic yields,

$$T_n^W \geq \frac{\sum_{i=1}^{n} (i - m_n)_+^{\gamma} \left(Y_{i,n} - \mathrm{E}[g_{\hat{\theta}_n}(Y_{i-1,n})] \right)}{(\sum_{i=1}^{n} (i-k)_+^{2\gamma})^{1/2}} \geq \frac{c n^{\gamma+1} + o_P(n^{\gamma+1})}{(\sum_{i=1}^{n} (i - m_n)_+^{2\gamma})^{1/2}}$$
$$\geq \frac{\sqrt{2\gamma+1} c n^{\gamma+1} + o_P(n^{\gamma+1})}{(n - m_n)^{\gamma+1/2}} = O_P(\sqrt{n}).$$

In this case, both a_n is of order $O(\sqrt{\log(\log(n))})$ and b_n is of order $O(\log(\log(n)))$, thus

$$a_n \frac{T_n^W}{\hat{\sigma}_n} - b_n = O(\sqrt{\log(\log(n))}) O_P(\sqrt{n}) - O(\log(\log(n)))$$
$$= O_P(\sqrt{n \log(\log(n))})$$

which proves the theorem. □

We now have two test statistics which not only behave as desired under the null hypothesis. Both can, under their own assumption regarding the change, yield a consistent test under the alternative. Moreover, we have shown the linear triangular Poisson-INARCH(1) process fulfills both of these assumptions. We can turn to the final numerical analysis and evaluate the performance of both test statistics, as well as comparing both of them with each other.

9.4 Numerical analysis of the CUSUM test statistic

We start the numerical analysis by examining how the weighted cumulative sums behave under this gradual change point model for a linear triangular Poisson-INARCH(1) process. Therefore, we simulate two examples, one with $\gamma = 2$, see Figure 9.1, and one with $\gamma = 0.5$, see Figure 9.2. When calculating the test statistic, we take the maximum over the weighted cumulative sums at $k = 1, ..., n$. Those are displayed in Figure 9.3 for $\gamma = 2$ and in Figure 9.4 for $\gamma = 0.5$ respectively. In the case of $\gamma = 2$, the cumulative sums behave similar as in the abrupt case, but more as if the change point would occur just after 150. The triangular shape is still present, but shifted to the right in comparison to the actual position of the change point. This gives a hint, that the relative position of the maximum is not anymore an estimator of the actual position of

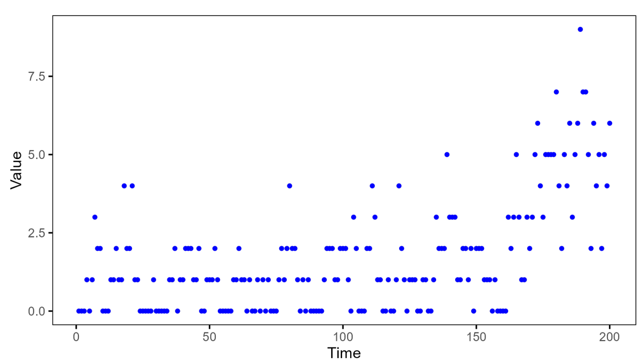

Figure 9.1: Example of a gradual change point with $\alpha = 0.5$, $\omega = 0.5$, $\delta = 8$, $\gamma = 2$ and $\tau = 0.5$.

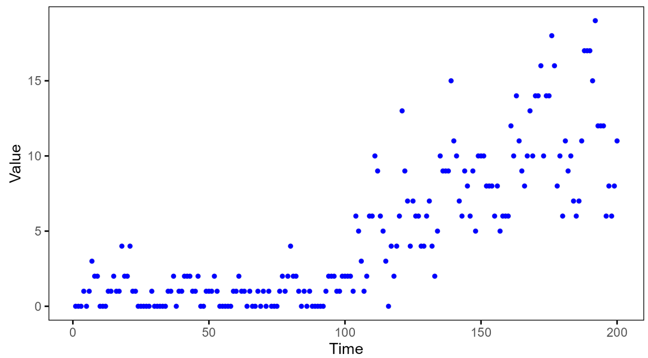

Figure 9.2: Example of a gradual change point with $\alpha = 0.5$, $\omega = 0.5$, $\delta = 8$, $\gamma = 0.5$ and $\tau = 0.5$.

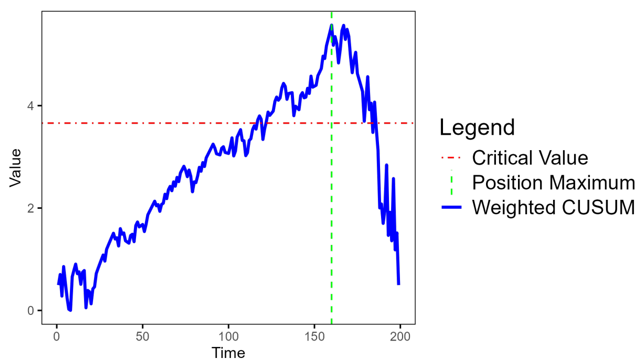

Figure 9.3: Example of the CUSUM test statistic for a gradual change point with $\alpha = 0.5$, $\omega = 0.5$, $\delta = 8$, $\gamma = 2$ and $\tau = 0.5$.

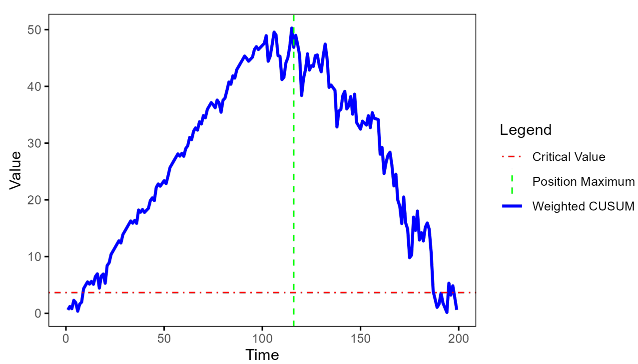

Figure 9.4: Example of the CUSUM test statistic for a gradual change point with $\alpha = 0.5$, $\omega = 0.5$, $\delta = 8$, $\gamma = 0.5$ and $\tau = 0.5$.

Number of observations	100	200	500	1000
H_0	2.41%	2.99%	3.93%	4.67%
H_1, $\tau = 0.5$ $\delta = 8$	15.34%	83.34%	100%	100%
H_1, $\tau = 0.75$ $\delta = 32$	24%	75.51%	99.91%	100%

Table 9.1: Power of the CUSUM test statistic for $\omega_0 = 0.5$, $\alpha_0 = 0.5$, $\gamma = 2$, a significance level of 5%, 10000 repetitions.

Number of observations	100	200	500	1000
H_1, $\tau = 0.5$ $\delta = 8$	84.28%	80.76%	80.28%	80.02%
H_1, $\tau = 0.75$ $\delta = 32$	89.54%	89.06%	89.73%	89.72%

Table 9.2: Relative position of the maximum T_n^C for $\omega_0 = 0.5$, $\alpha_0 = 0.5$, $\gamma = 2$, a significance level of 5%, 10000 repetitions.

the change point. Regarding the critical value, in this case the test unambiguously rejects the null hypothesis. In the case of $\gamma = 0.5$, i.e. Figure 9.3, the weighted cumulative sums do even more resemble those for an abrupt change. The position of the maximum of the weighted cumulative sums is even really close to the actual position of the change point. It is notable, that the weighted cumulative sums are scattered more widely directly after the change point than before it and towards the last values. This effect could arise from the characteristics of the change point in this case. For the square root function, the highest growth rate is right at the beginning and flattens thereafter. So the highest difference between consecutive values is directly after the change point. Again, the test clearly rejects the null hypothesis. These examples indicate, that for this gradual change point model, the CUSUM test statistic behaves similar as in the examples for abrupt changes. The weighted cumulative sums increase almost monotonically until some time point, and revert thereafter, returning back to 0. Now, the relative position of the maximum of the weighted cumulative sums is not an estimator of the relative position of the change point τ.

The next part of our analysis is calculating the rejection rates for some selected parameter sets, beginning with the results in Table 9.1. Note, that the values under the null distribution are only displayed for completeness, as the convergence under the null is already given from Franke et al. [2012]. The rejection rates under the alternative show a promising result. Both, for a change point at $\tau = 0.5$ and $\tau = 0.75$, the rejection rates are for $n = 200$ observations at 83.34% rejections and 75.51% rejections, respectively. At $n = 500$, both rejection rates already reach 100%. Here, we used different vales of the scale parameters δ, to make them more comparable. In Table 9.2, the relative posi-

Number of observations	100	200	500	1000
H_0	0.92%	1.44%	1.96%	2.22%
H_1, $\tau = 0.5$ $\delta = 8$	9.37%	99.23%	100%	100%
H_1, $\tau = 0.75$ $\delta = 32$	1.56%	3.08%	99.96%	100%

Table 9.3: Power of the CUSUM test statistic for $\omega_0 = 1$, $\alpha_0 = 0.3$, $\gamma = 0.5$, a significance level of 5%, 10000 repetitions.

