8-2018

Some Models for Count Time Series

Yisu Jia
Clemson University, yisujia@gmail.com

Some Models for Count Time Series

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Statistics

---

by
Yisu Jia
August 2018

Accepted by:
Dr. Robert Lund, Committee Chair
Dr. Peter Kiessler
Dr. Xin Liu
Dr. Brian Fralix

Abstract

There has been growing interest in modeling stationary series that have discrete marginal distributions. Count series arise when describing storm numbers, accidents, wins by a sports team, disease cases, etc.

The first count time series model introduced in this paper is the superpositioning methods. It have proven useful in devising stationary count time series having Poisson and binomial marginal distributions. Here, properties of this model class are established and the basic idea is developed. Specifically, we show how to construct stationary series with binomial, Poisson, and negative binomial marginal distributions; other marginal distributions are possible.

A second model class for stationary count time series – the latent Gaussian count time series model – is also proposed. The model uses a latent Gaussian sequence and a distributional transformation to build stationary series with the desired marginal distribution. This model has proven to be quite flexible. It can have virtually any marginal distribution, including generalized Poisson and Conway-Maxwell. It is shown that the model class produces the most flexible pairwise correlation structures possible, including negatively dependent series. Model parameters are estimated via two methods: 1) a Gaussian likelihood approach (GL), and 2) a particle filtering approach (PF).

Dedication

I dedicate this work to my loving parents – for their unconditional love and support. My family is the foundation for who I am.

Acknowledgments

First of all, I would like to thank my advisor Dr. Robert Lund. This work would not be completed without his guidance and support.

Secondly, I would like to thank my colleagues on the "Latent Gaussian Count Time Series Modeling" project: Dr. Vladas Pipiras, Dr. James Livsey, and Dr. Stefanos Kechagias. Their mathematical insights inspired me a lot in my research.

I would also like to thank Dr. Peter Kiessler, Dr. Xin Liu, and Dr. Brian Fralix for their insightful advice.

Finally, I would like to thank the Department of Mathematical Sciences for supporting me financially during my Ph.D studies.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Time Series Overview

A time series is a sequence of random variables collected over time. Most often, the measurements are made at regular time intervals. Time Series Analysis is used in many applications, including: economic forecasting, sales forecasting, stock market analysis, sports analysis, and many more.

The basic objectives of time series modeling are to fit a model that describes the structure of the time series and provides real-world interpretations. Uses for a fitted model are:

- To describe the important features of the time series, such as trend, seasonality, and change-points.
- To explain how the past affects the future, thus to permitting forecasting of future values of the series.

One difference of time series analysis from regression analysis is that the data are not necessarily independent. Let $X_1, X_2, \cdots, X_n$ be a time series of length n, denoted as $\{X_t\}_{t=1}^n$. The mean structure of $\{X_t\}_{t=1}^n$ is $\mu_t := \mathbb{E}[X_t]$. The covariance structure of $\{X_t\}_{t=1}^n$ is described by its autocovariance function (ACVF). The lag h ACVF at time t is defined as

$$\gamma_X(t, t+h) := \mathrm{Cov}(X_t, X_{t+h}) = \mathrm{E}(X_t X_{t+h}) - \mathrm{E}(X_t)\mathrm{E}(X_{t+h}).$$

A time series $\{X_t\}$ is said to be weakly stationary if it satisfies: 1) the mean $\mathrm{E}(X_t)$ is the same for all t, 2) the covariance between X_t and X_{t+h} is the same for all t and every $h \in \{0, 1, 2, \cdots\}$. Similarly, a time series $\{X_t\}$ is said to be strictly stationary if $(X_1, X_2, \cdots, X_n)'$ and $(X_{1+h}, X_{2+h}, \cdots, X_{n+h})'$ have the same joint distribution for all integer $h \geq 0$ and $n > 0$. Clearly, strictly stationary implies weakly stationary.

For a (weakly) stationary series $\{X_t\}$, the lag h ACVF does not depend on t for all h. In this setting,

$$\gamma_X(t, t+h) = \gamma_X(0, h).$$

For notational convenience, one can use a single argument in all ACVFs: $\gamma_X(h) := \gamma_X(0, h)$.

In some cases, it is easier to look at correlations instead of covariances. The autocorrelation function (ACF) of a stationary time series $\{X_t\}$ is defined as

$$\rho(h) := \mathrm{Corr}(X_t, X_{t+h}) = \frac{\gamma(h)}{\gamma(0)}.$$

Clearly, ACFs are between -1 and 1 by Cauchy-Schwarz inequality. In ACFs, the effect of dispersion in the series is removed. ACFs can be used to compare the level of dependency in different series.

For some series, it is worthwhile to pursue a partial autocorrelation function (PACF). In general, a partial correlation is a conditional correlation. For a time series, the partial autocorrelation between X_t and X_{t+h}, $h \geq 0$, is defined as the conditional correlation between X_t and X_{t+h}, conditional on $X_{t+1}, X_{t+2}, \cdots, X_{t+h-1}$:

$$\kappa(h) := \mathrm{Corr}(X_t, X_{t+h} | X_{t+1}, \cdots, X_{t+h-1}),$$

where the conditional correlation is taken between X_t and X_{t+h} after linear prediction of all variables between X_t and X_{t+h}.

1.1.1 ARMA and ARIMA Models

In stationary time series analysis, the most commonly-used model class is the autoregressive moving average (ARMA) model class. The general ARMA model was described in the thesis of Peter Whittle ([36]) and was popularized in the 1970 book by Box and Jenkins ([4]). The ARMA model

class relates the current observation of a series to past observations and past prediction errors. An ARMA(p, q) model includes autoregressive terms up to order p and moving-average terms up to order q. It obeys the recursion

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \cdots - \phi_p X_{t-p} = \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \cdots + \theta_q Z_{t-q},$$

where p and q are non-negative integers. The series $\{Z_t\}$ is white noises, that is often assumed to be independent and identically distributed in time t. When $p = 0$, an ARMA(p, q) model is referred to as a moving-average model of order q (MA(q)). Similarly, when $q = 0$, this model is called an autoregressive model of order p (AR(p)). For the model to remain stationary, some constraints need to be met on the values of the parameters. For example, the AR(1) model is stationary with $|\phi_1| \neq 1$ and not stationary with $|\phi_1| = 1$. A realization of 300 observations of an ARMA(1,2) series with $\phi_1 = 0.5, \theta_1 = 0.5, \theta_2 = 0.3$ is shown in Figure 1.1.

For non-stationary series, autoregressive integrated moving average (ARIMA) models can be used to describe patterns in the series. The elements in the model are specified in the form ARIMA(p, d, q), which include p autoregressive terms, q moving average terms, and d difference operations. More formally, a process $\{X_t\}$ is said to be ARIMA(p, d, q) if

$$(1 - B)^d X_t$$

is ARMA(p, q), where p, d, q are non-negative integers. Here, $(1 - B)^d$ is the dth order difference operator. For example, $(1 - B)X_t = X_t - X_{t-1}, (1 - B)^2 X_t = (1 - B)(X_t - X_{t-1}) = X_t - 2X_{t-1} + X_{t-2}, \cdots$. An ARIMA (1,1,1) model, for example, has one AR parameter and one MA parameter. A first-order difference allows for a linear trend in the data. Realizations of ARIMA(1,1,0) and ARIMA(0,1,1) series of length $n = 300$ are shown in Figure 1.2.

1.2 Count Time Series

There has been significant recent interest in modeling stationary series that have discrete marginal distributions. Often, the discreteness arises in the form of counts taking values in $\{0, 1, 2, \cdots\}$. Count series are widely used when describing storm numbers, accident tallies, wins

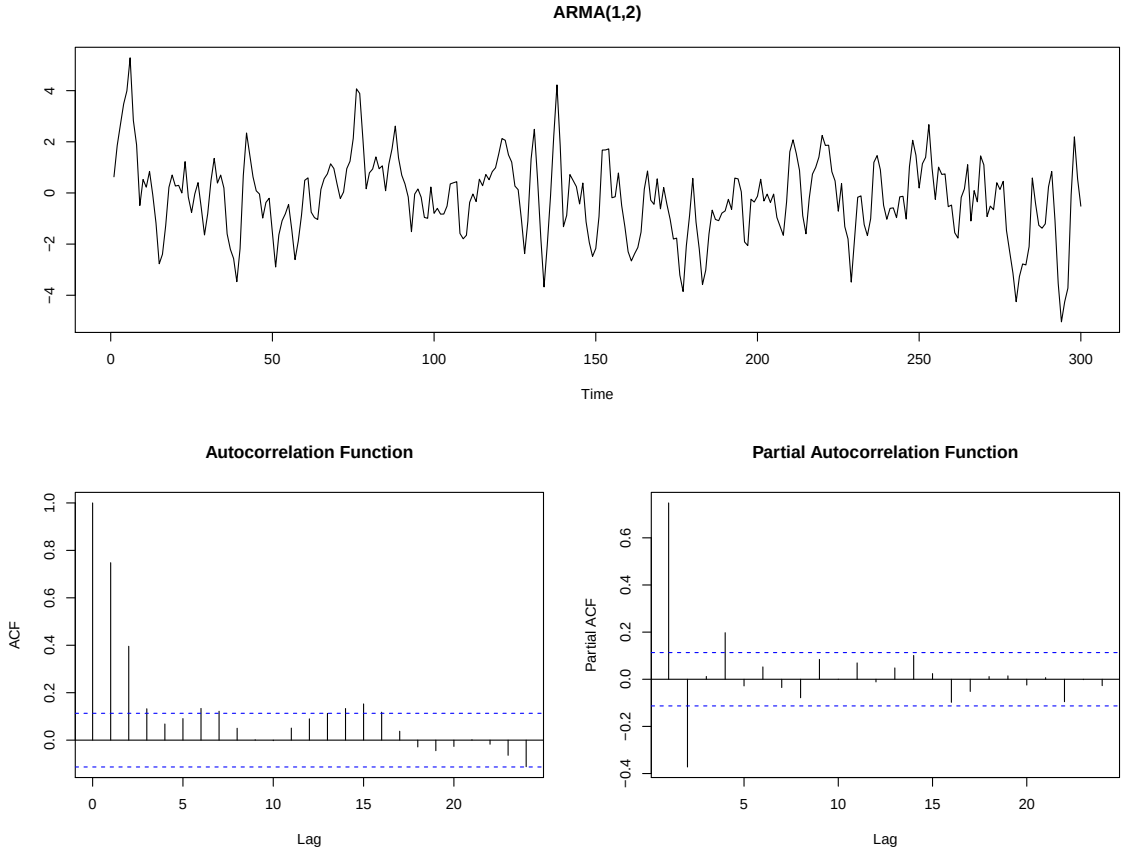

Figure 1.1: Realization of 300 observations of a ARMA(1,2) with $\phi_1 = 0.5, \theta_1 = 0.5, \theta_2 = 0.3$.

by a sports team, disease cases, etc. An example of a count time series is shown in Figure 1.3. It shows the annual number of Atlantic tropical cyclones from 1850 to 2011. The observation at each time is integer-valued, which clearly cannot be normally distributed.

The traditional ARMA/ARIMA model classes work well in describing series with Gaussian marginal distributions. However, no one definitive model class dominates the count series literature. In fact, the autocovariance function of many commonly used count models is deficient in some senses, as described below.

The theory of stationary Gaussian time series is well developed by now. However, there is no known result characterizing autocovariance functions of stationary count series. Elaborating, $\gamma_X(\cdot)$ is a symmetric non-negative definite function on the integers, if and only if there exists a stationary Gaussian sequence $\{X_t\}$ with $\mathrm{Cov}(X_t, X_{t+h}) = \gamma_X(h)$ for all integers h. Here, non-negative definite

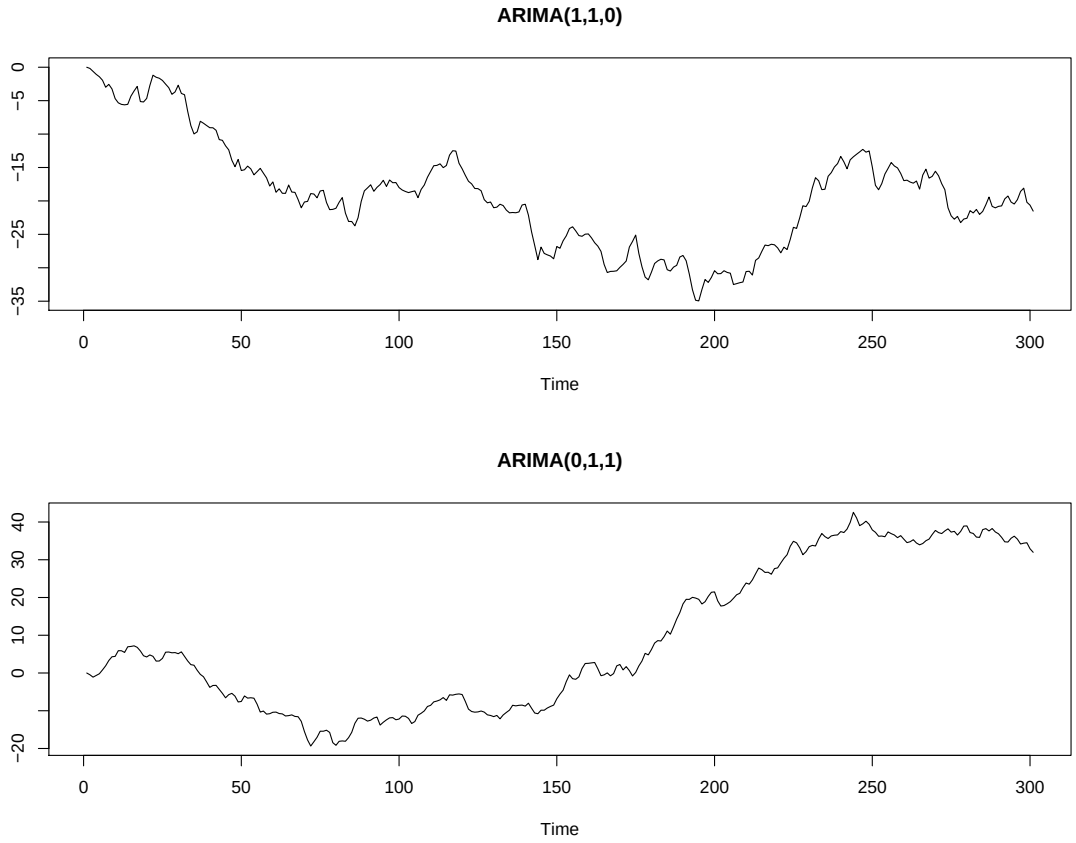

Figure 1.2: Top: realization of 300 observations of an ARIMA(1,1,0) series with $\phi_1 = 0.3$. Bottom: realization of 300 observations of an ARIMA(0,1,1) series with $\theta_1 = 0.5$.

is defined as

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i \gamma_X(t_i - t_j) a_j \geq 0$$

for every choice of $n \in \{1, 2 \ldots, \}$ and real numbers $a_1, \ldots, a_n$. Unfortunately, no analogous result exists for say, a stationary series with Poisson marginal distributions. In fact, restrictions on autocovariance functions of count time series are often more stringent than just non-negative definiteness. For example, it may not be possible to have a stationary count series having a specific marginal distribution that is highly negatively correlated at some lag while the autocovariance function can take on any value between -1 and 1 in a Gaussian process.