Number of observations	100	200	500	1000
H_1, $\tau = 0.5$ $\delta = 8$	57.9%	58.29%	58.08%	57.92%
H_1, $\tau = 0.75$ $\delta = 32$	96.93%	93.71%	77.55%	78.12%

Table 9.4: Relative position of the maximum T_n^C for $\omega_0 = 1$, $\alpha_0 = 0.3$, $\gamma = 0.5$, a significance level of 5%, 10000 repetitions.

tion of the maximum of the weighted cumulative sums are displayed. Clearly, they are not estimators for τ. They seemingly converge to some value behind the actual relative position of the change point. From how we defined the gradual change point, the gradual increase does not visibly accelerate until a later time point. So this maximum is more of an indicator for when the gradual growth starts to accelerate. Table 9.3 displays rejection rates for a different parametrization, in particular the shape parameter γ is now 0.5. The overall picture is similar as in the previous case. So seemingly, the test can handle different shapes of the gradual increase rate. Regarding the relative position of the maximum of the cumulative sums, see Table 9.4. The values are here closer to the actual value of τ. Still, the values do not seem to converge to the true value.

We go on with a more thorough analysis of the quality of the test under the alternative. First we inspect the consequences of increasing the scale parameter δ. For $\alpha = 0.5$, $\omega = 0.5$, $\gamma = 2$, $\tau = 0.5$ we repeated the test 1000 times for values of δ between 0 and 6. The resulting Figure 9.5 is an S-shaped curve, similar to results for the other examples of change point models before. If δ is small, the overall increases is also small, and it is hard to discern between the null and the alternative. But as δ increases, the test is better at rejecting the null hypothesis.

Regarding a changing value of the shape parameter γ, we approach the analysis slightly different. From Lemma 8.14(i), we know that the limit of the expectation of the sample mean depends on the shape parameter γ. The limit decreases if γ increases. Moreover, an increasing value of γ leads to a later start of the acceleration of the gradual increases, as noted before. Therefore, the time series

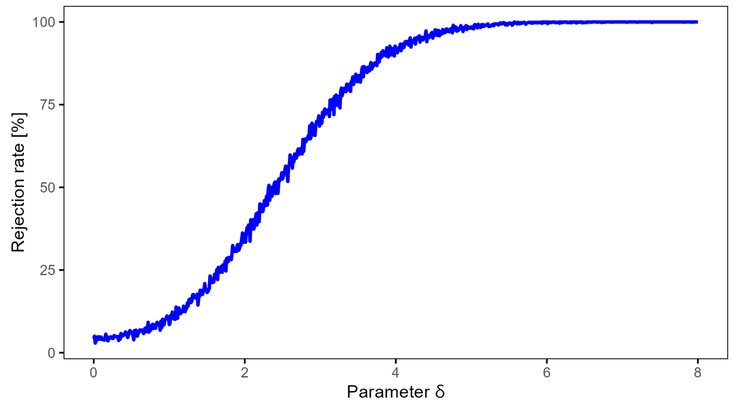

Figure 9.5: Rejection rates of the CUSUM test statistic for $\delta \in [0, 8]$ with $\alpha = 0.5$, $\omega = 0.25$, $\gamma = 2$, $\tau = 0.5$, $n = 500$, a significance level of 5% and 1000 repetitions.

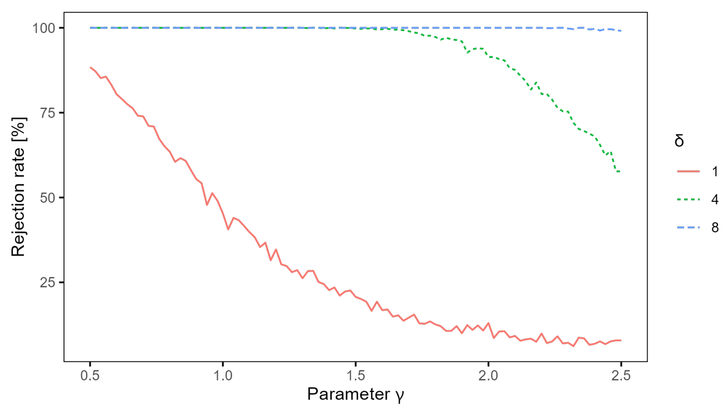

Figure 9.6: Rejection rates of the CUSUM test statistic for $\gamma \in [0.5, 2.5]$ with $\alpha = 0.5$, $\omega = 0.5$, $\tau = 0.5$, $n = 500$, a significance level of 5% and 1000 repetitions.

after the change point stays longer on roughly the same distribution as before the change point. Consequently, the increase in γ has an behaves similarly to an increase in τ, regarding only the start of the gradual increase. Thus, we calculate the rejection rates for different scale parameters δ, see Figure 9.6. For a given δ, the rejection rates decrease for increasing shape parameter γ. This reinforces the statements before, on how an increasing γ influences the gradual change, i.e., a reduction of the expectation of the sample mean. If the time series under the alternative deviates enough from the stationary case, the shape of the gradual change has not a big effect on the test.

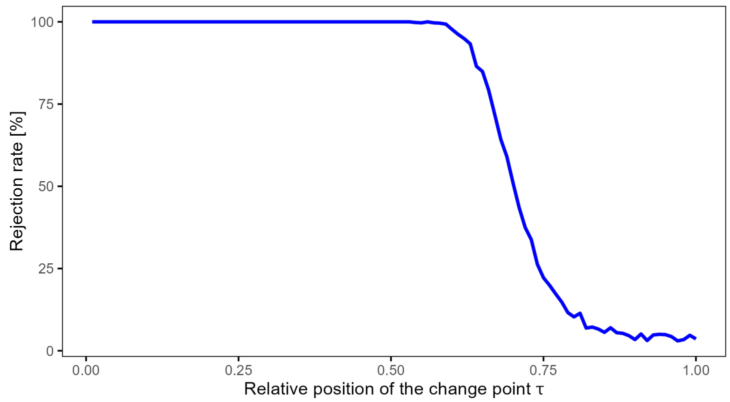

Figure 9.7: Rejection rates of the CUSUM test statistic for $\tau \in [0,1]$ with $\alpha = 0.5$, $\omega = 1$, $\gamma = 2$, $\delta = 4$, $n = 500$, a significance level of 5% and 1000 repetitions.

Lastly, we are interested in the results of moving the relative position of the change point between 0 and 1, displayed in Figure 9.7. We note, that the curve behaves very different to the same analyses for the prior examples of abrupt changes. Already for the earliest change points, the test rejects the null hypothesis reliably. This makes sense, since an early change point results in a time series which is only stationary for a small part. And we test against a null hypothesis of a completely stationary time series. If τ increases towards 1, we note that the decrease already starts shortly after 0.5. As discussed before, the steepest part of the gradual increase starts only after some time, at least for $\gamma > 1$. So the part with high deviations from the stationary distribution starts some time after the change point. Thus, there are less observations which help discern the null hypothesis from the alternative as for comparable values of τ in an abrupt change point model. So the early start of the decrease also makes sense.

In summary, the CUSUM test statistic for a linear Poisson-INARCH(1) process with a gradual change is able to reliably reject the null hypothesis. As in the previous examples, it is of course dependent on the deviation from a stationary distribution. The more it deviates from stationarity, the better the test works.

9.5 Numerical analysis of the gradually weighted test statistic

Now, we assess the quality of the gradually weighted test statistic. The analysis will mostly be the same as for the CUSUM test statistic. We still base the

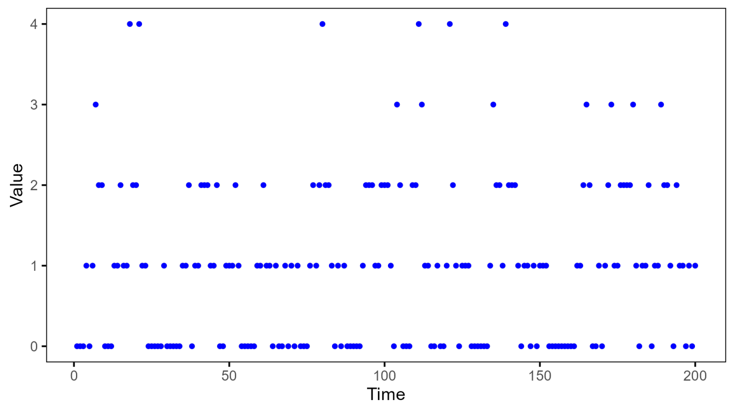

Figure 9.8: Example of a linear Poisson-INARCH(1) process with $\alpha = 0.5$ and $\omega = 0.5$.

analysis on a linear triangular Poisson-INARCH(1) process. In this case, we will put more emphasis on the analysis under the null hypothesis. In difference to the CUSUM case, we have proven the asymptotics under the null hypothesis in this thesis. Moreover, an implied assumption of the test is that we know the real value of γ. But in practice, this is not reliably the case. So, we will also test the quality of the test statistic when γ is misspecified, i.e., we intentionally use another γ in the test statistic, as we have used for simulating the time series.