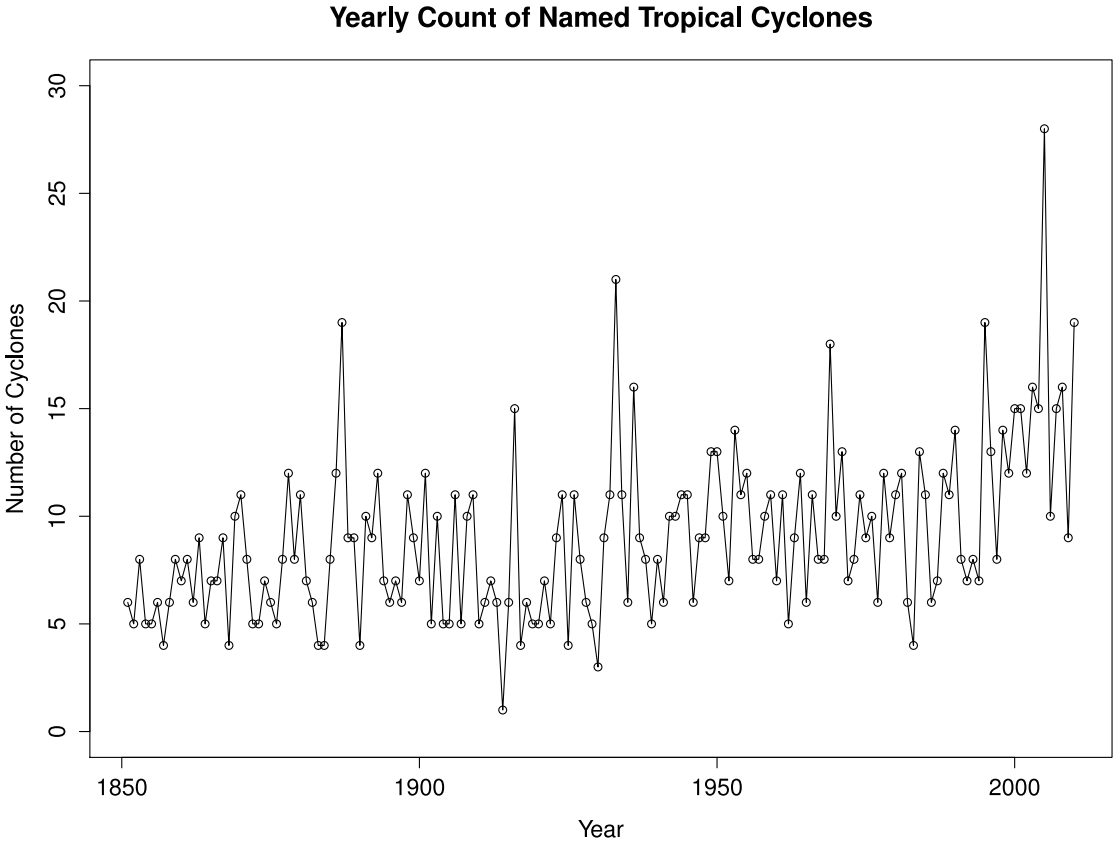

Figure 1.3: Annual number of Atlantic tropical cyclones from 1850 to 2011.

1.2.1 DARMA Models

Initial attempts to model stationary count series used the discrete-valued autoregressive moving-average (DARMA) methods introduced in the 1970s by Jacobs and Lewis ([14, 15]). The model used mixing techniques to build series with preset marginal distributions. For example, a first order discrete autoregressive (DAR(1)) series $\{X_t\}$ is built from independent and identically distributed (IID) count variables $\{Y_t\}$, say with marginal cumulative distribution $F(\cdot)$, and an IID Bernoulli thinning sequence $\{V_t\}$ with $\mathbb{P}(V_t = 1) := \rho \in [0, 1]$. The count series is initialized with $X_0 = Y_0$ and then recursively updated via

$$X_t = V_t X_{t-1} + (1 - V_t)Y_t, \quad t = 1, 2, \ldots.$$

Induction shows that X_t has distribution $F(\cdot)$ for every $t \in \{1, 2, \ldots\}$.

While one can build any marginal distribution for a DAR(1) series (in addition to discrete structures), there are some undesirable properties of DARMA series. Foremost, DAR(1) sample paths can remain constant for long runs. Since $\mathbb{P}(X_t = X_{t-1}) \geq \rho$, this issue becomes problematic for larger ρ. DAR(1) models cannot produce negatively correlated series either. This is because the thinning probability ρ must lie in $[0, 1]$. In fact, one can show that the DAR(1) autocorrelation has form $\text{Corr}(X_t, X_{t+h}) = \rho^h$ for all $h \geq 0$. While higher order autoregressions can be built by introducing additional IID Bernoulli trial sequences, negative correlations cannot be achieved with any DAR formulation. An example where negatively correlated count series are encountered with hurricane counts is introduced by Livsey in 2018 ([19]). In short, the DARMA covariance structure cannot be expected to describe all stationary count series. For these reasons and more, DARMA models fell out of favor in the 1980s.

1.2.2 INARMA Models

Another count time series model, and one that remains popular today, is the integer ARMA (INARMA) model class. INARMA models were introduced by Steutel in 1979 ([29]) and studied further in a series of papers in the 1980s ([1, 22, 23, 24]). For example, a first-order integer autoregressive (INAR(1)) model for $\{X_t\}$ obeys the recursion

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

Here, $\circ$ denotes a thinning operator that acts on a count-valued random variate Y via $\alpha \circ Y := \sum_{i=1}^{Y} B_i$, where $\{B_i\}_{i=1}^{\infty}$ is a sequence of zero/one Bernoulli trials with $\mathbb{P}(B_i = 1) = \alpha \in [0, 1]$ and $\{\epsilon_t\}$ is a sequence of IID count-valued random variables with mean μ_ϵ and finite variance σ_ϵ^2.

Unlike DARMA series, INARMA sample paths do not tend to stay constant for long runs; however, like the DARMA class, INARMA models cannot have negative correlations. In fact, the INAR(1) model has $\text{Corr}(X_t, X_{t+h}) = \alpha^h$ for any $h \geq 0$. One can construct higher order integer autoregressions and even add moving-average components as in [22, 23, 24]; however, one will not obtain a model with any negative correlations. Unlike DARMA series, it is not clear how to obtain any marginal distribution with INARMA methods — this, in fact, may be hard or impossible, depending on the marginal distribution desired.

1.2.3 Convolution-closed Infinitely Divisible Model Class

Another count series model type was proposed by Joe in 1996 ([16]). It produces stationary series whose marginal distribution lies in the so-called convolution-closed infinitely divisible class. Suppose that F_θ is a marginal distribution whose convolution, denoted by $*$, satisfies $F_{\theta_1} * F_{\theta_2} = F_{\theta_1+\theta_2}$. For the first order autoregressive case, the count series $\{X_t\}$ obeys the recursion

$$X_t = A_t(X_{t-1}; \alpha) + \epsilon_t,$$

where $\{\epsilon_t\}$ are IID variables having the marginal distribution $F_{(1-\alpha)\theta}$, $\alpha \in [0, 1]$, and ϵ_t is independent of X_j for $j < t$. The operator A_t, which is IID in time t, is defined so that $A_t(Y)$ is a random variable whose marginal distribution is $F_{\alpha\theta}$ (see [16] for details). These models capably describe many marginal distributions — both discrete and continuous — and include gamma, beta, normal, binomial, Poisson, negative binomial, and generalized Poisson. However, the marginal distribution must come from the convolution-closed class. Unfortunately, again the correlations of these models cannot be negative.

1.2.4 RINAR Models

Recently, [17] produced negatively correlated count series models by rounding solutions to Gaussian autoregressive models. For example, a rounded autoregressive model of order p with location parameter μ and autoregressive parameters $\phi_1, \ldots, \phi_p$ obeys

$$X_t = \left\langle \mu + \sum_{j=1}^{p} \phi_j Y_{t-j} \right\rangle + \epsilon_t,$$

where $\langle x \rangle$, for $x \geq 0$, rounds x to its nearest integer (round down should this be non-unique), $\{Y_t\}$ is a Gaussian autoregressive series, and $\{\epsilon_t\}$ is count-valued IID noise. While such $\{X_t\}$ can have negative correlations, due to the rounding, it is difficult to construct a pre-specified marginal distribution in this model class.

1.3 Research Motivations

All historical count models devised to date have some nice features and individualized drawbacks. Bayesian approaches also exist for many count analyses, which usually do not demand a fixed marginal distribution. This dissertation focuses on building stationary count models with possibly negative and positive autocovariances while maintaining a fixed marginal distribution.

Another drawback of previous models is that none of them can generate count time series with long memory autocovariances. A stationary count series $\{X_t\}$ is said to have long memory if $\sum_{h=0}^{\infty} |\text{cov}(X_t, X_{t+h})| = \infty$. Methods of constructing count series with long memory will also be illustrated.

In what follows, Chapter 2 and 3 will introduce two methods for devising count time series that have desirable properties. Chapter 2 will build on the work of Cui and Lund ([7]), devising some count models where explicit autocovariance expressions are achieved. Chapter 3 then presents a very general copula-based technique. Here, explicit autocovariance expressions are not produced; however, a Hermite polynomial expansion provides series expressions that are very useful numerically.

Chapter 2

Superpositioned Stationary Count Time Series

This chapter introduces a different approach to model stationary count time series. Our count time series will be built from a stationary zero-one (binary) random sequence $\{B_t\}$. This tactic was used in [3] and further developed in [7] and [21]. The idea can be viewed as the time series extension of the fact that any discrete-valued distribution can be constructed from fair coin flips.

For notation, let $p_B = \mathbb{E}[B_t] \equiv \mathbb{P}(B_t = 1)$ be the mean of $\{B_t\}$ and denote its lag-h autocovariance by $\gamma_B(h) = \text{Cov}(B_t, B_{t+h})$. Then

$$\gamma_B(h) = \mathbb{P}(B_t = 1 \cap B_{t+h} = 1) - p_B^2 = p_B[\mathbb{P}(B_{t+h} = 1|B_t = 1) - p_B]. \tag{2.1}$$

Two quantities that will be important later are the h-step-ahead transition probabilities to a unit point: $p_{1,1} := \mathbb{P}(B_{t+h,i} = 1|B_{t,i} = 1)$ and $p_{0,1} := \mathbb{P}(B_{t+h,i} = 1|B_{t,i} = 0)$. Obviously, the two h-step-ahead transition probabilities to a zero point are $p_{1,0} := \mathbb{P}(B_{t+h,i} = 0|B_{t,i} = 1) = 1 - p_{1,1}$ and $p_{0,0} := \mathbb{P}(B_{t+h,i} = 0|B_{t,i} = 0) = 1 - p_{0,1}$.

2.1 Stationary Zero-One Series

2.1.1 Renewal zero-one point processes

We now discuss two models to efficiently construct $\{B_t\}$; other methods are also possible. Our first model uses the renewal times in a stationary discrete-time renewal process as in [3, 7, 21]. A stationary renewal process employs an initial random "lifetime" $L_0 \in \{0, 1, \ldots\}$ (delay) and a sequence of IID aperiodic lifetimes $\{L_i\}_{i=1}^{\infty}$ supported in $\{1, 2, \ldots\}$ with $\mu_L := \mathbb{E}[L_1] < \infty$. The random walk $\{S_n\}_{n=0}^{\infty}$ associated with the renewal process obeys $S_n = \sum_{i=0}^{n} L_i$ for $n \in \{0, 1, \ldots\}$. The zero-one process $\{B_t\}$ is simply set to unity at all renewal times:

$$B_t = \mathbf{1}_{[\cup_{n=0}^{\infty}\{S_n=t\}]}.$$

To have $\{B_t\}$ stationary, L_0 needs to be specially selected as the so-called first-derived distribution from the tails of L_1 [12]:

$$\mathbb{P}(L_0 = k) = \frac{\mathbb{P}(L_1 > k)}{\mu_L}, \quad k \in \{0, 1, \ldots\}.$$

The Elementary Renewal Theorem gives $p_B = \mu_L^{-1}$ and $\mathbb{P}(B_{t+h} = 1 | B_t = 1) = u_h$, where u_h is the probability of a renewal at time h in a so called *zero-delayed* process ($L_0 = 0$). Equation (2.1) now yields

$$\gamma_B(h) = \mathrm{Cov}(B_t, B_{t+h}) = \frac{1}{\mu_L}\left(u_h - \frac{1}{\mu_L}\right), \tag{2.2}$$

Notice that $\gamma_B(h) < 0$ if and only if $u_h < \mu_L^{-1}$, which happens for many renewal sequences. The parameters in a renewal binary sequence are those that describe the lifetime L_1. In what follows, L will denote a lifetime whose distribution is equivalent to any of $L_1, L_2, \ldots$. Point probabilities are $p_{1,1} = u_h$ and $p_{0,1} = p_B(1 - u_h)/(1 - p_B)$.

2.1.2 Clipped Gaussian Sequences

A second construct for $\{B_t\}$ uses a correlated latent Gaussian process $\{Z_t\}$ as in [19]. Specifically, let $\{Z_t\}$ be a correlated zero-mean unit-variance Gaussian random processes with $\mathrm{Corr}(Z_t, Z_{t+h}) = \rho_Z(h)$. Set $B_t = \mathbf{1}_{(Z_t > \kappa)}$ for some preset real κ. Then $\{B_t\}$ is a strictly stationary

binary sequence with $p_B = 1 - \Phi(\kappa)$; here, $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal random variable. This construct is very similar to the clipping tactics introduced in [30].

The autocovariance function of $\{B_t\}$ can be derived from bivariate normal probability calculations. As an illustration, suppose that $\kappa = 0$ so that $p_B = 1/2$. Then a classical multivariate normal orthant probability calculation gives

$$\mathbb{E}(B_t B_{t+h}) = \mathbb{P}(Z_t > 0 \cap Z_{t+h} > 0) = \frac{1}{4} + \frac{\sin^{-1}(\rho_Z(h))}{2\pi},$$

where [28] is used. Since $\mathbb{E}[B_t] \equiv 1/2$,

$$\gamma_B(h) = \frac{\sin^{-1}(\rho_Z(h))}{2\pi}, \quad \rho_B(h) = \frac{2\sin^{-1}(\rho_Z(h))}{\pi}. \tag{2.3}$$

In this case, lag h autocovariances and autocorrelations are negative if and only if $\rho_Z(h) < 0$. Hence, this model can assume negative covariances. Notice that $\rho_B(h)$ can take on any value in $[-1, 1]$.

When $p_B \neq 1/2$, one will need to invert the standard normal cumulative distribution to find the desired quantile of the standard normal distribution corresponding to p_B. The covariance function in this case is harder to derive explicitly as it involves integrating the bivariate normal density over an infinite rectangle that is not an orthant. For notation, define $G(x_1, x_2, \rho) := \mathbb{P}(Z_1 > x_1, Z_2 > x_2)$ for a bivariate normal random vector $(Z_1, Z_2)'$ with zero-mean, unit variance, and correlation ρ. A recent result on bivariate normal quadrant probabilities [19] gives

$$\begin{aligned} G(x_1, x_2, \rho) = 1 - \Phi(x_1) - \Phi(x_2) + \frac{1}{4}\sum_{m=0}^{\infty} \frac{(2\rho)^{2m}}{(2m)!} \frac{2x_1 x_2 F_1(m+\frac{1}{2}; \frac{3}{2}; \frac{-x_1^2}{2}) F_1(m+\frac{1}{2}; \frac{3}{2}; \frac{-x_2^2}{2})}{\Gamma^2(\frac{1}{2} - m)} \\ + \frac{1}{4}\sum_{m=0}^{\infty} \frac{(2\rho)^{2m+1}}{(2m+1)!} \frac{F_1(m+\frac{1}{2}; \frac{3}{2}; \frac{x_1^2}{2}) F_1(m+\frac{1}{2}; \frac{3}{2}; \frac{x_2^2}{2})}{\Gamma^2(\frac{1}{2} - m)}, \end{aligned}$$

where $F_1(\cdot; a, b)$ is the confluent hypergeometric function of the first kind with parameters a and b: $F_1(x; a, b) = \sum_{n=0}^{\infty} (a)_n x!/[n!(b)_n]$, where $(a)_n$ is the rising factorial defined by $(a)_0 = 1$ and $(a)_n = a(a+1)(a+2)\cdots(a+n-1)$. In these expressions, $x! = \Gamma(x+1)$ for $x > 0$ and $\Gamma(-x)$ for a non-integer $x > 0$ is defined via the recursion $\Gamma(x+1) = x\Gamma(x)$.

Turning to covariances,

$$\mathbb{E}[B_t B_{t+h}] = G(\kappa, \kappa, \rho_Z(h)) = \int_\kappa^\infty \int_\kappa^\infty \frac{\exp\left(\frac{x^2 - 2\rho_Z(h)xy + y^2}{2[1-\rho_Z(h)^2]}\right)}{2\pi\sqrt{1-\rho_Z(h)^2}} dydx$$

and $\gamma_B(h) = G(\kappa, \kappa, \rho_Z(h)) - p_B^2$. Transition probabilities can be verified as

$$p_{1,1} = \frac{G(\kappa, \kappa, \rho_Z(h))}{1 - \Phi(\kappa)}, \quad p_{0,1} = \frac{1 - \Phi(\kappa) - G(\kappa, \kappa, \rho_Z(h))}{\Phi(\kappa)}.$$

2.2 Superpositioning

We now move to superpositioned count series. Let $\{B_{t,i}\}$ for $i \in \{1, 2, \ldots\}$ denote IID copies of $\{B_t\}$. Our count series $\{X_t\}$ is built by superimposing a random number of IID copies of $\{B_t\}$:

$$X_t = \sum_{i=1}^{M_t} B_{t,i}. \tag{2.4}$$

Here, $\{M_t\}$ is an IID count-valued random sequence that is independent of all $\{B_{t,i}\}$. See [2] for more when $\{M_t\}$ is a Poisson process.

Let $\mathbb{E}(M_t) = \mu_M$ and $\text{var}(M_t) = \sigma_M^2$. It is obvious that $\{X_t\}$ in (2.4) is a count-valued strictly stationary random sequence with mean $\mathbb{E}[X_t] \equiv p_B \mu_M$. The following result establishes additional properties of $\{X_t\}$.

Theorem 2.2.1. *Let $\{X_t\}$ be the strictly stationary count series in (2.4). Then*

a) The probability generating function of X_t has form $\psi_X(u) := \mathbb{E}[u^{X_t}] = \psi_M(1 - p_B + p_B u)$, where $\psi_M(u) := \mathbb{E}[u^{M_t}]$ is the probability generating function of M_t.

b) The dispersion of $\{X_t\}$ is $D_X := var(X_t)/\mathbb{E}[X_t] = p_B D_M + 1 - p_B$, where $D_M := \sigma_M^2/\mu_M$ is the dispersion of M_t. X_t is over/under dispersed if and only if M_t is over/under dispersed.

c) The lag h autocovariance of $\{X_t\}$ Has the form $\gamma_X(h) = \kappa\gamma_B(h)$, where $\kappa = \mathbb{E}[\min(M_1, M_2)]$ when $h \neq 0$, and $\gamma_X(0) = \gamma_B(0)\mu_M + p_B^2\sigma_M^2$.

d) The lag h bivariate probability distributions of $\{X_t\}$ have form

$$\mathbb{P}(X_t = x_t, X_{t+h} = x_{t+h}) = \sum_{m_t=0}^{\infty} \sum_{m_{t+h}=0}^{\infty} H_{m_t, m_{t+h}}(x_t, x_{t+h}) f_M(m_t) f_M(m_{t+h}),$$

where $f_M(k) = \mathbb{P}(M_t = k)$ *and* $H_{m_t,m_{t+h}}(x_t, x_{t+h})$ *has the form identified in the Appendix.*

e) In the renewal case, $\{X_t\}$ *has long memory if and only if* $\mathbb{E}[L^2] = \infty$*. In the clipped Gaussian case,* $\{X_t\}$ *has long memory if and only if* $\{Z_t\}$ *has long memory.*

This theorem is proven in appendix A.1. Attempts to derive higher order (beyond bivariate) joint process distributions have not produced tractable expressions to date. This is unfortunate as it precludes using the processes' joint distribution to construct likelihood-based parameter estimators; nonetheless, the bivariate distribution above allows one to compute composite likelihood estimators [26, 25] and the covariance structure of the model permits pseudo-Gaussian likelihood parameter estimation.

2.3 Classical Count Marginal Distributions

This section constructs stationary time series with the classical count marginal distributions: binomial, Poisson and negative binomial.

2.3.1 Binomial Marginals

Count time series with binomial marginal distributions with M trials and success probability p_B are easily obtained: just take M_t equal to the constant M. The binomial distribution is under-dispersed with $D_X = 1 - p_B$. This model was introduced by [3] and studied further in [7], [8], [33], and [32]. By part c) of Theorem 1, when $h \neq 0$, the lag h autocovariance and autocorrelation of $\{X_t\}$ are

$$\gamma_X(h) = M\gamma_B(h), \quad \rho_X(h) = \frac{\gamma_B(h)}{p_B(1-p_B)},$$

and $\gamma_X(0) = Mp_B(1-p_B)$, $\rho_X(0) = 1$. Then by (2.2), in the renewal case, the lag h autocovariance and autocorrelation of $\{X_t\}$ are

$$\gamma_X(h) = \frac{M}{\mu_L}\left(u_h - \frac{1}{\mu_L}\right), \quad \rho_X(h) = \frac{\frac{1}{\mu_L}\left(u_h - \frac{1}{\mu_L}\right)}{p_B(1-p_B)}.$$

From our derived expressions in the clipped Gaussian case, the lag h autocovariance and autocorrelation of $\{X_t\}$ are, for $h \neq 0$,

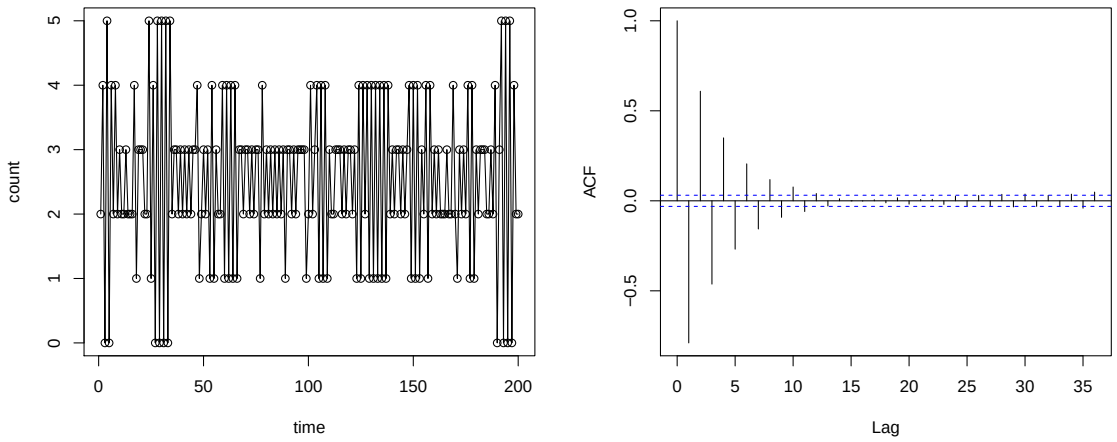

Figure 2.1: Two hundred points of a stationary count time series with Bin$(5, 0.5)$ marginal distribution. Sample autocorrelations and partial autocorrelations are shown with pointwise 95% confidence bands for white noise.

$$\gamma_X(h) = M\left[G(\kappa, \kappa, \rho_Z(h)) - p_B^2\right], \quad \rho_X(h) = \frac{G(\kappa, \kappa, \rho_Z(h)) - p_B^2}{p_B(1 - p_B)}.$$

Figure 2.1 shows a simulated realization of such a series. Here, the $B_{t,i}$s were generated from a renewal process with lifetime L supported on $\{1, 2, 3\}$, with $\mathbb{P}(L = 1) = \mathbb{P}(L = 3) = 0.1$ and $\mathbb{P}(L = 2) = 0.8$. Here, $\mathbb{E}[L] = 2$ and $p_B = 1/2$. From the sample autocorrelations plotted, it is evident that negative correlations are obtained.

2.3.2 Poisson Marginals

To construct a count time series $\{X_t\}$ with Poisson marginal distributions with mean $\lambda > 0$, let $\{M_t\}$ be an IID Poisson sequence with mean λ/p_B. It is easy to show that X_t in (2.4) has a Poisson distribution with mean λ (see [7]). The Poisson distribution has unit dispersion. The lag h autocovariance of this process has form $\gamma_X(h) = \kappa\gamma_B(h)$, for $h \neq 0$, where $\kappa = \mathbb{E}[\min(M_t, M_{t+h})]$ and M_t and M_{t+h} are independent Poisson random variables with mean λ/p_B. This is derived in [21] as

$$\kappa = \frac{\lambda\left\{1 - e^{-2\lambda/p_B}[I_0(2\lambda/p_B) + I_1(2\lambda/p_B)]\right\}}{p_B},$$

where $I_i(x)$ is the modified Bessel function

$$I_i(x) = \sum_{n=0}^{\infty} \frac{(x/2)^{2n+i}}{n!(n+i)!}, \quad i \in \{0, 1\}.$$

When $h = 0$, $\mathbb{E}[\min(M_t, M_{t+h})] = \lambda/p_B$. By Theorem 4.1, the lag h autocovariance of $\{X_t\}$ is $\gamma_X(h) = \kappa\gamma_B(h)$ when $h \neq 0$ and $\gamma_X(0) = \lambda$. In the renewal case, autocovariance and autocorrelation functions are

$$\gamma_X(h) = \frac{\kappa}{\mu_L}\left(u_h - \frac{1}{\mu_L}\right), \quad \rho_X(h) = \frac{\kappa}{\lambda\mu_L}\left(u_h - \frac{1}{\mu_L}\right).$$

Observe that $\gamma_X(h) < 0$ when $u_h < \mu_L^{-1}$, which happens for many renewal lifetimes L. In the clipped Gaussian case, the autocovariance and autocorrelation are

$$\gamma_X(h) = \kappa[G(\kappa, \kappa, \rho_Z(h)) - p_B^2], \quad \rho_X(h) = \frac{\kappa}{\lambda}[G(\kappa, \kappa, \rho_Z(h)) - p_B^2];$$

these are negative at lag h when $\rho_Z(h) < 0$.

Figure 2.2 shows a simulated realization of a stationary count time series with Poisson marginal distributions with mean $\lambda = 5$. The $B_{t,i}$s are generated from a clipped Gaussian process – a zero-mean unit variance AR(1) series with lag one autocorrelation of 0.9.

2.3.3 Negative Binomial Marginals

Count series with negative binomial marginal distributions are often used to model overdispersed count series [34, 31, 13]. The negative binomial distribution with parameters $r \in \{1, 2, \ldots\}$ and $p \in (0, 1)$ (NB(r, p)) has the probability mass function

$$\mathbb{P}(X_t = k) = \binom{r+k-1}{r-1} p^r (1-p)^k, \quad k \in \{0, 1, \cdots\}.$$

The dispersion of this distribution is $D_X = 1/(1-p) > 1$ and its probability generating function is

$$\psi_{X_t}(u) = \mathbb{E}[u^{X_t}] = \left(\frac{p}{1-(1-p)u}\right)^r, \quad |u| < (1-p)^{-1}.$$

To construct a negative binomial count series $\{X_t\}$ via the superposition in (2.4), apply part 1 of Theorem 4.1 to infer that the probability generating function of M_t must satisfy

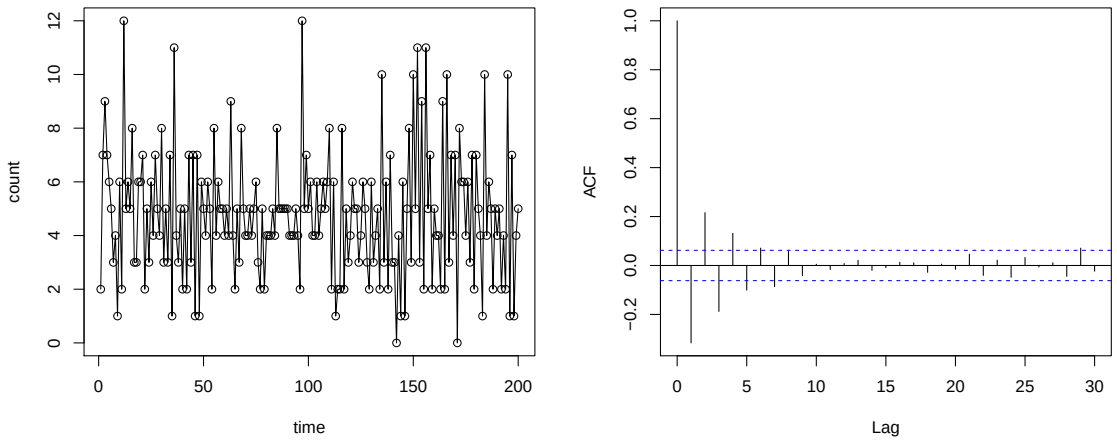

Figure 2.2: A realization of a stationary count time series with Poisson marginal distributions with mean 5. Sample autocorrelations and partial autocorrelations are shown with pointwise 95% critical intervals for white noise.

$$\psi_M(u) = \left(\frac{\frac{pp_B}{1-p+pp_B}}{1 - \frac{pp_B}{1-p+pp_B}u} \right)^r .$$

From this, it follows that the marginal distribution of $\{M_t\}$ is again negative binomial: $M_t \sim \text{NB}(r, \tilde{p})$, where $\tilde{p} = pp_B/(1-p+pp_B) \in [0,1]$.

Part c) of Theorem 1 shows that the lag h autocovariance of $\{X_t\}$ has form $\gamma_X(h) = \kappa\gamma_B(h)$ when $h \neq 0$ and $\gamma_X(0) = r(1-p)/p^2$, where $\kappa = \mathbb{E}[\min(M_t, M_{t+h})]$ and M_t and M_{t+h} are independent $\text{NB}(r, \tilde{p})$ variates. To find κ, note that

$$\kappa = \sum_{k=0}^{\infty} \mathbb{P}(\min(M_t, M_{t+h}) > k) = \sum_{k=0}^{\infty} \mathbb{P}(M_t > k)^2.$$

The tail probability $\mathbb{P}(M_t > k)$ for $M_t \sim \text{NB}(r, \tilde{p})$ can be calculated via a recursion in r. Specifically, M_t has the representation $M_t = A_1 + \cdots + A_r$, where the A_is are independent with tail distribution $\mathbb{P}(A_i > k) = (1-\tilde{p})^k$ for $k \in \{0, 1, 2, \ldots\}$ (a $\text{NB}(r = 1, \tilde{p})$ distribution). Let $\tilde{q} = 1 - \tilde{p}$ and condition on A_1 to get the recursion

$$\mathbb{P}(A_1 + \cdots + A_r > k) = \sum_{\ell=0}^{\infty} \mathbb{P}(A_2 + \cdots + A_r > k - \ell)\tilde{p}\tilde{q}^{\ell}.$$

With $\psi_r(k) = \mathbb{P}(A_1 + \cdots + A_r > k)$, we arrive at the difference equation

$$\psi_r(k) = \tilde{q}^{k+1} + \tilde{p} \sum_{\ell=0}^{k} \psi_{r-1}(k-\ell)\tilde{q}^{\ell},$$

which can be numerically evaluated recursively in r to obtain $\psi_r(k) = P(M_t > k)$, starting with $\psi_1(k) = \tilde{q}^{k+1}$.

Figure 2.3 shows a realization of stationary count time series with negative binomial marginal distribution with $r = 10$ and $p = 0.5$. The $B_{t,i}$s here are built from a renewal process whose lifetimes L have a Pareto distribution with parameter $\alpha = 2.1$: $\mathbb{P}(L = k) = C(\alpha)/k^{\alpha}$, where $k = 1, 2, \ldots$. Here, the constant $C(\alpha)$ makes the distribution of L sum to unity (there is no explicit form for $C(\alpha)$). In this case, L has a finite mean but an infinite second moment, implying from part e) of Theorem 4.1 that this series has long memory.

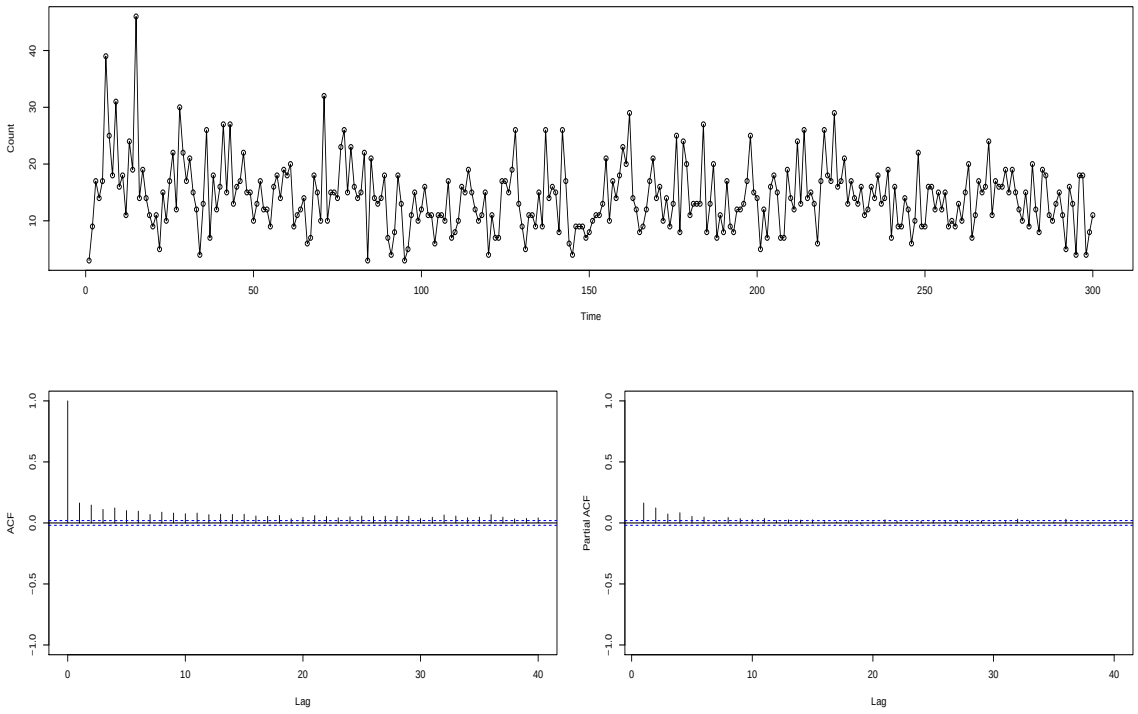

Figure 2.3: A realization of a long memory stationary count time series with $NB(10, 0.5)$ marginal distributions. Sample autocorrelations and partial autocorrelations are shown with pointwise 95% confidence bounds for white noise.