We start with an example of a stationary linear Poisson-INARCH(1) process, see Figure 9.8. We first calculate the gradually weighted test statistics with a shape parameter $\gamma = 2$, which are displayed in Figure 9.9. The gradually weighted cumulative sums produce a smooth line which is close to constant for the most part and clearly accepts the null hypothesis. Since in the stationary case, there is no correct or wrong value of γ, we do the same for a shape parameter of $\gamma = 0.5$. The result can be observed in Figure 9.10. The gradually weighted cumulative sums diverge more, nonetheless the test accepts the null hypothesis with a considerably large distance between the test statistic and the critical value. Regarding the alternative, we use the same examples as for the CUSUM test statistic, i.e., the observations of Figures 9.1 and 9.2. The gradually weighted test statistic with correctly specified shape parameter $\gamma = 2$ is displayed in Figure 9.11. The gradually weighted cumulative sums behave like a smoothed version of those in the CUSUM case. They increase until a maximum after the actual position of the change point is reached and revert thereafter towards 0. The null hypothesis is rejected, but not with the same margin as for the CUSUM based test. The gradually weighted cumulative sums are also calculated for the case of $\gamma = 0.5$. The behavior is similar as in the quadratic case, with a rougher path and a spike at the end. The null hypothesis is again rejected, only with a smaller distance between the critical value and the test

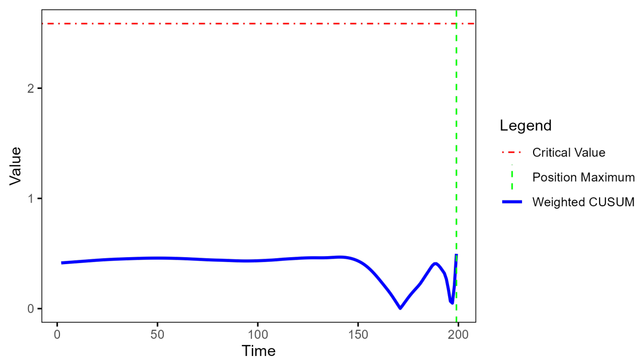

Figure 9.9: Example of the gradually weighted test statistic with $\alpha = 0.5$, $\omega = 0.5$ and $\gamma = 2$.

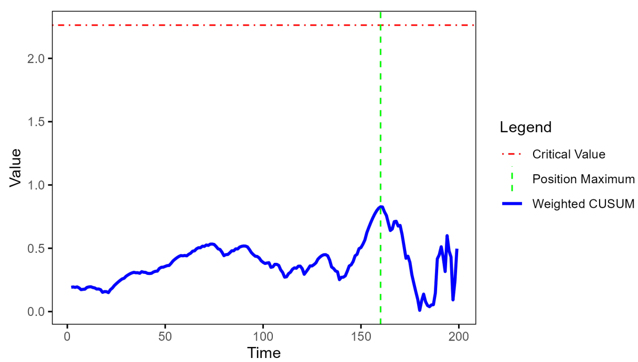

Figure 9.10: Example of the gradually weighted test statistic with $\alpha = 0.5$, $\omega = 0.5$ and $\gamma = 0.5$.

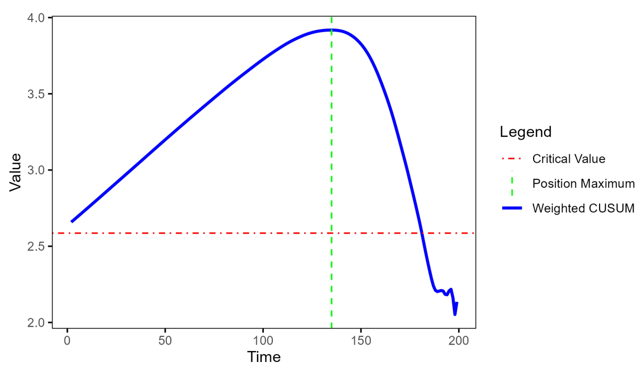

Figure 9.11: Example of the gradually weighted test statistic for a gradual change point with $\alpha = 0.5$, $\omega = 0.5$, $\delta = 8$, $\gamma = 2$ and $\tau = 0.5$.

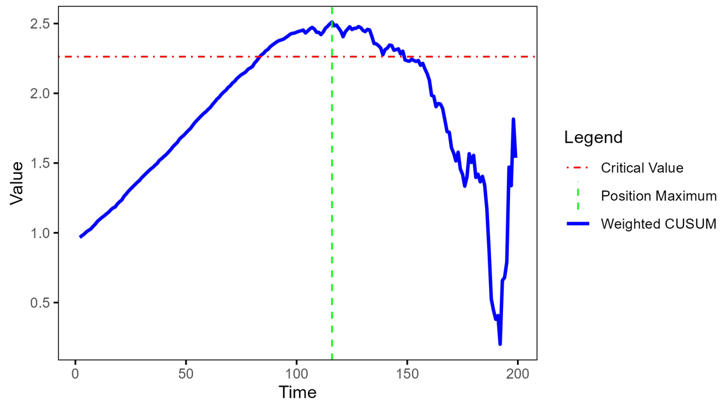

Figure 9.12: Example of the gradually weighted test statistic for a gradual change point with $\alpha = 0.5$, $\omega = 0.5$, $\delta = 8$, $\gamma = 0.5$ and $\tau = 0.5$.

Number of observations	100	200	500	1000
H_0	4.3%	4.39%	4.86%	4.84%
H_1, $\tau = 0.5$ $\delta = 8$	54.32%	94.56%	100%	100%
H_1, $\tau = 0.75$ $\delta = 32$	64.29%	92.06%	100%	100%

Table 9.5: Power of the gradually weighted test statistic for $\omega_0 = 0.5$, $\alpha_0 = 0.5$, $\gamma = 2$, a significance level of 5%, 10000 repetitions.

Number of observations	100	200	500	1000
H_1, $\tau = 0.5$ $\delta = 8$	76.95%	66.57%	63.74%	63.28%
H_1, $\tau = 0.75$ $\delta = 32$	83.66%	80.12%	78.92%	78.61%

Table 9.6: Relative position of the maximum of T_n^W for $\omega_0 = 0.5$, $\alpha_0 = 0.5$, $\gamma = 2$, a significance level of 5%, 10000 repetitions.

gradually weighted statistic.

Next, we examine the quality of the test for different parameter sets and under the different hypotheses for 10000 repetitions. Table 9.5 shows the rejection rates for $\omega_0 = 0.5$, $\alpha_0 = 0.5$ and $\gamma = 2$, where the shape parameter is correctly specified in the test statistic. Under the null hypothesis, the rejection rates are close to 5%. So, the test rejects only slightly less of the observed time series than desired. The rejection rates under the alternative converge fast to 100% and are higher for $n = 100$ and $n = 200$ compared to the same results

Number of observations	100	200	500	1000
H_0	9.49%	8.91%	8.94%	8.91%
H_1, $\tau = 0.5$ $\delta = 8$	78.92%	100%	100%	100%
H_1, $\tau = 0.75$ $\delta = 32$	42.7%	80.89%	100%	100%

Table 9.7: Power of the gradually weighted test statistic for $\omega_0 = 1$, $\alpha_0 = 0.3$, $\gamma = 0.5$, a significance level of 5%, 10000 repetitions.

Number of observations	100	200	500	1000
H_1, $\tau = 0.5$ $\delta = 8$	69.16%	61.11%	59.21%	59.09%
H_1, $\tau = 0.75$ $\delta = 32$	96.14%	84.89%	77.58%	75.96%

Table 9.8: Relative position of the maximum of T_n^W for $\omega_0 = 1$, $\alpha_0 = 0.3$, $\gamma = 0.5$, a significance level of 5%, 10000 repetitions.