Despite having long memory, the series in Figure 3 does not exhibit large autocorrelations

at any lag. In fact, additional simulations with other parameters reveal the same drawback; to get correlations larger than 1/4 in absolute value (at any lag), one must take r on the order of hundreds. In this sense, the superposition tactics for negative binomial do not seem to work well. This said, negative correlations with this model can be achieved akin to the Poisson case.

Another way of combining the zero-one processes to construct negative binomial marginals draws from a tactic in [7]. Using that a negative binomial draw is the first time that r heads are obtained in independent coin flips (minus r to render a variable supported on $\{0, 1, \ldots\}$), set $M_t^{(r)} = \inf\{k \geq 1 : \sum_{i=1}^{k} B_{t,i} = r\}$ and $X_t = M_t^{(r)} - r$. Then X_t has a NB(r, p_B) distribution by construction. It also has a superpositioned form in that

$$X_t = \sum_{i=1}^{M_t^{(r)}} (1 - B_{t,i}). \tag{2.5}$$

The difference between (2.4) and (2.5), besides the $B_{t,i}$ versus the $1 - B_{t,i}$, is that $M_t^{(r)}$ in (2.5) is not independent of the $B_{t,i}$s, but is rather a stopping time constructed from them.

Explicit evaluation of the autocovariance function of this model is difficult, but can be done recursively in the integer r. Let $\psi_{i,j}(h) := \mathbb{E}[M_t^{(i)} M_{t+h}^{(j)}]$, where $M_t^{(i)} = A_{t,1} + \cdots + A_{t,i}$ and $M_{t+h}^{(j)} = A_{t+h,1} + \cdots + A_{t+h,j}$ are ordinary geometric random variables supported on $\{1, 2, \ldots\}$ (not that this support set does not contain zero). Since $\gamma_X(h) = \mathrm{cov}(M_t^{(r)}, M_{t+h}^{(r)})$, $\gamma_X(h) = \psi_{r,r}(h) - r^2/p_B^2$. The Appendix establishes the recursion

$$\begin{aligned} \psi_{i,j}(h) = {} & \frac{\psi_{i-1,j-1}(h) p_B p_{1,1} + \psi_{i,j-1}(h)(1 - p_B) p_{0,1} + \psi_{i-1,j}(h) p_B p_{1,0}}{1 - (1 - p_B) p_{0,0}} \\ & + \left(\frac{i + j - 1}{p_B} - \frac{p_{1,1}}{1 - (1 - p_B) p_{0,0}} \right) \frac{1}{1 - (1 - p_B) p_{0,0}} + \frac{[2 - (1 - p_B) p_{0,0}]}{[1 - (1 - p_B) p_{0,0}]^2} \end{aligned} \tag{2.6}$$

in $i, j \in \{1, 2, \ldots\}$. Of course, $\psi_{i,j}(h) = \psi_{j,i}(h)$. Boundary conditions take $\psi_{0,i}(h) = \psi_{i,0}(h) = 0$ and start with

$$\psi_{1,1}(h) = \frac{1}{p_B p_{0,1}} - \frac{p_B p_{0,0}}{[1 - (1 - p_B) p_{0,0}]^2 p_{0,1}} + \frac{p_{1,0}}{[1 - (1 - p_B) p_{0,0}]^2}. \tag{2.7}$$

For example, to get $\psi_{2,2}(h)$, one first uses (2.6) to get $\psi_{1,2}(h) = \psi_{2,1}(h)$. Using the recursion again gives $\psi_{2,2}(h)$ from $\psi_{1,2}(h)$ and $\psi_{2,1}(h)$.

Figure 2.4 shows a realization of a stationary count time series built from the above tech-

niques with negative binomial marginal distribution with $r = 10$ and $p = 0.5$. As with the last example, the $B_{t,i}$s are taken from a renewal process whose lifetimes L have a Pareto distribution with $\alpha = 2.1$. This time, covariances are much larger than those in Figure 3. Again, the model can take on negative correlations.

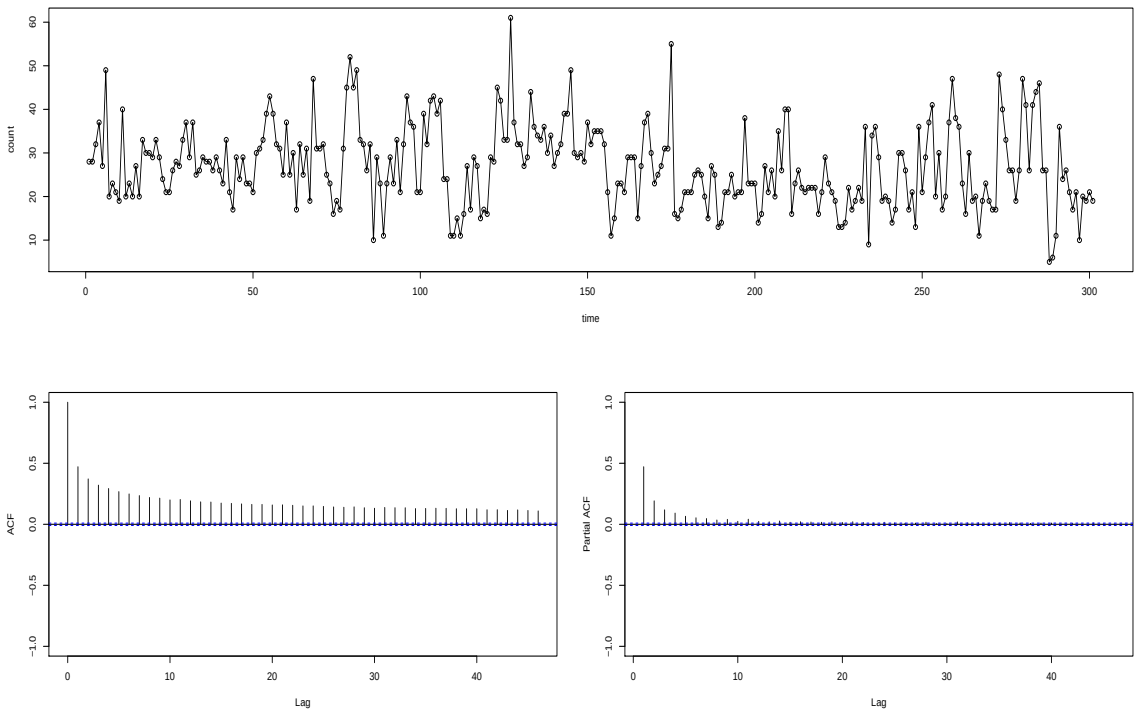

Figure 2.4: A realization of a stationary count time series with $NB(10, 0.5)$ marginal distributions. Sample autocorrelations and partial correlations are shown with pointwise 95% confidence bounds for white noise.

2.4 Other Marginal Distributions

This section turns to some other marginal distributions that can be built from our techniques. While any of the series in the last section can be built from either renewal or clipped Gaussian binary processes, in this section we find it more convenient to work with the same latent Gaussian processes, but to place them into more than two categories.

2.4.1 Discrete Uniform Marginals

The discrete uniform distribution on the categories $1, \ldots, M$, denoted by $\mathrm{DU}(M)$, has probably mass $\mathbb{P}(X_t = k) = 1/M$ for $k \in \{1, 2, \ldots, M\}$. This distribution might be useful in genetics ($M = 4$) (see [5]) and has dispersion $D_X = (M-1)/6$.

Let $\{Z_t\}$ be our latent stationary Gaussian process with zero mean, unit variance, and autocorrelation function $\rho_Z(\cdot)$. Let $A_1, A_2, \cdots, A_M$ be disjoint sets partitioning $\mathbb{R}$ into the equally likely intervals $A_1 = (-\infty, \Phi^{-1}(1/M)], A_2 = (\Phi^{-1}(1/M), \Phi^{-1}(2/M)]$,..., $A_M = (\Phi^{-1}((M-1)/M), \infty)$, where $\Phi^{-1}(\cdot)$ denotes the inverse of the standard normal cumulative distribution function. Now define

$$X_t = \sum_{i=1}^{M} i \mathbf{1}_{A_i}(Z_t).$$

By construction, $\{X_t\}$ is a series having discrete uniform marginal distributions. Non-equally likely categories can be produced by altering the A_i's to have non-equal standard normal probabilities.

The mean of the series is $\mathbb{E}[X_t] \equiv (M+1)/2$. To get the lag h covariance, note that

$$\mathbb{E}\left(X_t X_{t+h}\right) = \mathbb{E}\left[\left(\sum_{i=1}^{M} i \mathbf{1}_{A_i}(Z_t)\right)\left(\sum_{j=1}^{M} j \mathbf{1}_{A_j}(Z_{t+h})\right)\right] = \sum_{i=1}^{M}\sum_{j=1}^{M} ij \mathbb{P}\left(Z_t \in A_i, Z_{t+h} \in A_j\right)$$

The joint probability $\mathbb{P}(Z_t \in A_i, Z_{t+h} \in A_j)$ can be expressed in terms of $G(x_1, x_2, \rho_Z(h))$. For example, if $A_i = (\Phi^{-1}((i-1)/M), \Phi^{-1}(i/M))$ and $A_j = (\Phi^{-1}((j-1)/M), \Phi^{-1}(j/M))$,

$$\begin{aligned} K(M; i, j) &:= \mathbb{P}(Z_t \in A_i, Z_{t+h} \in A_j) \\ &= G(\Phi^{-1}(i/M), \Phi^{-1}(j/M), \rho_Z(h)) - G(\Phi^{-1}((i-1)/M), \Phi^{-1}(j/M), \rho_Z(h)) \\ &- G(\Phi^{-1}(i/M), \Phi^{-1}(j-1)/M), \rho_Z(h)) + G(\Phi^{-1}((i-1)/M), \Phi^{-1}((j-1)/M), \rho_Z(h)). \end{aligned}$$

When $h \neq 0$, the autocovariance and autocorrelation functions are

$$\gamma_X(h) = \left(\sum_{i=1}^{M}\sum_{j=1}^{M} ijK(M; i, j)\right) - \left(\frac{M+1}{2}\right)^2, \quad \rho_X(h) = \frac{\left(\sum_{i=1}^{M}\sum_{j=1}^{M} ijK(M; i, j)\right) - \left(\frac{M+1}{2}\right)^2}{(M^2-1)/12}.$$

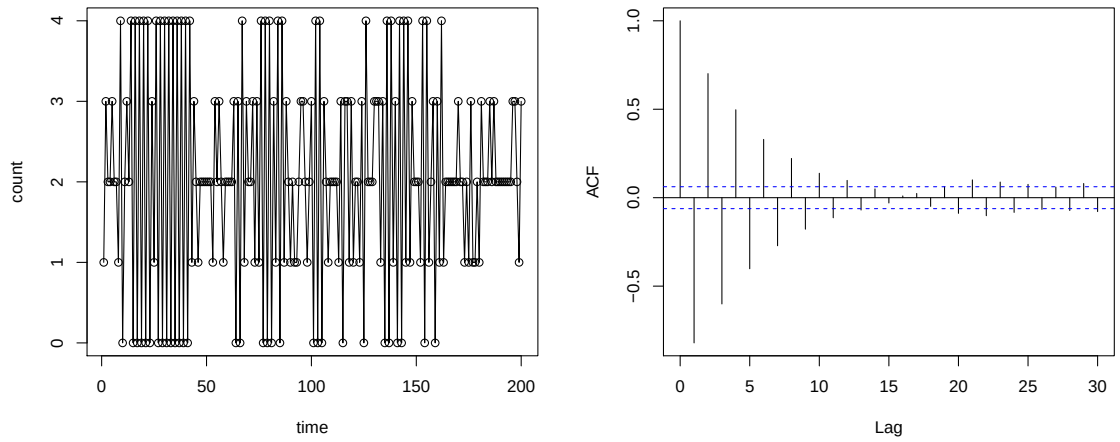

Figure 2.5: A realization of a stationary count time series with discrete uniform marginal supported on $\{1, 2, 3, 4, 5\}$. Sample autocorrelations and partial autocorrelations are shown with pointwise 95% confidence bounds for white noise.

Figure 2.5 shows a simulated realization of a count time series with DU(4) marginal distributions. Here, $\{Z_t\}$ is a stationary first order autoregression with $\rho_Z(1) = -0.9$. Negative correlations arise here whenever $\rho_Z(1) < 0$.

2.4.2 Multinomial Marginals

Our goal here is to construct a J-dimensional time series with multinomial marginal distributions with M trials and success probability vector $(p_1, p_2, \ldots, p_J)$ with $p_1 + p_2 + \ldots + p_J = 1$. We reduce to the case with one trial — results for a general M simply add M independent draws of one trial.

Now partition $\mathbb{R}$ into the J sets — call these $A_1, A_2, \ldots, A_J$ — so that $P(Z_1 \in A_j) = p_j$ for $j = 1, 2, \ldots, J$. Set $X_{t,j} = \mathbf{1}_{A_j}(Z_t)$ for $j = 1, 2, \ldots, J$. By construction, $\mathbf{X}_t := (X_{t,1}, X_{t,2}, \ldots, X_{t,j})$ has a multinomial distribution with one trial and success probabilities $p_1, \ldots, p_J$.

For a general M, $\mathbb{E}[X_{t,j}] = Mp_j$. To find lag h autocovariances, observe that $\mathbb{E}\left(X_{t,i}X_{t+h,j}\right) = M\mathbb{P}(Z_t \in A_i \cap Z_{t+h} \in A_j)$. It now follows that

$$\operatorname{cov}(X_{t,i}, X_{t+h,j}) = M\left[\mathbb{P}(Z_t \in A_i \cap Z_{t+h} \in A_j) - \mathbb{P}(Z_t \in A_i)P(Z_{t+h} \in A_j)\right]$$

We omit a graphic showing a sample path of $\{\mathbf{X}_t\}$ and its sample autocorrelations and partial autocorrelations. Again, one can have negative autocovariances.

2.5 Comments

This chapter presents methods to build stationary time series with common count marginal distributions that can take on very general autocovariance features, including negative correlations and long-memory. The methods build the series by combining correlated zero-one binary series in various ways. A superpositioning tactic worked well for producing Binomial and Poisson marginal distributions; however, a coin-tossing paradigm worked better for building negative binomial series. A distribution not pursued here but is worthy of further research is generalized Poisson. Also, statistical methods to fit these series are in need of development.

Chapter 3

Latent Gaussian Count Time Series Modeling

Another tactic used to generate a variety of series is a copula approach. Some words on this merit mention. Suppose $F(\cdot)$ is the desired marginal distribution of a count series. If $\{Z_t\}$ is a Gaussian process with standard normal marginal and autocorrelation function $\rho_Z(\cdot)$, then the sequence $\{X_t\}$ defined pointwise by $X_t = F^{-1}(\Phi(Z_t))$ will have marginal distribution $F(\cdot)$. This model can construct count series with any desired marginal distributions. However, the autocovariance function of $\{X_t\}$ is hard to explicitly derive in practice. Statistically, this would not matter if one could evaluate the data's likelihood function. However, the second drawback is that the joint distribution needed to derive the likelihood is difficult to obtain because of the discrete nature of F^{-1}. If F were continuous, then a simple Jacobian transformation method would suffice to evaluate the likelihood. But, the discrete nature of F^{-1} makes one have to quantify a discrete joint transformation, which appears difficult. In this chapter, we will introduce a method to use Hermite polynomial expansion to evaluate the autocovariance function of the count series $\{X_t\}$ numerically. Then we introduce particle filtration method to help us to approximate the full likelihood of the series.