Number of observations	100	200	500	1000
H_0, $\gamma = 2$, $\alpha = 0.3$, $\omega = 0.3$	5.17%	4.74%	5.1%	5.05%
H_0, $\gamma = 2$, $\alpha = 0.5$, $\omega = 1$	3.34%	3.9%	3.99%	4.38%
H_0, $\gamma = 0.5$, $\alpha = 0.5$, $\omega = 0.5$	9.51%	9.45%	9.39%	9.58%
H_0, $\gamma = 0.5$, $\alpha = 0.5$, $\omega = 1$	9.28%	9.23%	8.73%	8.76%

Table 9.9: Rejection rates of the gradually weighted test statistic for a significance level of 5% and 10000 repetitions.

for the CUSUM test statistic. We also tested, what the relative position of the maximum of the gradually weighted cumulative sums were. The results are displayed in Table 9.6. For this test statistic, we do not get an estimator of τ with the relative position of this maximum. The results for a shape parameter $\gamma = 0.5$ are similar under the alternative, see Tables 9.7 and 9.8. Under the null hypothesis on the other hand, the rejection rate is way above the significance level of 5%, only slowly decreasing for increasing n. Hence, we further investigate the behavior under the null. In Table 9.9, we see the rejection rates under the null for different parameter configuration of α, ω and γ. Again, for $\gamma = 2$, the rejection rates are close to the aimed significance level. For $\gamma = 0.5$, more than the desired 5% are rejected. The shape parameter influences the quality of the test under the null. Therefore, for the parameters $\alpha = 0.5$ and $\omega = 0.5$, we calculated the rejection rate for 1000 repetitions and varying values of γ. The

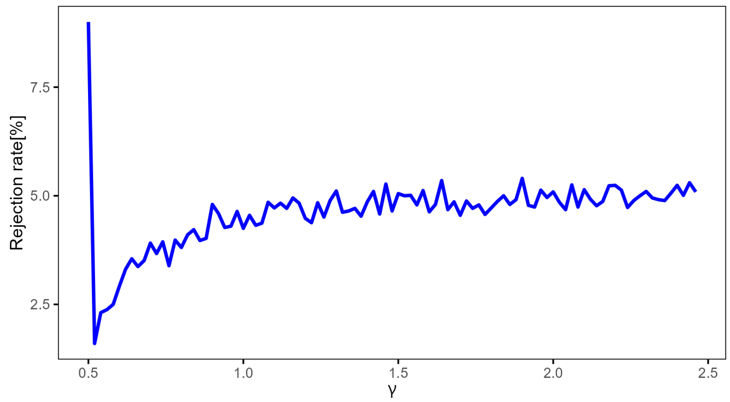

Figure 9.13: Rejection rates under the null hypothesis for the gradually weighted test statistic for $\gamma \in [0.5, 2.5]$ with $\alpha = 0.5$, $\omega = 0.5$, $n = 500$, a significance level of 5% and 10000 repetitions.

Number of observations	100	200	500	1000
H_0, $\gamma = 0.6$, $\alpha = 0.5$, $\omega = 0.5$	2.37%	2.57%	3.2%	2.94%
H_0, $\gamma = 0.7$, $\alpha = 0.5$, $\omega = 0.5$	3%	3.26%	3.73%	3.63%
H_0, $\gamma = 0.8$, $\alpha = 0.5$, $\omega = 0.5$	3.34%	3.65%	3.94%	4.05%
H_0, $\gamma = 0.8$, $\alpha = 0.5$, $\omega = 0.3$	4.3%	3.88%	4.94%	4.69%

Table 9.10: Rejection rates of the gradually weighted test statistic for a significance level of 5% and 10000 repetitions.

result is displayed in Figure 9.13. For $\gamma \geq 1$, the results vary around the goal of 5%. For $\gamma = 0.5$, we see the same result as before, whereas for $0.5 < \gamma < 1$ the test is conservative with values below 5%. Under the null hypothesis, the shape parameter influences how fast the rejection rates converge to the significance level. A possible explanation for this behavior for $\gamma = 0.5$ could be the four times iterated logarithm in b_n. Asymptotically, the this logarithm will converge to positive infinity. For $n < e^{e^{e}}$ this logarithm is negative and we have $b_n < 2\log(\log(n))$, while it asymptotically holds that $b_n > 2\log(\log(n))$. For our relatively small values of n, the negative for times iterated logarithm reduces the critical value and the null hypothesis gets rejected more often. If this is true, the slow growth of this four times iterated logarithm could explain, why for our sample lengths the rejection rate seemingly stays constant. Moreover,

Number of observations	100	200	500	1000
H_1, $\gamma_{real} = 2$, $\gamma_{miss} = 0.8$	13.78%	54%	100%	100%
H_1, $\gamma_{real} = 2$, $\gamma_{miss} = 3$	54.89%	94.19%	100%	100%
H_1, $\gamma_{real} = 0.5$, $\gamma_{miss} = 0.8$	10.07%	14.82%	100%	100%
H_1, $\gamma_{real} = 0.5$, $\gamma_{miss} = 2$	13.93%	32.08%	100%	100%

Table 9.11: Power of the gradually weighted test statistic for misspecified γ and $\omega_0 = 0.5$, $\alpha_0 = 0.5$, $\delta = 8$, $\tau = 0.5$, a significance level of 5% and 10000 repetitions.

we note that for $\gamma \searrow 0.5$, the latter part of

$$b_n = 2\log(\log(n)) + \log\left(\frac{1}{4\pi}\left(\frac{2\gamma+1}{2\gamma-1}\right)^{1/2}\right)$$

grows to infinity. The two times iterated logarithm needs to grow large enough such that the ratio between both logarithms behaves like it does asymptotically. To test this we calculated the rejection rates for different values of γ between 0.5 and 1. The results are displayed in Table 9.10. Apart from single outliers, they mostly make the impression to increase to 5% for increasing sample lengths. At least smaller sample lengths, this suggests to overestimate the shape parameter γ if it is unknown. Then, the possibility to either over- or undershoot the desired significance level under H_0 are minimized.

As discussed in the beginning of the subsection, we assume to know the real value of γ in the test statistic. As this is not always fulfilled in practice, we analyze the effect of a misspecified γ. In Table 9.11, the results of such an experiment are displayed for $\omega = 0.5$, $\alpha = 0.5$, $\delta = 8$ and $\tau = 0.5$. The correct shape parameters used for simulating the time series are given as γ_{real}. The misspecified values of the shape parameter are denoted by γ_{miss}. For a real shape parameter $\gamma_{real} = 2$, the results are nearly the same as for the correct value if $\gamma_{miss} = 3$, see Table 9.5. In the case of $\gamma_{miss} = 0.8$, the rejection rates are for $n = 100$ and $n = 200$ around 40% lower as in the correctly specified case. For $\gamma_{real} = 0.5$, the rejection rates are for $n = 100$ and $n = 200$ much lower than in Table 9.7. Again, a higher value of the misspecified γ_{miss} leads to higher rejection rates. In applications it makes a difference if the expected gradual change is more in the shape of a square root or a parabola. To further investigate this effect, we did an experiment, where the misspecified γ takes on values between 0.5 and 3, see Figure 9.14. We chose an example where the rejection rates are for the most part below 100%, to be able to recognize, how the rejection rates change. Besides a peak at 0.5, the rejection rates increase for

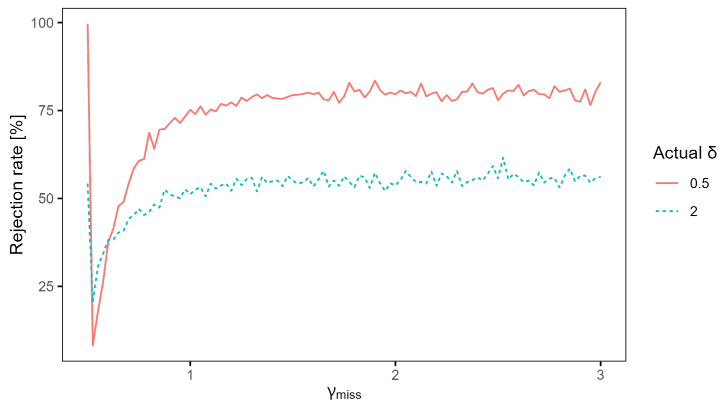

Figure 9.14: Rejection rates of the gradually weighted test statistic for $\gamma_{miss} \in [0.5, 3]$ with $\alpha = 0.5$, $\omega = 0.5$, $\delta = 4$, $n = 200$, a significance level of 5% and 1000 repetitions.

increasing γ. If the actual shape parameter is unknown, choosing $\gamma = 0.5$ yields the best rejection rates under the alternative. But this is accompanied with the cost of an higher rejection rate under the null hypotheses, as we have seen above. For the range we have tested, this again suggests to prefer larger values of the shape parameter, if it is unknown. The power under the alternative is still sufficient even if misspecified, while the rejection rates under the null are close to the significance level.

Figure 9.15 shows the rejection rates under the alternative, while varying the scale parameter δ between 0 and 6. We have included the rejection rates for the CUSUM test, to make them easier to compare. For the gradually weighted test statistic we also have an S-shaped curve for increasing δ. It is similar to the behavior of the CUSUM test, while always being a few percent above the rejection rate of the CUSUM test. The gradually weighted test statistic is a bit more sensitive regarding the correct rejection of the null hypothesis for small scale parameters.

The results for varying the shape parameter γ are displayed in Figure 9.16. This is the same experiment as in Figure 9.6, regarding the CUSUM test. We show only the results for the gradually weighted test statistic for a better readability of the figure. Besides the downward spike in at the start of the curve for $\delta = 1$, the resulting figure is similar to the CUSUM case. Again, the rejection rates are a bit higher regarding the gradually weighted test statistic. The spike at $\gamma = 0.5$ and the sudden drop thereafter reflects what we have seen under the null hypothesis in Figure 9.13. For $\gamma = 0.5$, the test tends to reject more cases as desired under the null. Hence, the test statistic is larger than expected from the Gumbel distribution, exceeding the critical value. For $0.5 < \gamma < 1$, this is

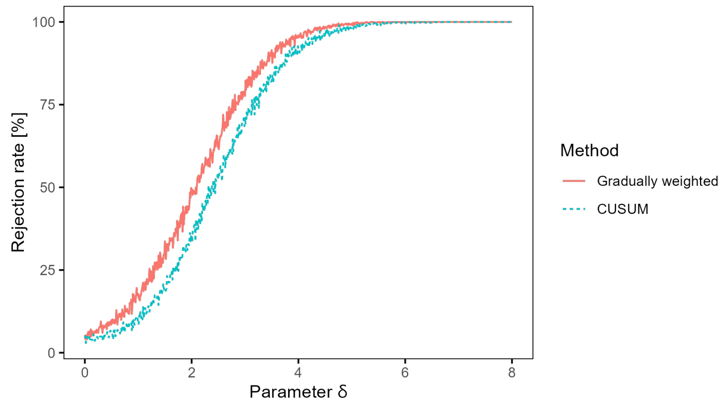

Figure 9.15: Rejection rates of the gradually weighted test statistic for $\delta \in [0, 8]$ $\alpha = 0.5$, $\omega = 0.5$, $\gamma = 2$, $\tau = 0.5$, $n = 500$, a significance level of 5% and 1000 repetitions.

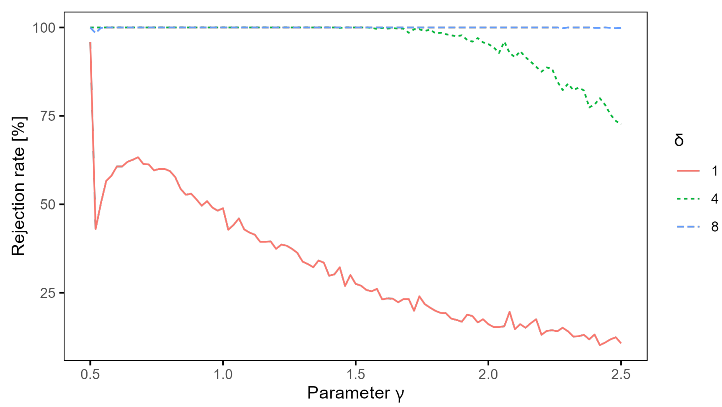

Figure 9.16: Rejection rates of the gradually weighted test statistic for $\gamma \in [0.5, 2.5]$ with $\alpha = 0.5$, $\omega = 0.5$, $\tau = 0.5$, $n = 500$, a significance level of 5% and 1000 repetitions.

reverted. The test keeps this property under the alternative. For $\gamma = 0.5$ this is positive, since the null hypothesis is correctly rejected more frequently. But for $0.5 < \gamma < 1$, this leads to less rejections as the test is more conservative.

We do a final experiment for different relative positions of the change point τ between 0 and 1. In Figure 9.17, we show the results in comparison to the same experiment for the CUSUM case. The pattern of the two previous experiments repeats itself. The overall curves are similar, while the rejection rate of the

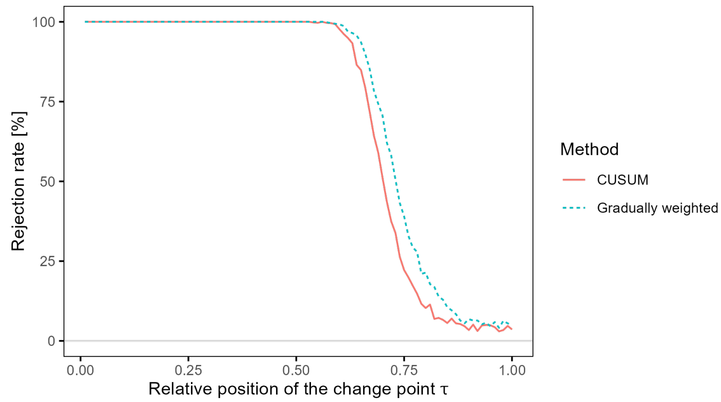

Figure 9.17: Rejection rates of the gradually weighted test statistic for $\tau \in [0, 1]$ and $\alpha = 0.5$, $\omega = 0.5$, $\gamma = 2$, $\delta = 4$, $n = 500$, a significance level of 5% and 1000 repetitions.

gradually weighted test statistic are higher or equal as for the CUSUM test.

The test based on gradually weighted test statistic demonstrates its consistency in simulation. The rejection rates under the null hypothesis are close to the significance level, at least for $\gamma \geq 1$. Under the alternative, the test is slightly better than the CUSUM based test, at least for correctly specified γ. Depending on the actual shape of the time series, the quality under the alternative is reduced, if the shape parameter is misspecified.

9.6 Discussion

We have introduced two change point tests for gradual changes in the Poisson-INARCH(1) process. Under the null hypothesis we have the same setting as in the abrupt case and thus, the CUSUM test statistic fulfills the same properties. This means, we have an asymptotic level α-test, where α is the significance level. Regarding the gradually weighted test statistic, we have proven, that it has asymptotic level α. For a new assumption for each of the test statistics, regarding the expectation of the test under the alternative, we get for both cases asymptotic power 1 under the alternative. Both of these new assumptions are fulfilled by a linear triangular Poisson-INARCH(1) process. In a numerical analysis, the CUSUM based test with a linear triangular Poisson-INARCH(1) process yields good results under the alternative. Compared to the abrupt case, we lose the property, that the relative position of the maximum of the weighted cumulative sums is an estimator for the relative position of the change point. At best, we get an indicator for the time point, where the gradual increase accelerates. The gradually weighted test statistic also proved reliable in the simulation study. Compared to the CUSUM test statistic, it is even slightly better in correctly rejecting the null hypothesis. On the other hand, the CUSUM test statistic has the advantage of being independent of the prior knowledge of the shape parameter.

10 Conclusion and Outlook

In this work, we introduced the concept of change point testing and the CUSUM test statistic, based on an example for i.i.d. normally distributed random variables. Thereafter, we established a model for count time series with dependent observations. We familiarized us with the concept of strong mixing time series, which provided us with useful properties, in particular regarding asymptotic results of mixing time series. Section 4 and 5 contained theoretical results from Markov chain theory and about the convergence of estimators for stationary and non-stationary time series. The review of existing literature ended with a change point test for abrupt changes for the Poisson-INARCH(1), as this test is the foundation for the gradual change point test introduced in this work. We proposed two candidates for gradual change point models for the Poisson-INARCH(1) process. The first approach was based on including the transition between two stationary distributions instead of jumping directly to the new distribution. We came to the conclusion, that fixed regression parameters smaller than 1 do not yield the desired kind of transition. Therefore, we introduced the logistic Poisson-INARCH(1) process to overcome this problem, through a growth rate, which is larger than 1 for small values. Only after a certain threshold the growth rate is smaller than 1, converging to 0 even. We have seen, that this process is still stationary and strongly mixing, using the results from Markov chain theory. But since it is a complicated intensity function, we were not able to prove the assumptions of the abrupt change point test, disregarding the transition between stationary distributions at first. Hence, we proposed a second approach, which is based on the addition of a time dependent and deterministic function on the intensity function. For this new model, we proved asymptotic results of the CLS estimates, given three new assumptions. We could accomplish this with a triangular representation of the time series with a gradual change and the CLS function. We showed, that an appropriate SLLN for triangular arrays can be applied, which we used with an equicontinuity condition to get an ULLN. This ULLN together with a suitable assumption on the identifiability of the CLS estimates was used to obtain the convergence of the CLS estimates. We have seen that the new assumptions are fulfilled by a model based on a linear Poisson-INARCH(1) process. We calculated the explicit limits of the CLS estimates in that example and reinforced the validity of these results in a simulation study. In Section 9, we introduced a change point test for gradual changes for the Poisson-INARCH(1) process. Since the null hypothesis stays unchanged, compared to the abrupt change point test, we have an asymptotic level α test regarding the CUSUM test statistic. In addition we introduced a new test statistic besides the CUSUM test statistic. It is based on the form of the gradual increase in this model. We generalized the asymptotic results of a similar test statistic for independent observations to our case of strongly mixing observations. From this we could deduce, that under the null hypothesis, this test statistic yields also an asymptotic level α test. Using the results for the CLS estimates from the prior section, we could prove that both tests have asymptotic power 1 under the alternative. Both test statistics proved reliable in a simulation study. Therefore, we have achieved our goal to extend the theory of change point testing for gradual change points in count time series, by proposing a new model for gradual change points. Moreover, we have two test statistics with different strengths, both yielding consistent tests.