We are interested in constructing stationary time series $\{X_t\}$ that have marginal distributions from several families of count structures supported in $\{0, 1, \ldots\}$, including:

- Binomial (Bin(N, p)): $\mathbb{P}[X_t = n] = \binom{N}{n} p^n (1 - p^{N-n})$, $n = 0, \ldots, N$, $p \in (0, 1)$;
- Poisson (Pois(λ)): $\mathbb{P}[X_t = n] = e^{-\lambda}\lambda^n/n!$, with $\lambda > 0$;

- Mixture Poisson (mixPois($\boldsymbol{\lambda}, \boldsymbol{p}$)): $\mathbb{P}[X_t = n] = \sum_{m=1}^{M} p_m e^{-\lambda_m} \lambda_m^n / n!$, where $\boldsymbol{p} = (p_1, \ldots, p_M)$ with the mixture probabilities $p_m > 0$ such that $\sum_{m=1}^{M} p_m = 1$ and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_M)$ with $\lambda_m > 0$;
- Negative binomial (NB(r, p)): $\mathbb{P}[X_t = n] = \frac{\Gamma(r+n)}{n!\Gamma(r)}(1-p)^r p^n$, with $r \geq 0$ and $p \in (0, 1)$;
- Generalized Poisson (GPois(λ, w)): $\mathbb{P}[X_t = n] = e^{-(\lambda + wn)} \lambda (\lambda + wn)^{n-1} / n!$, with $\lambda > 0$ and $w \in (0, 1)$;
- Conway-Maxwell-Poisson (CMP(λ, ν)): $\mathbb{P}[X_t = n] = \frac{\lambda^n}{(n!)^\nu C(\lambda, \nu)}$, with $\lambda > 0$, $\nu > 0$, and a normalizing constant $C(\lambda, \nu)$ making the probabilities sum to unity.

The negative binomial, generalized Poisson, and Conway-Maxwell-Poisson distributions are over-dispersed in that their variances are larger than their respective means. This is the case for sample variances and means of many observed count time series.

3.1 Theory

Let $\{X_t\}_{t \in \mathbb{Z}}$ be the stationary count time series of interest. Suppose that one wants the marginal cumulative distribution function (CDF) of X_t for each t of interest to be $F_X(x) = \mathbb{P}(X_t \leq x)$, depending on a vector $\boldsymbol{\theta}$ containing all model parameters. The series $\{X_t\}$ will be modeled through the copula type transformation

$$X_t = G(Z_t). \tag{3.1}$$

Here,

$$G(x) = F_X^{-1}(\Phi(x)), \quad x \in \mathbb{R}, \tag{3.2}$$

where $\Phi(\cdot)$ is the CDF of a standard normal variable and

$$F_X^{-1}(u) = \inf\{t : F_X(t) \geq u\}, \quad u \in (0, 1),$$

is the generalized inverse (quantile function) of the non-decreasing CDF F_X. The process $\{Z_t\}_{t\in\mathbb{Z}}$ is assumed to be standard Gaussian, but possibly correlated in time t:

$$\mathbb{E}[Z_t] = 0, \quad \mathbb{E}[Z_t^2] = 1; \tag{3.3}$$

that is, each $Z_t \sim \mathcal{N}(0,1)$ for each t. This approach was recently used by [11] in spatial settings with good results. The autocovariance function (ACVF) of $\{Z_t\}$ at lag h is denoted by

$$\gamma_Z(h) = \mathbb{E}[Z_{t+h}Z_t]; \tag{3.4}$$

the autocovariance and autocorrelation of $\{Z_t\}$ coincide due to the standard normal assumptions.

The construction in (3.1) ensures that the marginal CDF of X_t is indeed $F_X(\cdot)$. Elaborating, the probability integral transformation theorem shows that $\Phi(Z_t)$ has a uniform distribution on $(0,1)$ for each t; a second application of the result justifies the claimed marginal distribution. Temporal dependence in $\{Z_t\}$ will induce temporal dependence in $\{X_t\}$ as quantified in the next section. For autocovariace notation, let

$$\gamma_X(h) = \mathbb{E}[X_{t+h}X_t] - \mathbb{E}[X_{t+h}]\mathbb{E}[X_t] \tag{3.5}$$

denote the ACVF of $\{X_t\}$, that depends on another vector $\boldsymbol{\eta}$ of parameters.

3.1.1 Relationship between Autocovariances

The autocovariace functions (ACVFs) of $\{X_t\}$ and $\{Z_t\}$ can be related using Hermite expansions (see Chapter 5 of Pipiras and Taqqu [27]). More specifically, let

$$G(z) = \mathbb{E}[G(Z_0)] + \sum_{k=1}^{\infty} g_k H_k(z) \tag{3.6}$$

be the expansion of $G(x)$ in terms of the Hermite polynomials

$$H_k(z) = (-1)^k e^{z^2/2} \frac{d^k}{dz^k} e^{-z^2/2}, \quad z \in \mathbb{R}. \tag{3.7}$$

The first three Hermite polynomials are $H_0(z) \equiv 1$, $H_1(z) = z$, and $H_2(z) = z^2 - 1$; higher order polynomials can be obtained from the recursion $H_k(z) = zH_{k-1}(z) - H'_{k-1}(z)$. The *Hermite coefficients* are

$$g_k = \frac{1}{k!}\int_{-\infty}^{\infty} G(z)H_k(z)\frac{e^{-z^2/2}dz}{\sqrt{2\pi}} = \frac{1}{k!}\mathbb{E}[G(Z_0)H_k(Z_0)]. \tag{3.8}$$

The relationship between $\gamma_X(\cdot)$ and $\gamma_Z(\cdot)$ is extracted from Chapter 5 of [27] as

$$\gamma_X(h) = \sum_{k=1}^{\infty} k!g_k^2\gamma_Z(h)^k := g(\gamma_Z(h)), \tag{3.9}$$

where the power series is

$$g(u) = \sum_{k=1}^{\infty} k!g_k^2u^k. \tag{3.10}$$

In particular,

$$\mathrm{Var}(X_t) = \gamma_X(0) = \sum_{k=1}^{\infty} k!g_k^2 \tag{3.11}$$

depends only on the parameters in the marginal distribution $F_X(\cdot)$. Note also that

$$\rho_X(h) = \sum_{k=1}^{\infty} \frac{k!g_k^2}{\gamma_X(0)}\gamma_Z(h)^k = h(\rho_Z(h)), \tag{3.12}$$

where ρ refers to autocorrelations and

$$h(u) = \sum_{k=1}^{\infty} \frac{k!g_k^2}{\gamma_X(0)}u^k := \sum_{k=1}^{\infty} h_k u^k. \tag{3.13}$$

The function h maps $[-1,1]$ into (but not necessarily onto) $[-1,1]$. For future reference, note also that $h(0) = 0$,

$$h(1) = \sum_{k=1}^{\infty} h_k = 1. \tag{3.14}$$

Using (3.6) and $E[H_k(Z_0)H_\ell(-Z_0)] = (-1)^k k!1_{[k=\ell]}$ gives

$$h(-1) = \mathrm{Corr}(G(Z_0), G(-Z_0)); \tag{3.15}$$

however, $h(-1)$ is not necessarily -1 in general. As such, $h(\cdot)$ "starts" at $(-1, h(-1))$, passes through $(0,0)$, and connects to $(1,1)$. Examples will be given in Section 3.1.4.

The quantities $g(\cdot)$ and $h(\cdot)$ are called *link* functions, and ACVF–link and ACF–link func-

tions, respectively, if additional precision is required. Similarly, $k!g_k^2$ and $h_k = k!g_k^2/\gamma_X(0)$ are called *link coefficients*. A key feature in (3.9) is that the effects of the marginal CDF $F_X(\cdot)$ and the ACVF $\gamma_Z(\cdot)$ are "decoupled" in the sense that the correlation parameters in $\{Z_t\}$ do not influence the g_k coefficients in (3.9) — this will be very useful in our ensuing estimation work.

Remark 3.1.1. The relationship (3.9) between the ACVFs of $\{X_t\}$ and $\{Z_t\}$ can be used to gauge short- and long-range dependence properties. Recall that a time series $\{Z_t\}$ is short-range dependent (SRD) if $\sum_{h=-\infty}^{\infty} |\gamma_Z(h)| < \infty$. According to one definition, a series $\{Z_t\}$ is long-range dependent (LRD) if $\gamma_Z(h) = L(h)h^{2d-1}$, where $d \in (0, 1/2)$ is the LRD parameter and L is a slowly varying function at infinity [27]. The ACVF of such LRD series satisfies $\sum_{h=-\infty}^{\infty} |\gamma_Z(h)| = \infty$. If $\{Z_t\}$ is SRD, then so is $\{X_t\}$. To see this, when $\mathbb{E}[Z_t^2] = 1$ and $\mathrm{Var}(X_t) < \infty$, this follows directly from (3.9): since $|\gamma_Z(h)|^k \leq |\gamma_Z(h)|$, note that

$$\sum_{h=-\infty}^{\infty} |\gamma_X(h)| \leq \sum_{h=-\infty}^{\infty} \sum_{k=1}^{\infty} g_k^2 k! |\gamma_Z(h)|^k \leq \sum_{k=1}^{\infty} g_k^2 k! \sum_{h=-\infty}^{\infty} |\gamma_Z(h)| = \mathrm{Var}(X_t) \sum_{h=-\infty}^{\infty} |\gamma_Z(h)| < \infty.$$

On the other hand, if $\{Z_t\}$ is LRD with parameter d, then $\{X_t\}$ can be either LRD or SRD. The conclusion depends, in part, on the Hermite rank of $G(\cdot)$, which is defined as $r = \min\{k \geq 1 : g_k \neq 0\}$. Specifically, if $d \in (0, (r-1)/2r)$, then $\{X_t\}$ is SRD; if $d \in ((r-1)/2r, 1/2)$, then $\{X_t\}$ is LRD with parameter $r(d - 1/2) + 1/2$ (see Pipiras and Taqqu [27], Proposition 5.2.4). For example, when the Hermite rank is unity, $\{X_t\}$ is LRD with parameter d for all $d \in (0, 1/2)$; when $r = 2$, $\{X_t\}$ is LRD with parameter $2d - 1/2$ for $d \in (1/4, 1/2)$.

Remark 3.1.2. The construction in (3.1)–(3.2) yields models with very flexible autocorrelations. In fact, the methods achieve the most flexible correlation possible for $\mathrm{Corr}(X_{t_1}, X_{t_2})$ when X_{t_1} and X_{t_2} have the same marginal distribution F_X. Indeed, let $\rho_- = \min\{\mathrm{Corr}(X_{t_1}, X_{t_2}) : X_{t_1}, X_{t_2} \sim F_X\}$ and define ρ_+ similarly with min replaced by max. Then, as shown in Theorem 2.5 of Whitt [35],

$$\rho_+ = \mathrm{Corr}(F_X(U), F_X(U)) = 1, \quad \rho_- = \mathrm{Corr}(F_X(U), F_X(1-U)),$$

where U is a uniform random variable over $(0, 1)$. Since $U \stackrel{\mathcal{D}}{=} \Phi^{-1}(Z)$ and $1 - U \stackrel{\mathcal{D}}{=} \Phi^{-1}(-Z)$ for a standard normal random variable Z, the maximum and minimum correlations ρ_+ and ρ_- are indeed achieved with (3.1)–(3.2) when $Z_{t_1} = Z_{t_2}$ and $Z_{t_1} = -Z_{t_2}$, respectively. The preceding statements are non-trivial for ρ_- only since $\rho_+ = 1$ is attained whenever $X_{t_1} = X_{t_2}$. It is worthwhile to compare

this to the discussion surrounding (3.15). Finally, all correlations in $(\rho_-, \rho_+) = (\rho_-, 1)$ are achievable since $h(u)$ in (3.13) is continuous in u.

The preceding remark all but settles flexibility of autocovariance debates for stationary count series. Flexibility is a concern when the count series is negatively correlated, an issue arising in the hurricane data in [19]. Since a general count marginal distribution can also be achieved, the model class appears quite general.

3.1.2 Covariates

There are situations where stationarity is not desired. Such scenarios can often be accommodated with simple variants of the above setup. For concreteness, consider a situation where L non-random covariates are available to explain the series at time t — call these $M_{1,t}, \ldots, M_{L,t}$. If one wants X_t to have the marginal distribution $F_{\boldsymbol{\theta}(t)}(\cdot)$, where $\boldsymbol{\theta}(t)$ is a vector-valued function of t containing parameters, then simply set

$$X_t = F^{-1}_{\boldsymbol{\theta}(t)}(\Phi(Z_t)). \tag{3.16}$$

and reason as before.

Link functions can be used when bounds are needed for parametric support sets. As an example, a Poisson regression-type model is easily constructed as

$$\boldsymbol{\theta}(t) = E[X_t] = \exp\left(\beta_0 + \sum_{i=1}^{L} \beta_i M_{i,t}\right).$$

Here, the exponential link guarantees that the Poisson parameter is positive and $\beta_0, \ldots, \beta_L$ are regression coefficients. The above construct requires the covariates to be non-random Should covariates be random, the marginal distribution may change.

3.1.3 Calculation and properties of Hermite coefficients

Several strategies for computing the Hermite coefficients are available. We consider the stationary setting here for simplicity. Because G in (3.2) is discrete, the following approach proved to be simple, stable, and revealing. Let $\boldsymbol{\theta}$ denote all parameters appearing in the marginal distribution

of F_X. For $\boldsymbol{\theta}$ fixed, define the mass and cumulative probabilities of F_X via

$$p_n = \mathbb{P}[X_t = n], \quad C_n = \mathbb{P}[X_t \leq n] = \sum_{j=0}^{n} p_j, \quad n \in \{0, 1, \ldots\}. \tag{3.17}$$

Note that

$$G(z) = \sum_{n=0}^{\infty} n \, 1_{\{C_{n-1} \leq \Phi(z) < C_n\}} = \sum_{n=0}^{\infty} n \, 1_{\left[\Phi^{-1}(C_{n-1}), \Phi^{-1}(C_n)\right)}(z) \tag{3.18}$$

(take $C_{-1} = 0$ as a convention). When $C_n = 0$, we take $\Phi^{-1}(C_n) = -\infty$ and, for $C_n = 1$, $\Phi^{-1}(C_n) = \infty$. Using this in (3.8) provides, for $k \geq 1$,

$$g_k = \frac{1}{k!}\mathbb{E}[G(Z_0)H_k(Z_0)] = \frac{1}{k!}\sum_{n=0}^{\infty} n\mathbb{E}\left[1_{\left[\Phi^{-1}(C_{n-1}), \Phi^{-1}(C_n)\right)}(Z_0)H_k(Z_0)\right].$$

Using (3.7) and simplifying provides

$$\begin{aligned}
g_k &= \frac{1}{k!}\sum_{n=0}^{\infty} \frac{n}{\sqrt{2\pi}} \int_{\Phi^{-1}(C_{n-1})}^{\Phi^{-1}(C_n)} H_k(z) e^{-z^2/2} dz \\
&= \frac{1}{k!}\sum_{n=0}^{\infty} \frac{n}{\sqrt{2\pi}} \int_{\Phi^{-1}(C_{n-1})}^{\Phi^{-1}(C_n)} (-1)^k \Big(\frac{d^k}{dz^k} e^{-z^2/2}\Big) dz \\
&= \frac{1}{k!}\sum_{n=0}^{\infty} \frac{n}{\sqrt{2\pi}} (-1)^k \Big(\frac{d^{k-1}}{dz^{k-1}} e^{-z^2/2}\Big)\Big|_{z=\Phi^{-1}(C_{n-1})}^{\Phi^{-1}(C_n)} \\
&= \frac{1}{k!}\sum_{n=0}^{\infty} \frac{n}{\sqrt{2\pi}} (-1) e^{-z^2/2} H_{k-1}(z)\Big|_{z=\Phi^{-1}(C_{n-1})}^{\Phi^{-1}(C_n)} \\
&= \frac{1}{k!\sqrt{2\pi}}\sum_{n=0}^{\infty} n \left[e^{-\Phi^{-1}(C_{n-1})^2/2} H_{k-1}(\Phi^{-1}(C_{n-1})) - e^{-\Phi^{-1}(C_n)^2/2} H_{k-1}(\Phi^{-1}(C_n))\right].
\end{aligned} \tag{3.19}$$

Using the telescoping nature of the series (3.19) reveals that

$$g_k = \frac{1}{k!\sqrt{2\pi}}\sum_{n=0}^{\infty} e^{-\Phi^{-1}(C_n)^2/2} H_{k-1}(\Phi^{-1}(C_n)) \tag{3.20}$$

(convergence issues are dealt with in Remark 3.1.3 below). When $\Phi^{-1}(C_n) = \pm\infty$ (that is, $C_n = 0$ or 1), the summand $e^{-\Phi^{-1}(C_n)^2/2} H_{k-1}(\Phi^{-1}(C_n))$ is interpreted as 0. Before proceeding, the following remarks clarify a number of issues related to these coefficients. As noted in these remarks and also in the next section, (3.20) is particularly appealing from a numerical standpoint and also sheds light on the behavior of the Hermite coefficients.