Based on these results, further research questions arise. As of now, we do not have an estimator of the relative position of the change point, which is usually interesting in the application of change point tests. Also, one could analyze the asymptotic behavior of the relative position of the maximum of the weighted cumulative sums of the CUSUM test statistic. As we have seen, this is probably not an estimator of the relative position of the change point. On the other hand, it seems to indicate the time point, where the gradual increase starts to accelerate. This could be even more important in practice, since in this model, the first observations after the change point do not deviate strongly from the stationary distribution before the change point. Regarding count time series, the question arises if it is possible to translate the results from this work to a Poisson-INGARCH(1,1) process. More generally, the theory developed in Section 8, in particular about the convergence of the CLS estimates could be applied to other time series models with gradual changes. The strategy of adding a time-dependent function to shape the gradual change on a process and treating it as a triangular random array is not restricted to the Poisson-INARCH(1) process. The theory which we applied on this triangular random array is also not restricted to the Poisson-INARCH(1) process or count time series and could be used for other types of strongly mixing processes. Lastly, there is the question if the first approach to gradual change models can be used in change point testing. On the one hand, this could be pursued with logistic Poisson-INARCH(1). First by proving the assumptions of the abrupt change point test and using this as a base to include the transitory part between stationary distributions. On the other hand, one could approach this by introducing an intensity function with similar properties regarding the growth rate, which is not as difficult to handle as the logistic function.

Appendices

A Important distributions

Definition A.1. A random variable X is called Poisson distributed $Poi(\lambda)$ with intensity $\lambda > 0$, if it has the probability mass function

$$P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}, \qquad k \in \mathbb{N}_0.$$

Lemma A.2. *Let X and Y be two independent Poisson distributed random variables with parameters $\lambda, \mu > 0$, i.e. $X \sim Poi(\lambda)$ and $Y \sim Poi(\mu)$. Then*

- $E[X] = \mathrm{Var}(X) = \lambda$
- $X + Y \sim Poi(\lambda + \mu)$.
- $E[X^k] \leq \lambda^k \exp(\frac{k^2}{2\lambda})$, *for all $k > 0$.*

Definition A.3. A random variable X is called Gumbel distributed with location parameter $\mu \in \mathbb{R}$ and scale parameter $\beta > 0$, if it for $x \in \mathbb{R}$ has the cumulative distribution function

$$F(x|\mu, \beta) = e^{-e^{-(x-\mu)/\beta}}.$$

If we set $\beta = 1$ and $\mu = \log(2)$, we get the version from the theorems under the null hypothesis, i.e.

$$F(x|\mu, \beta) = e^{-2e^{-x}}.$$

B Stochastic Landau symbols

Definition B.1. Let $(X_n)_{n\in\mathbb{N}}$ and $(U_n)_{n\in\mathbb{N}}$ be time series with $U_n > 0$ for all $n \in \mathbb{N}$, then

(i) $X_n = o_P(U_N) \Leftrightarrow \frac{X_n}{U_n} \xrightarrow{P} 0$,

(ii) $X_n = o_{a.s.}(U_N) \Leftrightarrow \frac{X_n}{U_n} \xrightarrow{a.s.} 0$,

(iii) $X_n = O_P(U_N) \Leftrightarrow \forall \varepsilon > 0, \exists C > 0 : \forall n \in \mathbb{N} P\left(|X_n| > CU_n\right) < \epsilon$,

(iv) $X_n = O_{a.s.}(U_N) \Leftrightarrow \exists C > 0 : \forall n \in \mathbb{N}\, |X_n| > CU_n$ *a.s.*.

Lemma B.2. *Let $(X_n)_{n\in\mathbb{N}}$, $(Y_n)_{n\in\mathbb{N}}$, $(U_n)_{n\in\mathbb{N}}$ and $(V_n)_{n\in\mathbb{N}}$ be time series with $U_n, V_n > 0$ for all $n \in \mathbb{N}$ and X a different random variable. Then for both in probability and almost surely, it holds,*

(i) $X_n = o(U_N) \Rightarrow X_n = O(U_N)$,

(ii) $X_n = O(U_N), Y_n = O(V_N) \Rightarrow X_n + Y_n = O(\max(U_n, V_n))$, $X_nY_n = O(U_n, V_n)$,

(iii) $X_n = o(U_N), Y_n = o(V_N) \Rightarrow X_n + Y_n = o(\max(U_n, V_n)),\; X_n Y_n = o(U_n, V_n)$,

(iv) $X_n = O(U_N), Y_n = O(V_N) \Rightarrow X_n = O(V_n)$, *or if at least one* O *is replaced by* o, *then* $X_n = o(V_n)$.

(v) $X_n \xrightarrow{a.s.} X \Rightarrow X_n = O_{a.s.}(1)$,

We refer to Mann and Wald [1943] for the introduction of the concept and Section 2.2 of van der Vaart [1998] and section 6.9 of Hayashi [2000] for the latter results.

C Brownian Motion

Definition C.1. Let $(\Omega, \mathcal{A}, P)$ be a probability space. The time continuous process $(B_t)_{t\geq 0}$ is called Brownian motion (or Wiener process) if all the following properties hold,

(i) $B_0 = 0$ almost surely,

(ii) $(B_t)_{t\geq 0}$ has independent increments, i.e. for $t > 0$ the difference $B_{t+u} - B_t$, $u \geq 0$, are independent of B_s, $s \leq t$,

(iii) the difference $B_{t+u} - B_t$, $u, t \geq 0$ is normally distributed, i.e. $B_{t+u} - B_t \sim \mathcal{N}(0, u)$,

(iv) $(B_t)_{t\geq 0}$ is almost surely continuous in t.

D Proof of Lemma 8.12

(i) We prove the claim by induction. We get for $i = m + 1$

$$
\begin{aligned}
E[Y_{m+1}] &= E[E[Y_{m+1}|\mathcal{F}_m^n]] = E[g_\theta(Y_{m,n}, m+1, |n)] \\
&= \omega + \alpha E[Y_{m,n}] + \delta \left(\frac{1}{n}\right)^\gamma \\
&= \omega \frac{1-\alpha}{1-\alpha} + \omega \frac{\alpha}{1-\alpha} + \delta \left(\frac{1}{n}\right)^\gamma = E[Y_0] + \delta \left(\frac{1}{n}\right)^\gamma
\end{aligned}
$$

with $E[Y_{m,n}] = E[Y_0]$. Now let $i = 1, ..., n-m-1$, then

$$E[Y_{m+i+1,n}] = E[E[Y_{m+i+1}|\mathcal{F}^n_[m+i]]] = E[g_\theta(Y_{m+i,n}, m+i+1, |n)]$$
$$= \omega + \alpha E[Y_{i,n}] + \delta \left(\frac{i+1}{n}\right)^\gamma$$
$$= \omega + \alpha E[Y_0] + \delta \sum_{j=1}^{i} \alpha^{i-j+1} \left(\frac{j}{n}\right)^\gamma + \delta \left(\frac{i+1}{n}\right)^\gamma$$
$$= E[Y_0] + \delta \sum_{j=1}^{i+1} \alpha^{i-j+1} \left(\frac{j}{n}\right)^\gamma$$

which proves the claim.

(ii) For this case, we can use the result from (i) and calculate the arithmetic mean, resulting in

$$E[\overline{Y}_n] = \frac{m}{n} E[Y_0] + \frac{1}{n} \sum_{i=1}^{n-m} \left(E[Y_0] + \delta \sum_{j=1}^{i} \alpha^{i-j} \left(\frac{j}{n}\right)^\gamma \right)$$
$$= E[Y_1] + \frac{\delta}{n^{\gamma+1}} \sum_{i=1}^{n-m} \sum_{j=1}^{i} \alpha^{i-j} j^\gamma$$
$$= E[Y_1] + \frac{\delta}{n^{\gamma+1}} \sum_{i=1}^{n-m} i^\gamma \sum_{j=0}^{n-m-i} \alpha^j$$
$$= E[Y_1] + \frac{\delta}{n^{\gamma+1}} \sum_{i=1}^{n-m} i^\gamma \frac{1 - \alpha^{n-m-i+1}}{1-\alpha},$$

which was the claim.