Remark 3.1.3. One obtains (3.20) from (3.19) if, after changing $k-1$ to k for notational simplicity,

$$\sum_{n=0}^{\infty} e^{-\Phi^{-1}(C_n)^2/2}\Big|H_k(\Phi^{-1}(C_n))\Big| < \infty. \tag{3.21}$$

Such finiteness holds when $\{X_t\}$ has a finite variance. To see this, suppose that $C_n < 1$ for all n, since otherwise the sum in (3.21) has a finite number of terms. Since $H_k(z)$ is a polynomial of degree k, $|H_k(z)| \leq \kappa(1+|z|^k)$ for some constant κ that depends on k. The sum in (3.21) can hence be bounded (up to a constant) by

$$\sum_{n=0}^{\infty} e^{-\Phi^{-1}(C_n)^2/2}(1+|\Phi^{-1}(C_n)|^k). \tag{3.22}$$

To show that (3.22) converges, it suffices to show that

$$\sum_{n=0}^{\infty} e^{-\Phi^{-1}(C_n)^2/2}|\Phi^{-1}(C_n)|^k < \infty \tag{3.23}$$

since $|\Phi^{-1}(C_n)|^k \uparrow \infty$ as $C_n \uparrow 1$. Mill's ratio for a standard normal distribution states that $1-\Phi(x) \sim e^{-x^2/2}/(\sqrt{2\pi}x)$ as $x \to \infty$. Substituting $x = \Phi^{-1}(y)$ gives $1-y \sim e^{-\Phi^{-1}(y)^2/2}/(\sqrt{2\pi}\Phi^{-1}(y))$ as $y \uparrow 1$. Taking logs in the last relation and ignoring constant terms, order arguments show that $\Phi^{-1}(y) \sim \sqrt{2}|\log(1-y)|^{1/2}$ as $y \uparrow 1$. Substituting $\Phi^{-1}(C_n) \sim \sqrt{2}|\log(1-C_n)|^{1/2}$ in (3.23) provides

$$\sum_{n=0}^{\infty} e^{-\Phi^{-1}(C_n)^2/2}|\Phi^{-1}(C_n)|^k \leq \sum_{n=0}^{\infty} |\log(1-C_n)|^{k/2}(1-C_n). \tag{3.24}$$

For any $\delta > 0$ and $x \in (0,1)$, one can verify that $-\log(x) \leq x^{-\delta}/\delta$. Using this in (3.24) and $C_n = 1-\mathbb{P}[X > n]$, it suffices to prove that

$$\sum_{n=0}^{\infty} \mathbb{P}[X > n]^{1-\delta k/2} < \infty \tag{3.25}$$

for some $\delta > 0$. Since $X \geq 0$ and $\mathbb{E}[X^2] < \infty$ are assumed, the Markov inequality gives $\mathbb{P}[X > n] = \mathbb{P}[X^2 > n^2] \leq \mathbb{E}[X^2]/n^2$. Thus the sum in (3.25) is bounded by

$$\mathbb{E}[X^2]^{1-\delta k/2} \sum_{n=0}^{\infty} \frac{1}{n^{2-\delta k}}. \tag{3.26}$$

But (3.26) converges whenever $\delta < 2/k$. Choosing such a δ proves (3.21).

Remark 3.1.4. From a numerical standpoint, the expression (3.20) is evaluated as follows. The families of marginal distributions considered in this work have fairly "light" tails. This means that C_n approaches 1 rapidly as $n \to \infty$. In fact, for many distribution families, C_n becomes *exactly* 1 numerically for small to moderate values of n. Let $n(\boldsymbol{\theta})$ be the smallest such value. For example, for the Poisson distribution with parameter $\boldsymbol{\theta} = \lambda$ and Matlab software, $n(0.1) = 10$, $n(1) = 19$, and $n(10) = 47$. For $n \geq n(\boldsymbol{\theta})$, the numerical value of $\Phi^{-1}(C_n)$ is infinite and the terms $e^{-\Phi^{-1}(C_n)^2/2} H_{k-1}(\Phi^{-1}(C_n))$ in (3.20) are numerically zero and can be discarded. Thus, (3.20) becomes

$$g_k = \frac{1}{k!\sqrt{2\pi}} \sum_{n=0}^{n(\boldsymbol{\theta})-1} e^{-\Phi^{-1}(C_n)^2/2} H_{k-1}(\Phi^{-1}(C_n)). \tag{3.27}$$

Remark 3.1.5. Assuming that the g_k are evaluated through (3.27), their asymptotic behavior as $k \to \infty$ can be quantified. We focus on $g_k(k!)^{1/2}$, whose squares are the link coefficients. The asymptotic relation for Hermite polynomials states that $H_m(x) \sim e^{x^2/4}(m/e)^{m/2}\sqrt{2}\cos(x\sqrt{m} - m\pi/2)$ as $m \to \infty$ for each $x \in \mathbb{R}$. Using this and Stirling's formula, $k! \sim k^k e^{-k}\sqrt{2\pi k}$ as $k \to \infty$, show that

$$g_k(k!)^{1/2} \sim \frac{1}{2^{1/4}\pi^{3/4}} \frac{1}{k^{3/4}} \sum_{n=0}^{n(\boldsymbol{\theta})-1} e^{-\Phi^{-1}(C_n)^2/4} \cos\left(\Phi^{-1}(C_n)\sqrt{k-1} - \frac{(k-1)\pi}{2}\right). \tag{3.28}$$

Numerically, this approximation, which does not involve Hermite polynomials, was found to be accurate for even moderate values of k. It also suggests that $k!g_k^2$ decay at most (up to a constant) as $k^{-3/2}$. While this might appear slow, these coefficients are multiplied by $\gamma_Z(h)^k$ in (3.9), which decays geometrically fast in k to zero, except in degenerate cases when $|\gamma_Z(h)| = 1$.

The computation and behavior of the link coefficients $h_k = k!g_k^2/\gamma_X(0)$ are now examined for several families of marginal distributions. Figure 3.1 shows plots of h_k on a vertical log scale over a range of parameter values for $k = 1, \ldots, 5$ for the Poisson and negative binomial (with $r = 3$) distributions. A number of observations are worth making.

Since $\sum_{k=1}^{\infty} h_k = 1$ and $h_k \geq 0$ by construction, the parameter values in Figure 3.1 with $\log(h_1)$ close to 0 (or h_1 close to 1) means that most of the "weight" in the link coefficients is contained in the first coefficient, with higher order coefficients being considerably smaller and decaying with increasing k. This takes place in the approximate ranges $\lambda > 1$ for the Poisson distribution and $p \in (0.1, 0.9)$ in the negative binomial distribution with $r = 3$. Such cases will be called *"condensed"*.

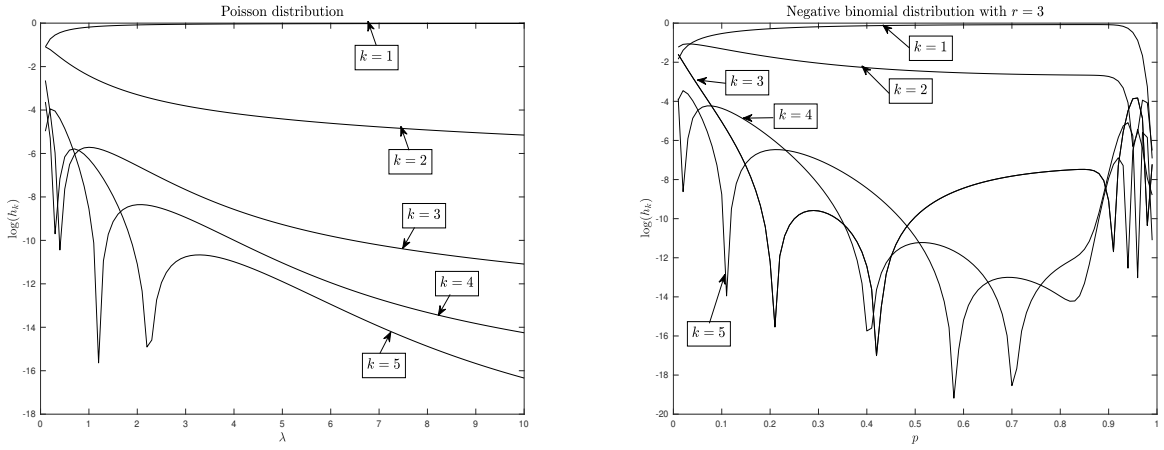

Figure 3.1: The link coefficients h_k on a log-vertical scale for the Poisson (left) and negative binomial (right) distributions.

As shown in Section 3.1.4 below, $h(z)$ in the condensed case is close to z. In the condensed case, the implication is that correlations in $\{Z_t\}$ and $\{X_t\}$ are similar.

Non-condensed cases are referred to as *"diffuse"*. Here, weights are spread to many link coefficients. This happens in the approximate ranges $\lambda < 1$ for the Poisson distribution and $p < 0.1$ and $p > 0.9$ for the negative binomial distribution with $r = 3$. This was expected for small λs and small ps: these cases correspond to discrete random structures that are nearly degenerate in the sense that they concentrate at 0 (as $\lambda \to 0$ or $p \to 0$). For such cases, large negative correlations in (3.15) are impossible; hence, $h(z)$ cannot be close to z and correlations in $\{Z_t\}$ and $\{X_t\}$ are different. The diffuse range $p > 0.9$ for the negative binomial distribution remains to be understood, although it can also probably be attributed to some form of "degeneracy."

3.1.4 Calculation and properties of link functions

We now study calculation of $h(u)$ in (3.13), which requires truncation of the sum to $k \in \{1, \ldots, K\}$ for some K. Note again that the link coefficients h_k are multiplied by $\gamma_Z(h)^k$ in (3.9), which decays geometrically fast in k to zero for most stationary $\{Z_t\}$ of interest when $h \neq 0$. The link coefficients for large k are therefore expected to play a minor role. We now set $K = 25$ and explore the consequences of this choice.

Remark 3.1.6. An alternative procedure would bound (3.28) by $(2\pi^3 k^3)^{-1/4} \sum_{n=0}^{n(\boldsymbol{\theta})-1} e^{-\Phi^{-1}(C_n)^2/4}$. Now let $K = K(\boldsymbol{\theta})$ be the smallest k for which the bound is smaller than some predetermined error

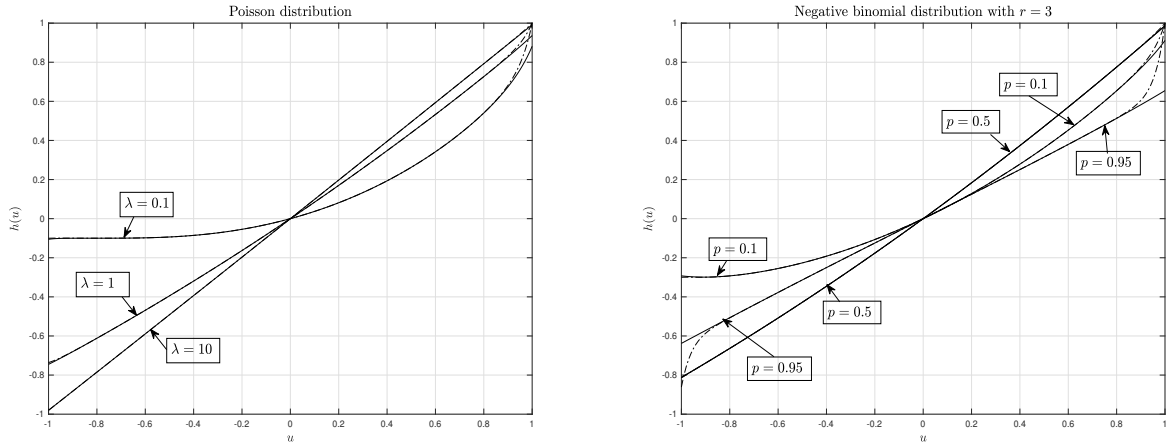

Figure 3.2: The link function $h(u)$ for the Poisson distribution with $\lambda = 0.1$, 1, and 10 (left) and the negative binomial distribution with $r = 3$ and $p = 0.1$, 0.5, and 0.95 (right).

tolerance ϵ. In the Poisson case with $\epsilon = 0.01$, for example, such K are $K(0.01) = 29, K(0.1) = 27$, and $K(1) = 25$. These are around the chosen value of $K = 25$.

Figure 3.2 plots $h(u)$ (solid line) for the Poisson and negative binomial distributions for several parameter values. The link function is computed by truncating its expansion to $k \leq 25$ as discussed above. The condensed cases $\lambda = 10$ and $\lambda = 1$ (perhaps this case is less condensed), and $p = 0.85$ lead to curves that are close to $h(u) \approx u$. However, the diffuse cases appear more delicate. Diffusivity and truncation of the infinite series in (3.13) lead to a computed link function that does not have $h(1) = 1$, in contrast to theoretical properties in (3.14); in this case, one should increase the number of terms in the summation.

Though deviations from $h(1) = 1$ might seem large (most notably for the negative binomial distribution with $p = 0.95$), this seems to arise only in the more degenerate cases associated with diffusivity; moreover, this occurs only when linking an ACVF of $\{Z_t\}$ for lags h for which $\rho_Z(h)$ is close to unity. For example, note that if the link deviation is 0.2 from 1 at $u = 1$ (as it is approximately for the negative binomial distribution with $p = 0.95$), the error for linking $\rho_Z(h)$ as 0.8 (or smaller but positive) would be no more than $0.2 \cdot (0.8)^{26} = 0.0006$! In practice, any link deviation could be partially corrected by adding one extra pseudo link coefficient, in our case the 26th coefficient, which would make the link function pass through $(1, 1)$. The resulting link function is depicted in the dashed line in Figure 3.2 around the point $(1, 1)$ and nearly coincides with the original link function for u values that are close to unity. It is this dashed link function connecting

to $(1, 1)$ that is used in practice for positive u.

The situation for negative u and, in particular, around $u = -1$ is different: the theoretical value of $h(-1)$ in (3.15) is not explicitly known. However, a similar correction could be achieved by first estimating $h(-1)$ through a Monte-Carlo simulation and adding a pseudo 26th coefficient making the computed link function connect to the desired value at $u = -1$. This is again depicted for negative u via the dashed lines in the Figure 3.2, which is visually distinguishable only near $u = -1$ (and then only in some cases). Again, it is this link function connecting to $(-1, h(-1))$ that is used in practice for negative values of u.

Remark 3.1.7. In estimation (Section 3.3 below), a link function needs to be evaluated multiple times; hence, running Monte-Carlo simulations to evaluate $h(-1)$ can become computationally expensive. In this case, the estimation procedure is fed precomputed values of $h(-1)$ on a grid of parameter values and interpolation is used for any intermediate parameter values.