(iii) We calculate the mean of the squared expectation with the same approach and get

$$\frac{1}{n} \sum_{i=1}^{n} E[Y_{i,n}]^2 = \frac{m}{n} E[Y_0]^2 + \frac{1}{n} \sum_{i=1}^{n-m} \left(E[Y_0] + \delta \sum_{j=1}^{i} \alpha^{i-j} \left(\frac{j}{n}\right)^\gamma \right)^2$$
$$= E[Y_0]^2 + \frac{2\delta E[Y_0]}{n^{\gamma+1}} \sum_{i=1}^{n-m} \sum_{j=1}^{i} \alpha^{i-j} j^\gamma + \frac{\delta^2}{n^{2\gamma+1}} \sum_{i=1}^{n-m} \left(\sum_{j=1}^{i} \alpha^{i-j} j^\gamma \right)^2$$
$$= E[Y_0]^2 + \frac{2\delta E[Y_0]}{n^{\gamma+1}(1-\alpha)} \sum_{i=1}^{n-m} i^\gamma (1 - \alpha^{n-m-i+1}) + \frac{\delta^2}{n^{2\gamma+1}} \sum_{i=1}^{n-m} \left(\sum_{j=1}^{i} \alpha^{i-j} j^\gamma \right)^2.$$

(iv) We get the closed formula of the variance by induction. We start with $i = m + 1$:

$$
\begin{aligned}
\mathrm{Var}(Y_{m+1,n}) &= \mathrm{Var}(E[Y_{m+1,n}|\mathcal{F}^n_m]) + E[\mathrm{Var}(Y_{m+1,n}|\mathcal{F}^n_m)] \\
&= \mathrm{Var}(g_\theta(Y_{m,n}, m+1, |n)) + E[g_\theta(Y_{m,n}, m+1, |n)] \\
&= \alpha^2 \mathrm{Var}(Y_{m,n}) + \alpha E[Y_{m,n}] + \omega + \delta \left(\frac{1}{n}\right)^\gamma \\
&= \alpha^2 \mathrm{Var}(Y_{m,n}) + \alpha E[Y_0] + \omega + \delta \left(\frac{1}{n}\right)^\gamma \\
&= \alpha^2 \mathrm{Var}(Y_0) + E[Y_0] + \delta \left(\frac{1}{n}\right)^\gamma \\
&= \alpha^2 \mathrm{Var}(Y_0) + E[Y_{m+1,n}].
\end{aligned}
$$

The induction step proves that the variance is in fact given by (iv).

$$
\begin{aligned}
\mathrm{Var}(Y_{m+i+1,n}) &= \alpha^2 \mathrm{Var}(Y_{m+i,n}) + \alpha E[Y_{m+i,n}] + \omega + \delta \left(\frac{i+1}{n}\right)^\gamma \\
&= \alpha^2 \left(\alpha^{2i} \mathrm{Var}(Y_0) + \sum_{j=1}^{i} \alpha^{2(i-j)} E[Y_{m+j,n}] \right) \\
&\quad + \alpha E[Y_0] + \omega + \delta \sum_{j=1}^{i} \alpha^{i-j+1} \left(\frac{j}{n}\right)^\gamma + \delta \left(\frac{i+1}{n}\right)^\gamma \\
&= \alpha^{2(i+1)} \mathrm{Var}(Y_0) + \sum_{j=1}^{i} \alpha^{2(i-j+1)} E[Y_{m+j,n}] \\
&\quad + E[Y_0] + \delta \sum_{j=1}^{i+1} \alpha^{i-j+1} \left(\frac{j}{n}\right)^\gamma \\
&= \alpha^{2(i+1)} \mathrm{Var}(Y_0) + \sum_{j=1}^{i} \alpha^{2(i-j+1)} E[Y_{m+j,n}] + E[Y_{m+i+1,n}] \\
&= \alpha^{2(i+1)} \mathrm{Var}(Y_0) + \sum_{j=1}^{i+1} \alpha^{2(i-j+1)} E[Y_{m+j,n}]
\end{aligned}
$$

(v) To prove this representation of the arithmetic mean of the variances, we first calculate the arithmetic mean of the latter part of (iv).

$$\frac{1}{n}\sum_{i=1}^{n-m}\sum_{j=1}^{i}\alpha^{2(i-j)}E[Y_{m+j,n}]$$
$$= \frac{E[Y_0]}{n}\sum_{i=1}^{n-m}\sum_{j=1}^{i}\alpha^{2(i-j)} + \frac{\delta}{n^{\gamma+1}}\sum_{i=1}^{n-m}\sum_{j=1}^{i}\alpha^{2(i-j)}\sum_{k=1}^{j}k^{\gamma}\alpha^{j-k}$$
$$= \frac{E[Y_0]}{n}\sum_{i=1}^{n-m}\frac{1-\alpha^{2i}}{1-\alpha^2} + \frac{\delta}{n^{\gamma+1}}\sum_{i=1}^{n-m}\sum_{j=1}^{i}\alpha^{2(i-j)}\sum_{k=1}^{j}k^{\gamma}\alpha^{j-k}$$
$$= \frac{E[Y_0]}{n(1-\alpha^2)}\left(n-m-\alpha^2\sum_{i=0}^{n-m-1}\alpha^{2i}\right) + \frac{\delta}{n^{\gamma+1}}\sum_{i=1}^{n-m}\sum_{j=1}^{i}\alpha^{2(i-j)}\sum_{k=1}^{j}k^{\gamma}\alpha^{j-k}$$
$$= \frac{\mathrm{Var}(Y_0)}{n}\left(n-m-\alpha^2\frac{1-\alpha^{2(n-m)}}{1-\alpha^2}\right)$$
$$+ \frac{\delta}{n^{\gamma+1}}\sum_{i=1}^{n-m}\sum_{j=1}^{i}\alpha^{2(i-j)}\sum_{k=1}^{j}k^{\gamma}\alpha^{j-k}$$

Setting this in the formula of the arithmetic mean of the variance results in

$$\frac{1}{n}\sum_{i=1}^{n}\mathrm{Var}(Y_{i,n}) = \frac{m}{n}\mathrm{Var}(Y_0) + \frac{1}{n}\sum_{i=1}^{n-m}\alpha^{2i}\mathrm{Var}(Y_0)$$
$$+ \frac{1}{n}\sum_{i=1}^{n-m}\sum_{j=1}^{i}\alpha^{2(i-j)}E[Y_{m+j,n}]$$
$$= \frac{m}{n}\mathrm{Var}(Y_0) + \frac{\alpha^2\mathrm{Var}(Y_0)}{n}\frac{1-\alpha^{2(n-m)}}{1-\alpha^2}$$
$$+ \frac{\mathrm{Var}(Y_0)}{n}\left(\alpha(n-m)-\alpha^2\frac{1-\alpha^{2(n-m)}}{1-\alpha^2}\right)$$
$$+ \frac{\delta}{n^{\gamma+1}}\sum_{i=1}^{n-m}\sum_{j=1}^{i}\alpha^{2(i-j)}\sum_{k=1}^{j}k^{\gamma}\alpha^{j-k}$$
$$= \mathrm{Var}(Y_0) + \frac{\delta}{n^{\gamma+1}}\sum_{i=1}^{n-m}\sum_{j=1}^{i}\alpha^{2(i-j)}\sum_{k=1}^{j}k^{\gamma}\alpha^{j-k}$$

which is the claim.

(vi) First we note, that we can represent the expectation of $Y_{i,n}Y_{i+1,n}$ as

$$E[Y_{i,n}Y_{i-1,n}] = E[E[Y_{i,n}Y_{i-1,n}|\mathcal{F}_{i-1}^{n}]] = E[Y_{i-1,n}g_{\theta}(Y_{i-1,n}, i|n)]$$
$$= \alpha E[Y_{i-1,n}^2] + \left(\omega + \delta\left(\frac{m-i}{n}\right)_{+}^{\gamma}\right)E[Y_{i-1,n}].$$

Calculating the arithmetic mean then yields

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} E[Y_{i,n}Y_{i-1,n}] = \\
&\quad \alpha\frac{1}{n}\sum_{i=1}^{n} E[Y_{i-1,n}^2] + \frac{1}{n}\sum_{i=1}^{n}\left(\omega + \left(\frac{m-i}{n}\right)_+^{\gamma}\right) E[Y_{i-1,n}] \\
&\quad = \alpha\frac{1}{n}\sum_{i=1}^{n} E[Y_{i-1,n}^2] + \frac{\omega}{n}\sum_{i=1}^{n} E[Y_{i-1,n}] \\
&\qquad + \frac{\delta}{n^{\gamma+1}}\sum_{i=1}^{n-m} i^{\gamma}\left(E[Y_0] + \frac{\delta}{n^{\gamma}}\sum_{j=1}^{i-1} j^{\gamma}\alpha^{i-j}\right) \\
&\quad = \alpha\frac{1}{n}\sum_{i=1}^{n} E[Y_{i-1,n}^2] + \frac{\omega}{n}\sum_{i=1}^{n} E[Y_{i-1,n}] + \frac{\delta E[Y_0]}{n^{\gamma}}\sum_{i=1}^{n-m} i^{\gamma} \\
&\qquad + \frac{\delta^2}{n^{2\gamma+1}}\sum_{i=1}^{n-m} i^{\gamma}\sum_{j=1}^{i-1} j^{\gamma}\alpha^{i-j}.
\end{aligned}
$$

□

Akademischer Lebenslauf

10/2012 - 06/2015	Bachelor of Science in Mathematik, TU Kaiserslautern
10/2015 - 04/2018	Master of Science in Mathematik, TU Kaiserslautern
04/2022 - 12/2025	Promotion am Fachbereich Mathematik, RPTU Kaiserslautern-Landau

Academic curriculum vitae

10/2012 - 06/2015	Bachelor of Science in Mathematics, TU Kaiserslautern
10/2015 - 04/2018	Master of Science in Mathematics, TU Kaiserslautern
04/2022 - 12/2025	PhD at the department of mathematics, RPTU Kaiserslautern-Landau

References

T. Aach and A. Kaup. Bayesian algorithms for adaptive change detection in image sequences using markov random fields. *Signal Processing: Image Communication*, 7(2):147–160, 1995.