3.2 Particle filtering and the HMM connection

This section studies the implications of the latent structure of our model, especially as it relates to hidden Markov models (HMMs). Our main reference is [10]. As in that monograph, the observations are taken to start at time zero. The following prediction and notations are key: let $\widehat{Z}_{t+1} = \widehat{z}_{t+1}(Z_0, \ldots, Z_t)$ denote the one-step-ahead linear prediction of the latent Gaussian series Z_{t+1} from the history $Z_0, Z_1, \ldots, Z_t$. This will be expressed as $\widehat{Z}_{t+1} = \phi_{t0} Z_t + \ldots + \phi_{tt} Z_0$. The weights ϕ_{ts}, $s = 0, \ldots, t$, can be computed recursively in t and efficiently from the ACVF of $\{Z_t\}$ by using the Durbin-Levinson (DL) algorithm, for example. By convention, $\widehat{Z}_0 = 0$. Let also $r_t^2 = \mathbb{E}[(Z_t - \widehat{Z}_t)^2]$ be the corresponding mean-squared error.

We are interested in the following problems:

$$\begin{aligned} \text{Filtering} \quad &: \quad \text{the distribution of } \widehat{Z}_{t+1|t}, \text{ i.e. } \widehat{Z}_{t+1} \text{ conditional on } X_0 = x_0, \ldots, X_t = x_t, \\ \text{Prediction} \quad &: \quad \text{the distribution of } \widehat{X}_{t+1|t}, \text{ i.e. } X_{t+1} \text{ conditional on } X_0 = x_0, \ldots, X_t = x_t, \end{aligned}$$

as well as computing numerically the quantities $\mathbb{E}_X[f(\widehat{Z}_{t+1|t})]$ and $\mathbb{E}_X[f(\widehat{X}_{t+1|t})]$ for some function f, where $\mathbb{E}_X$ refers to an expectation conditioned on $X_0 = x_0, \ldots, X_t = x_t$. These quantities will be needed to evaluate model likelihoods in inference (Section 3.3.2) and the probability integral

transform in model diagnostics (Section 3.3.3).

Remark 3.2.1. Several comments are in place regarding the problems of interest above. Smoothing is another commonly considered problem in connection to HMMs. It is not considered here for several reasons. First, in the HMM setting, smoothing is commonly used in likelihood evaluation through versions of the EM algorithms (see, for example, Section 12 in Douc et al. [10]). But as explained in Remark 3.3.3 below, even when our model is an HMM, this standard approach does not apply to it because of the lack of the partial dominance property. Second, we work here with general underlying Gaussian series which are not necessarily Markovian. Our algorithms below extend filtering algorithms from HMM to the situations where the latent process is stationary Gaussian and possibly non-Markovian. We presently do not have efficient smoothing algorithms in the latter setting that extend those used in the HMM setting. Another comment related to these points concerns our filtering problem above. Note that $\widehat{Z}_{t+1} = \widehat{z}_{t+1}(Z_0, \ldots, Z_t)$ involves all previous values $Z_0, \ldots, Z_t$ and hence their conditional distribution given $X_0 = x_0, \ldots, X_t = x_t$. This suggests that evaluating the distribution of $\widehat{Z}_{t+1|t}$ could potentially be carried out through the smoothing problem. But since the operation $\widehat{z}_t = \widehat{z}_t(z_0, \ldots, z_{t-1})$ commonly acts as an exponential-type averaging of the previous values, we expect that direct, filtering-like approach in evaluating the distribution of $\widehat{Z}_{t+1|t}$ would be sufficient. In particular, one special case of the underlying Gaussian series Z_t is an autoregressive model of order 1 (AR(1)) satisfying $Z_t = \phi Z_{t-1} + (1-\phi)^{1/2}\epsilon_t$, where $|\phi| < 1$, $\{\epsilon_t\}$ consists of i.i.d. $\mathcal{N}(0,1)$ random variables and the presence of $(1-\phi)^{1/2}$ ensures that $\mathbb{E}Z_t^2 = 1$. In this case, $\widehat{Z}_{t+1} = \phi Z_t$ and then our filtering problem is equivalent to the conventional filtering problem of finding the distribution of Z_t given $X_0 = x_0, \ldots, X_t = x_t$.

For later reference, we first derive expressions for the above distributions of interest. It will be convenient here and below in the paper to use the notation

$$A(x) = \{z : \Phi^{-1}(f_{x-1}) \le z < \Phi^{-1}(f_x)\}. \tag{3.29}$$

Its role for our model stems from the fact that

$$x = G(z) \Leftrightarrow z \in A(x) \tag{3.30}$$

(see (3.18)). The proof of the following result can be found in Appendix A.2, and uses a more general

auxiliary result of interest (see Lemma A.2.1).

Lemma 3.2.1. *With the above notation,*

$$\mathbb{E}_X[f(\widehat{Z}_{t+1|t})] = \frac{\int_{z_s\in A(x_s),s=0,\dots,t} f(\widehat{z}_{t+1})e^{-\frac{1}{2}\sum_{s=0}^{t}(z_s-\widehat{z}_s)^2/r_s^2}dz_0\dots dz_t}{\int_{z_s\in A(x_s),s=0,\dots,t} e^{-\frac{1}{2}\sum_{s=0}^{t}(z_s-\widehat{z}_s)^2/r_s^2}dz_0\dots dz_t} \tag{3.31}$$

and

$$\mathbb{E}_X[f(\widehat{X}_{t+1|t})] = \frac{\int_{z_s\in A(x_s),s=0,\dots,t} f(G(z_{t+1}))e^{-\frac{1}{2}\sum_{s=0}^{t+1}(z_s-\widehat{z}_s)^2/r_s^2}dz_0\dots dz_{t+1}}{\int_{z_s\in A(x_s),s=0,\dots,t} e^{-\frac{1}{2}\sum_{s=0}^{t+1}(z_s-\widehat{z}_s)^2/r_s^2}dz_0\dots dz_{t+1}}, \tag{3.32}$$

where $\mathbb{E}_X$ refers to the expectation conditioned on $X_0 = x_0, \dots, X_t = x_t$. We also have

$$\mathbb{E}_X[f(\widehat{X}_{t+1|t})] = \mathbb{E}_X[g_{f,t+1}(\widehat{Z}_{t+1|t})], \tag{3.33}$$

where

$$g_{f,t+1}(z) = \int_{\mathbb{R}} f(G(z_{t+1}))\frac{1}{\sqrt{2\pi r_{t+1}^2}}e^{-\frac{(z_{t+1}-z)^2}{2r_{t+1}^2}}dz_{t+1}. \tag{3.34}$$

Our algorithm related to the filtering problem is next described, and its connection to the HMM literature is clarified in a subsequent remark. The name of the algorithm is also motivated by the HMM connection described in the remark.

Sequential Importance Sampling (SIS) particle filtering: For $i = 1, \dots, N$, initialize the underlying Gaussian series Z_t^i by

$$Z_0^i \stackrel{d}{=} \Big(\mathcal{N}(0,1)\big|G(\mathcal{N}(0,1)) = x_0\Big), \tag{3.35}$$

that is, generate Z_t^i at time 0 assuming $X_0 = G(Z_0) = x_0$. In view of (3.30), this is equivalent to generating

$$Z_0^i \stackrel{d}{=} \Big(\mathcal{N}(0,1)\big|\Phi^{-1}(f_{x_0-1}) \le \mathcal{N}(0,1) < \Phi^{-1}(f_{x_0})\Big). \tag{3.36}$$

Set also $w_0^i = 1$. Then, recursively in $t = 1, \dots, T$, perform the following steps:

1: Compute $\widehat{Z}_t^i$ by using the DL algorithm and the previous values $Z_0^i, \dots, Z_{t-1}^i$.

2: Sample an error ϵ_t^i conditionally on having $X_t = x_t$, that is,

$$\epsilon_t^i \stackrel{d}{=} \Big(\mathcal{N}(0,1)\big|G(\widehat{Z}_t^i + r_t\mathcal{N}(0,1)) = x_t\Big) \tag{3.37}$$

or, in view of (3.30),

$$\epsilon_t^i \stackrel{d}{=} \Big(\mathcal{N}(0,1)\big|r_t^{-1}(\Phi^{-1}(f_{x_t-1}) - \widehat{Z}_t^i) \le \mathcal{N}(0,1) < r_t^{-1}(\Phi^{-1}(f_{x_t}) - \widehat{Z}_t^i)\Big), \tag{3.38}$$

where $f_{-1} = 0$ by convention (and $\Phi^{-1}(0) = -\infty$).

3: Update the underlying Gaussian series as

$$Z_t^i = \widehat{Z}_t^i + r_t\epsilon_t^i, \tag{3.39}$$

and also set

$$w_t^i = w_{t-1}^i w_t(\widehat{Z}_t^i), \tag{3.40}$$

where

$$\begin{aligned} w_t(z) &= \int_{A(x_t)} \frac{1}{\sqrt{2\pi r_t^2}} e^{-\frac{1}{2r_t^2}(z'-z)^2}\, dz' \\ &= \Phi(r_t^{-1}(\Phi^{-1}(f_{x_t}) - z)) - \Phi(r_t^{-1}(\Phi^{-1}(f_{x_t-1}) - z)). \end{aligned} \tag{3.41}$$

Then, the following approximation can be used

$$\mathbb{E}_X f(\widehat{Z}_{t+1|t}) \approx \sum_{i=1}^N \frac{w_t^i}{\Omega_{N,t}} f(\widehat{Z}_{t+1}^i) = \frac{\frac{1}{N}\sum_{i=1}^N w_t^i f(\widehat{Z}_{t+1}^i)}{\frac{1}{N}\sum_{i=1}^N w_t^i} =: \widehat{\mathbb{E}}_X f(\widehat{Z}_{t+1|t}), \tag{3.42}$$

where $\Omega_{N,t} = \sum_{i=1}^N w_t^i$. This approximation is based on the following basic result, showing by the law of large numbers that the limit of the right-hand side of (3.42) is indeed $\mathbb{E}_X f(\widehat{Z}_{t+1|t})$. The proof can be found in Appendix A.2.

Proposition 3.2.1. *With the above notation, we have*

$$\mathbb{E}_X w_t^i f(\widehat{Z}_{t+1}^i) = \mathbb{E}_X f(\widehat{Z}_{t+1|t}) \frac{\int_{z_s \in A(x_s), s=0,\dots,t} \frac{e^{-\frac{1}{2}\sum_{s=0}^t (z_s - \widehat{z}_s)^2/r_s^2}}{(2\pi)^{(t+1)/2} r_0 \dots r_t} dz_0 \dots dz_t}{\int_{z_0 \in A(x_0)} \frac{e^{-z_0^2/2}}{(2\pi)^{1/2} r_0} dz_0}. \tag{3.43}$$

Remark 3.2.2. By (3.43) and (A.10),

$$\mathbb{P}(X_0 = x_0)\mathbb{E}_X w_T^i = \mathbb{P}(X_0 = x_0, \ldots, X_T = x_T). \tag{3.44}$$

We shall use the left-hand side of this relation to approximate the model likelihood when using the SIS algorithm.

Remark 3.2.3. Depending on model parameters, the SIS algorithm may, in fact, not generate a particle path Z_t^i for some value(s) of i, that is, a "NaN" may be obtained numerically. For example, supposing that the underlying Gaussian model is AR(1) with parameter $-1 < \phi < 0$ and $\widehat{Z}_{t+1} = \phi Z_t$ (see Remark 3.2.1), this happens in the following scenario. It can happen that a value of $\epsilon_t^i = \infty$ is generated numerically in Step 2 of the SIS algorithm, leading to $Z_t^i = \infty$ and hence also $\widehat{Z}_{t+1}^i = \phi Z_t^i = -\infty$. But then at the next time $t+1$, if $x_{t+1} = 0$ and hence $x_{t+1} - 1 = -1$, we have $\Phi^{-1}(f_{x_{t+1}-1}) = \Phi^{-1}(0) = -\infty$ and the term $\Phi^{-1}(f_{x_{t+1}-1}) - \widehat{Z}_{t+1}^i$ in Step 2 of the algorithm becomes $-\infty + \infty =$ NaN. When the algorithm leads to a "NaN", one could either regard these model parameters as very unlikely or ignore these particles, e.g., in likelihood calculations. We found these two approaches to lead to comparable results in practice.

The next two remarks and a subsequent discussion draw connections between our model and the algorithm above to HMMs and particle filtering.

Remark 3.2.4. When the underlying Gaussian series Z_t is autoregressive of order p (AR(p)), the vector series $(Z_t, \ldots, Z_{t-p+1})'$ is a Markov chain on $\mathbb{R}^p$, and our model $X_t = G(Z_t)$ is an HMM. (The same also happens with ARMA($p.q$) models with an appropriate enlargement of the state space.) Indeed, here are more details on this statement when $p = 1$, in which case $Z_t = \phi Z_{t-1} + (1-\phi)^{1/2}\epsilon_t$, where $|\phi| < 1$, $\{\epsilon_t\}$ consists of i.i.d. $\mathcal{N}(0,1)$ random variables and the presence of $(1-\phi)^{1/2}$ ensures that $\mathbb{E}Z_t^2 = 1$. The resulting series $X_t = G(Z_t)$ is then an HMM in the sense of Definition 9.3 of Douc et al. [10], p. 295, with a Markov kernel on $\mathbb{R}$ given by

$$M(z, dz') = \frac{1}{(2\pi(1-\phi^2))^{1/2}} e^{-\frac{(z'-\phi z)^2}{2(1-\phi^2)}} dz' \tag{3.45}$$

governing the transition of the Markov AR(1) series Z_t, and a Markov kernel from $\mathbb{R}$ to $\mathbb{N}_0$

$$G(z, dx) = \delta_{G(z)}(dx) = \text{point mass at } G(z) \tag{3.46}$$

governing the transition from Z_t to X_t. Thus, many of the developments for HMMs (see e.g. Chapters 9–13 in Douc et al. [10]) are expected to apply directly to our model in the special case of the underlying Gaussian AR series. One important feature of our model when viewed as an HMM is that it is *not* partially dominated (in the sense described following Definition 9.3 of Douc et al. [10]). Though a number of developments described, in particular, in Douc et al. [10] apply or extend easily to partially non-dominated models (as in the next remark), there are also issues that make these models more difficult to handle (see, in particular, Remark 3.3.2 below).

Remark 3.2.5. When our model is an HMM with, for example, the underlying Gaussian AR(1) series (as in the preceding remark), the algorithm described in (3.35)–(3.42) above is the Sequential Importance Sampling (SIS) algorithm for particle filtering discussed in Section 10.2 of Douc et al. [10] with the choice of the optimal kernel and the associated weight function in Eqs. (10.30) and (10.31) of Douc et al. [10]. Indeed, this could be seen from the following observations. For AR(1) series, the one-step-ahead prediction is $\widehat{Z}_{t+1} = \phi Z_t$ (and $\widehat{z}_{t+1} = \phi z_t$). Though as noted in the preceding remark, our model as an HMM is not partially dominated and hence a transition density function $g(z, x)$ (defined following Definition 9.3 of Douc et al. [10]) is not available, a number of formulas for partially dominated HMMs given in Douc et al. [10] also apply to our model by taking

$$g(z, x) = 1_{A(x)}(z), \tag{3.47}$$

where the set $A(x)$ is defined in (3.29). This is the case, in particular, for the developments in Section 10.2 on SIS in Douc et al. [10]. For example, one could check with the choice (3.47) that the filtering distribution ϕ_t in Eq. (10.23) of Douc et al. [10] is exactly that appearing in (3.31) above. The kernel $Q_t(z, A)$ appearing in Section 10.2 of Douc et al. [10] is then

$$Q_t(z, A) = \int_A M(z, dz')g(z', x_t) = \int_{A\cap A(x_t)} \frac{1}{(2\pi(1-\phi^2))^{1/2}} e^{-\frac{(z'-\phi z)^2}{2(1-\phi^2)}} dz', \tag{3.48}$$

where we used (3.45) and (3.47). Sampling Z_t^i from the optimal kernel $Q_t(Z_{t-1}^i, \cdot)/Q_t(Z_{t-1}^i, \mathbb{R})$ (see p. 330 in Douc et al. [10]) can then be checked to be the same as defining Z_t^i through Steps 2 and 3 of our particle filtering algorithm above. The optimal weight function $Q_t(z, \mathbb{R})$ can also be checked to be that in (3.41) above.

Following particle filtering developments in the HMM literature (Sections 10.4.1 and 10.4.2

in Douc et al. [10]), our SIS filtering algorithm could also be modified to the following two algorithms. These will also be examined in our simulation study.

Sequential Importance Sampling with Resampling (SISR) particle filtering: Proceed as in SIS but modify Step 3 and add resampling Step 4 as follows:

3: Modify Step 3 of SIS by setting

$$\widetilde{Z}_t^i = \widehat{Z}_t^i + r_t \epsilon_t^i, \quad \widetilde{w}_t^i = w_{t-1}^i w_t(\widehat{Z}_t^i) \tag{3.49}$$

and also $\widetilde{\Omega}_{N,t} = \sum_{i=1}^N \widetilde{w}_t^i$.