R. P. Adams and D. J. C. MacKay. Bayesian online changepoint detection. *Preprint*, 2007.

S. Aminikhanghahi and D. Cook. A survey of methods for time series change point detection. *Knowledge and Information Systems*, 51:339–367, 2017.

D. W. K. Andrews. Laws of large numbers for dependent non-identically distributed random variables. *Econometric Theory*, 4(3):458–467, 1988.

R. C. Bradley. *Introduction to Strong Mixing Conditions: Volume 1*. Kendrick Press, 2002.

R. C. Bradley. Basic properties of strong mixing conditions. A survey and some open questions. *Probability Surveys*, 2, 2005.

K. Brännäs and A.M.M. Shahiduzzaman Quoreshi. Integer-valued moving average modelling of the number of transactions in stocks. *Applied Financial Economics*, 20:1429–1440, 2010.

S. Chattopadhyay, R. Maiti, S. Das, and A. Biswas. Change-point analysis through integer-valued autoregressive process with application to some covid-19 data. *Statistica Neerlandica*, 76(1):4–34, 2022.

J. Chen and A. K. Gupta. *Parametric Statistical Change Point Analysis: With Applications to Genetics, Medicine, and Finance*. Birkhäuser, 2012.

C.-S. J. Chu and H. White. A direct test for changing trend. *Journal of Business & Economic Statistics*, 10(3):289–299, 1992.

Y. Cui, R. Wu, and Q. Zheng. Estimation of change-point for a class of count time series models. *Scandinavian Journal of Statistics*, 48(4):1277–1313, 2021.

S. Cunze, G. Schneider, N. Peter, and S. Klimpel. Linking patterns to processes: Using hunting bag data to classify raccoon (procyon lotor) invasion stages in germany since the 2000s. *Ecological Indicators*, 175, 2025.

D. Daly, W. Brown, H. Ingo, J. O'Leary, and D. Bradford. Change point detection in software performance testing. *Preprint*, 2020.

J. Davidson. *Stochastic Limit Theory: An Introduction for Econometricians*. Oxford University Press, 2021.

R. M. de Jong. A strong law of large numbers for triangular mixingale arrays. *Statistics & Probability Letters*, 27(1):1–9, 1996.

H. Dehling and W. Philipp. Almost sure invariance principles for weakly dependent vector-valued random variables. *The Annals of Probability*, 10(3): 689–701, 1982.

M. Diop and W. Kengne. Testing for parameter change in general integer-valued time series. *Journal of Time Series Analysis*, 38:880–894, 2016.

P. Doukhan and W. Kengne. Inference and testing for structural change in general Poisson autoregressive models. *Electronic Journal of Statistics*, 9(1): 1267–1314, 2015.

P. Doukhan, M. H. Neumann, and A. Leucht. Mixing properties of non-stationary INGARCH(1,1) processes. *Bernoulli*, 28:663–688, 2022.

R. Ferland, A. Latour, and D. Oraichi. Integer-valued GARCH processes. *Journal of Time Series Analysis*, 27:923–942, 2006.

J. Franke. Weak dependence of functional INGARCH processes. *Report in Wirtschaftsmathematik (WIMA Report)*, 126, 2010.

J. Franke and T. Seligmann. Conditional maximum likelihood estimates for INAR(1) processes and their application to modelling epileptic seizure counts. *Developments in Time Series Analysis*, pages 310–330, 1993.

J. Franke, W.K. Härdle, and C.M. Hafner. *Statistics of Financial Markets: An Introduction*. Universitext. Springer Berlin Heidelberg, 2010.

J. Franke, C. Kirch, and J. Kamgaing. Changepoints in time series of counts. *Journal of Time Series Analysis*, 33:757–770, 2012.

F. Hayashi. *Econometrics*. Princeton University Press, 2000.

L. Horvath. The Maximum Likelihood Method for Testing Changes in the Parameters of Normal Observations. *The Annals of Statistics*, 21(2):671–680, 1993.

P. J. Huber. Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35(1):73–101, 1964.

Š. Hudecová, M. Huskova, and S. Meintanis. Change detection in INARCH time series of counts. *Nonparametric Statistics*, 175:47–58, 2016.

M. Hušková and J. Steinebach. Limit theorems for a class of tests of gradual changes. *Journal of Statistical Planning and Inference*, 89(1):57–77, 2000.

W. Kengne and I. Ngongo. Inference for nonstationary time series of counts with application to change-point problems. *Annals of the Institute of Statistical Mathematics*, 74:801–835, 2022.

J. Kuelbs and W. Philipp. Almost sure invariance principles for partial sums of mixing b-valued random variables. *The Annals of Probability*, 8(6):1003–1036, 1980.

A. K. Lawrence and I. N. Paul. On conditional least squares estimation for stochastic processes. *The Annals of Statistics*, 6(3):629–642, 1978.

H. B. Mann and A. Wald. On Stochastic Limit and Order Relationships. *The Annals of Mathematical Statistics*, 14(3):217 – 226, 1943.

D. L. McLeish. A maximal inequality and dependent strong laws. *The Annals of Probability*, 3(5):829–839, 1975.

D. L. McLeish. On the Invariance Principle for Nonstationary Mixingales. *The Annals of Probability*, 5(4):616–621, 1977.

S. Meyn and R. Tweedie. *Markov Chains and Stochastic Stability*. Springer London, 2012.

M. H. Neumann. Absolute regularity and ergodicity of poisson count processes. *Bernoulli*, 17(4):1268–1284, 2011.

M. H. Neumann. Bootstrap for integer-valued GARCH(p, q) processes. *Statistica Neerlandica*, 75(3):343–363, 2021.

E. S. Page. Continuous inspection schemes. *Biometrika*, 41(1-2):100–115, 1954.

S. Plomer, T. Ernst, P. Gebhardt, E. Schleiff, R. Neininger, and G. Schneider. Bivariate change point detection in movement direction and speed. *Electronic Journal of Statistics*, 19(1):2121–2170, 2025.

B. M. Pötscher and I.R. Prucha. *Dynamic Nonlinear Econometric Models: Asymptotic Theory*. Springer Berlin Heidelberg, 1997.

B. M. Pötscher and I. R. Prucha. A uniform law of large numbers for dependent and heterogeneous data processes. *Econometrica*, 57(3):675–683, 1989.

B. M. Pötscher and I. R. Prucha. Generic uniform convergence and equicontinuity concepts for random functions: An exploration of the basic structure. *Journal of Econometrics*, 60(1):23–63, 1994.

M. A. Quddus. Time series count data models: An empirical application to traffic accidents. *Accident Analysis & Prevention*, 40(5):1732–1741, 2008.

J.-F. Quessy. Consistent nonparametric tests for detecting gradual changes in the marginals and the copula of multivariate time series. *Statistical Papers*, 60:717–746, 2019.

M. R. Csörgő and L. Horváth. *Limit Theorems in Change-Point Analysis*. John Wiley & Sons Ltd., 1997.

R. Ranga Rao. Relations between weak and uniform convergence of measures with applications. *The Annals of Mathematical Statistics*, 33(2):659–680, 1962.

M. Rosenblatt. A central limit theorem and a strong mixing condition. *Proceedings of the National Academy of Sciences of the United States of America*, 42 (1):43–47, 1956.

M. Staudacher, S. Telser, A. Amann, H. Hinterhuber, and M. Ritsch-Marte. A new method for change-point detection developed for on-line analysis of the heart beat variability during sleep. *Physica A: Statistical Mechanics and its Applications*, 349(3):582–596, 2005.

B. Straub and G. Schneider. A model for the study of the increase in stimulus and change point detection with small and variable spiking delays. *Neural Computation*, 32:1277–1321, 2020.

C. Truong, L. Oudre, and N. Vayatis. Selective review of offline change point detection methods. *Signal Processing*, 167, 2020.

G. J. J. van den Burg and C. K. I. Williams. An evaluation of change point detection algorithms. *Preprint*, 2020.

A. W. van der Vaart. *Asymptotic Statistics.* Cambridge University Press, 1998.

M. Vogt and H. Dette. Detecting gradual changes in locally stationary processes. *The Annals of Statistics*, 43(2):713–740, 2015.

P. Walters. *An Introduction to Ergodic Theory.* Springer New York, 1982.

C. Weiß. The INARCH(1) model for overdispersed time series of counts. *Communications in Statistics-simulation and Computation*, 39:1269–1291, 2010.

C. Weiß. *An Introduction to Discrete-Valued Time Series.* John Wiley & Sons Ltd., 2018.

C. Weiß and M. Testik. CUSUM monitoring of first-order integer-valued autoregressive processes of poisson counts. *Journal of Quality Technology*, 41: 389–400, 2009.

P. Yang and G. Dumont. Adaptive change detection in heart rate trend monitoring in anesthetized children. *IEEE transactions on bio-medical engineering*, 53, 2006.

Y. Zheng, L. Liu, L. Wang, and X. Xie. Learning transportation mode from raw GPS data for geographic applications on the web. *Proceedings of the 17th International Conference on World Wide Web*, page 247–256, 2008.