4: Draw, conditionally independently given $\{(Z_s^i, w_s^i), s \leq t-1, i = 1, \ldots, N\}$ and $\{\widetilde{Z}_t^i, i = 1, \ldots, N\}$, a multinomial trial $\{I_t^i, i = 1, \ldots, N\}$ with probabilities of success $\{\widetilde{w}_t^i / \widetilde{\Omega}_{N,t}\}$ and set $Z_t^i = \widetilde{Z}_t^{I_t^i}$ and $w_t^i = 1$, $i = 1, \ldots, N$.

Auxiliary particle filtering (APF): Proceed as in SIS but modify Step 3 as follows:

3: Denote the distribution of Z_t^i in (3.39) as $R_t(\widehat{Z}_t^i, \cdot)$. Then, draw conditionally independently pairs $\{(I_t^i, Z_t^i), i = 1, \ldots, N\}$ of indices and particles from the distribution

$$\mu(\{i\} \times A) = \frac{w_t(\widehat{Z}_t^i)}{\sum_{i=1}^N w_t(\widehat{Z}_t^i)} R_t(\widehat{Z}_t^i, A), \tag{3.50}$$

where $w_t(z)$ is defined in (3.41). Discard the indices to take $\{Z_t^i, i = 1, \ldots, N\}$ for the particles at time t. Also, set $w_t^i = 1$ for all $i = 1, \ldots, N$.

We now turn from the filtering problem to the prediction problem, namely, that of evaluating $\mathbb{E}_X f(\widehat{X}_{t+1|t})$. This can be addressed by relating prediction to the filtering problem as in (3.33). For example, if using the SIS particle filtering, we would approximate $\mathbb{E}_X f(\widehat{X}_{t+1|t})$ as

$$\mathbb{E}_X f(\widehat{X}_{t+1|t}) \approx \sum_{i=1}^N \frac{w_t^i}{\Omega_{N,t}} g_{f,t+1}(\widehat{Z}_{t+1}^i) \tag{3.51}$$

(see (3.42)). The SISR and APF algorithms could be used as well.

3.3 Inference

The model (3.1) involves unknown parameters $\boldsymbol{\theta}$, associated with the marginal density and $\boldsymbol{\eta}$, associated with the temporal dependence structure. Several inference questions for this model are discussed here, namely, estimation of parameters and assessment of goodness-of-fit.

3.3.1 Pseudo Gaussian likelihood estimation

An is Section 3.2, suppose the observations x_t are given for times $t \in \{0, \dots T\}$ and set $\mathbf{X} = (x_0, \dots, x_T)'$. Denote the exact likelihood of the model (3.1) as

$$\mathcal{L}_T(\boldsymbol{\theta}, \boldsymbol{\eta}) = \mathbb{P}(X_0 = x_0, X_1 = x_1, \dots, X_T = x_T). \tag{3.52}$$

The likelihood has proven difficult to directly calculate for most count time series models [**?**]. In Section 3.3.2, it is approximated by using particle filtering and the resulting approximate MLE can be computationally intensive. A computationally superior but potentially statistically less efficient approach is to use a pseudo Gaussian likelihood (GL).

In the pseudo GL approach, the parameters are estimated as

$$(\widehat{\boldsymbol{\theta}}_{gl}, \widehat{\boldsymbol{\eta}}_{gl}) = \underset{\boldsymbol{\theta},\boldsymbol{\eta}}{\operatorname{argmax}} \frac{e^{-\frac{1}{2}(\mathbf{X}-\boldsymbol{\mu}_{\boldsymbol{\theta}})'\Gamma_T(\boldsymbol{\theta},\boldsymbol{\eta})^{-1}(\mathbf{X}-\boldsymbol{\mu}_{\boldsymbol{\theta}})}}{(2\pi)^{(T+1)/2}|\Gamma_T(\boldsymbol{\theta}, \boldsymbol{\eta})|^{1/2}}, \tag{3.53}$$

that is, by maximizing the likelihood assuming the model is Gaussian with the model mean $\boldsymbol{\mu}_{\boldsymbol{\theta}}$ and the model covariance matrix $\Gamma_T(\boldsymbol{\theta}, \boldsymbol{\eta}) = (\gamma_X(i-j))_{i,j\in\{0,\dots,T\}}$. It should be noted that for large T, the pseudo GL approach is equivalent to least squares, where the regression is carried out (e.g. through the Durbin-Levinson algorithm) for X_t on the previous values $X_{t-1}, \dots, X_0$ (see e.g. Section 5.2 in [6]). Two other points related to the pseudo GL approach should be emphasized. First, the covariance function for the model (3.1) could be evaluated efficiently as described in Section 2. The model mean $\boldsymbol{\mu}_{\boldsymbol{\theta}}$, on the other hand, can often be expressed explicitly for the marginal distributions of interest. Second, the numerical optimization of (3.53) yields the numerical Hessian that can be used to set confidence intervals for the parameters of the model.

3.3.2 MLE through particle approximation

By using the notation and result (3.33) appearing in Lemma 3.2.1, the likelihood (3.52) can be expressed as

$$\begin{aligned}\mathcal{L}_T(\boldsymbol{\theta},\boldsymbol{\eta}) &= \mathbb{P}(X_0 = x_0)\prod_{s=1}^{T}\mathbb{P}(X_s = x_s|X_0 = x_0,\ldots,X_{s-1} = x_{s-1}) \\ &= \mathbb{P}(X_0 = x_0)\prod_{s=1}^{T}\mathbb{E}_X(1_{\{x_s\}}(\widehat{X}_{s|s-1})) \\ &= \mathbb{P}(X_0 = x_0)\prod_{s=1}^{T}\mathbb{E}_X(g_{1_{\{x_s\}},s}(\widehat{Z}_{s|s-1})). \end{aligned} \tag{3.54}$$

Note from (3.34) that

$$g_{1_{\{x_s\}},s}(z) = w_s(z), \tag{3.55}$$

where $w_s(z)$ is defined and can be computed numerically as in (3.41). The particle approximation of the likelihood is then defined as

$$\widehat{\mathcal{L}}_T(\boldsymbol{\theta},\boldsymbol{\eta}) = \mathbb{P}(X_0 = x_0)\prod_{s=1}^{T}\widehat{\mathbb{E}}_X(w_s(\widehat{Z}_{s|s-1})), \tag{3.56}$$

by using the notation in (3.42) and supposing that particles are generated by one of the methods (SIS, SISR or APF) discussed in Section 3.2. The ML estimators through particle approximations are then defined as

$$(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\eta}}) = \underset{\boldsymbol{\theta},\boldsymbol{\eta}}{\operatorname{argmax}}\,\widehat{\mathcal{L}}_T(\boldsymbol{\theta},\boldsymbol{\eta}). \tag{3.57}$$

Remark 3.3.1. For the SIS algorithm, the expression (3.56) reduces to

$$\widehat{\mathcal{L}}_T(\boldsymbol{\theta},\boldsymbol{\eta}) = \mathbb{P}(X_0 = x_0)\frac{1}{N}\sum_{i=1}^{N} w_T^i, \tag{3.58}$$

which is consistent with what was noted in Remark 3.2.2.

Remark 3.3.2. In the HMM literature, superior algorithms to the likelihood approximation (3.56) are available, based on e.g. versions of the EM algorithm (see Section 12 in Douc et al. [10]). These algorithms do not apply directly to our context even when our model is an HMM. Indeed, as noted in Remark 3.2.4, our model is not partially dominated. As a consequence, even the most basic

developments of Douc et al. [10] do not carry over to our model. For example, for the underlying AR(1) model Z_t, the complete likelihood for our model is

$$\mathbb{P}(Z_0 = z_0, \ldots, Z_T = z_T, X_0 = x_0, \ldots, X_T = x_T)$$

$$= \frac{e^{-z_0^2/2}}{(2\pi)^{1/2}} \frac{e^{-\frac{1}{2}\sum_{t=1}^{T}(z_t - \phi z_{t-1})^2}}{(2\pi(1-\phi^2))^{T/2}} \prod_{t=0}^{T} 1_{\{z_t \in A(x_t)\}}.$$

For example, taking the logarithm of this likelihood is not quite possible in order to get the form of the complete likelihood appearing in Eq. (12.9) of Douc et al. [10], on which their subsequent EM algorithms are based.

When using MLE through particle approximation, confidence intervals for model parameters will be set through a block bootstrap. In this regard, note that under the model (3.1), a block bootstrap of the series X_t corresponds to that of the underlying Gaussian series Z_t, which is quite well understood.

3.3.3 Model diagnostics

The goodness-of-fit of count models is commonly assessed through probability integral transform (PIT) histograms and related tools (e.g. Czado et al. [9], Kolassa [18]). These are based on the predictive distributions of $\{X_t\}$, defined for time t as

$$P_t(y) = \mathbb{P}_X(\widehat{X}_{t|t-1} \leq y) = \mathbb{P}(X_t \leq y | X_0 = x_0, \ldots, X_{t-1} = x_{t-1}), \quad y \in \{0, 1, \ldots\}. \tag{3.59}$$

This quantity can be estimated through the particle methods discussed in Section 3.2, namely, as

$$\widehat{P}_t(y) = \sum_{x=0}^{y} \widehat{\mathbb{E}}_X 1_{\{x\}}(\widehat{X}_{t|t-1}) = \sum_{x=0}^{y} \widehat{\mathbb{E}}_X g_{1_{\{x\}},t}(\widehat{Z}_{t|t-1}), \tag{3.60}$$

by using the notation $g_{f,t}$ in (3.34) and $\widehat{\mathbb{E}}_X$ in (3.42), supposing the particles are generated by the SIS, SISR or APF method. Similarly to (3.55), note also that

$$g_{1_{\{x\}},t}(z) = \widetilde{w}_{x,t}(z), \tag{3.61}$$

where

$$\widetilde{w}_{x,t}(z) = \Phi(r_t^{-1}(\Phi^{-1}(f_x) - z)) - \Phi(r_t^{-1}(\Phi^{-1}(f_{x-1}) - z)) \tag{3.62}$$

(and $\widetilde{w}_{x_t,t}(z) = w_t(z)$).

The (non-randomized) mean PIT is defined as

$$\overline{F}(u) = \frac{1}{T+1} \sum_{t=0}^{T} F_t(u|x_t), \quad u \in [0,1], \tag{3.63}$$

where

$$F_t(u|y) = \begin{cases} 0, & \text{if } u \le P_t(y-1), \\ \frac{u - P_t(y-1)}{P_t(y) - P_t(y-1)}, & \text{if } P_t(y-1) < u < P_t(y), \\ 1, & \text{if } u \ge P_t(y), \end{cases} \tag{3.64}$$

which is estimated by replacing P_t by $\widehat{P}_t$ in practice. The PIT histogram with H bins is defined as a histogram with the height $\overline{F}(h/H) - \overline{F}((h-1)/H)$ for bin $h \in \{1, \ldots, H\}$.

As a more elementary diagnostics tool, another possibility is to consider the model residuals defined as follows. Let

$$\widehat{Z}_t = \mathbb{E}(Z_t|X_t = x_t) = \frac{1}{\sqrt{2\pi}} \frac{\exp(-\Phi^{-1}(C_{k-1})^2/2) - \exp(-\Phi^{-1}(C_k)^2/2)}{C_k - C_{k-1}} \tag{3.65}$$

be the estimated mean value of the latent Gaussian process at time t given the observation x_t, where the formula (3.65) follows by direct calculations for the model (3.1) (assuming the estimated parameter values $\boldsymbol{\theta}$ of the marginal distribution entering C_k's). For a fitted underlying time series model with parameter values $\boldsymbol{\eta}$, the residuals are then defined as the residuals $\widehat{\epsilon}_t$ of this model fitted to the series $\widehat{Z}_t$, after centering it by the sample mean. In more formal terms (omitting the sample mean for simplicity),

$$\widehat{\epsilon}_t = \widehat{Z}_t - \mathbb{E}_{\boldsymbol{\eta}}(\widehat{Z}_t|\widehat{Z}_{t-1}, \ldots, \widehat{Z}_0),$$

where $\mathbb{E}_{\boldsymbol{\eta}}$ indicates the linear prediction under the fitted time series model with parameter values $\boldsymbol{\eta}$.

3.3.4 Inclusion of covariates

Covariates can be included in the model (3.1) through the parameter $\boldsymbol{\theta}$ of the marginal distribution, as discussed in Section 3.1.2. With the covariates, the parameter $\boldsymbol{\theta}$ depends on time t

as $\boldsymbol{\theta}(t)$. The GL and particle inference procedures are modified for $\boldsymbol{\theta}(t)$ in the following ways.

For the GL procedure, the covariance $\mathrm{Cov}(X_{t_1}, X_{t_2}) = \mathrm{Cov}(G_{\boldsymbol{\theta}(t_1)}(Z_{t_1}), G_{\boldsymbol{\theta}(t_2)}(Z_{t_2}))$ is needed, where $\boldsymbol{\theta}(t)$ is added as subscript in G to indicate the dependence on t. But as in (3.9), one has

$$\mathrm{Cov}(X_{t_1}, X_{t_2}) = \mathrm{Cov}(G_{\boldsymbol{\theta}(t_1)}(Z_{t_1}), G_{\boldsymbol{\theta}(t_2)}(Z_{t_2})) = \sum_{k=1}^{\infty} k! g_{\boldsymbol{\theta}(t_1),k} g_{\boldsymbol{\theta}(t_2),k} \gamma_Z(t_1 - t_2)^k, \tag{3.66}$$

where again, the subscript $\boldsymbol{\theta}(t)$ is added to g_k to indicate dependence on t. Numerically, evaluating (3.66) is not much different from that of (3.9), since both are based on calculating the Hermite coefficients g_k.

For the particle filtering approach, the modification is somewhat simpler: one just needs to replace $\boldsymbol{\theta}$ by $\boldsymbol{\theta}(t)$ at time t when generating the underlying particles. For example, for the SIS algorithm in Section 3.2, $\boldsymbol{\theta}(t)$ would enter through f_x in (3.36), (3.38), (3.41). This modification is justified from the structure of the model, where the covariates enter only through the parameter $\boldsymbol{\theta}$ controlling the marginal distribution.

3.4 A Simulation Study

To study the developed methods, a simulation study considering multiple marginal distributions and ACVFs was conducted. Here, the classic Poisson and negative binomial count distributions with ARMA $\{Z_t\}$ are considered.

3.4.1 Poisson AR(1)

We begin with the simple case where X_t has a Poisson marginal distribution for each t with mean $\lambda > 0$. An AR(1) $\{Z_t\}$ satisfying

$$Z_t = \phi Z_{t-1} + (1 - \phi)^{1/2} \epsilon_t \tag{3.67}$$

is considered, where $\{\epsilon_t\}$ is i.i.d. standard normal as in Remark 3.2.4. Series of lengths $T = 100, 200,$ and 400 are simulated from (3.67) and then transformed via (3.1) to obtain the Poisson count series. The Gaussian likelihood (GL) and particle filtering (PF) likelihood estimation methods of Section 3.3 are applied. All simulations were replicated 200 times to assess bias and variability of the estimators. All combinations of $\lambda \in \{2, 5, 10\}$ and $\phi \in \{\pm 0.25, \pm 0.75\}$ were considered.

Box plots of the resulting parameter estimates for the models with $\lambda = 2$ and $\phi = \pm 0.75$ are displayed in Figure 3.3. Observe that the GL estimates of ϕ tend to be slightly less than ϕ (biased towards zero), while having similar variance than the approximately unbiased PF estimates. When the lag-one correlation in $\{Z_t\}$ (and hence also that in $\{X_t\}$) is negative, both estimation methods have smaller variances than their positively correlated counterparts. This is expected: the mean of this process is λ, and the variability of the sample mean, one good estimator of the mean, for stationary series is comparatively smaller for negatively correlated series than for positively correlated series. Smaller λ values yield estimates with smaller variabilities. Again this is expected: the variance of the Poisson distribution is also λ. In estimating ϕ, the GL is again biased toward zero, while the PF estimates show little, if any, bias; however, the variance of the GL and PF estimators are now similar. In this case, both estimation methods have yielded reasonable results. All other simulation runs mimicked these results and the resulting box plots are omitted here.

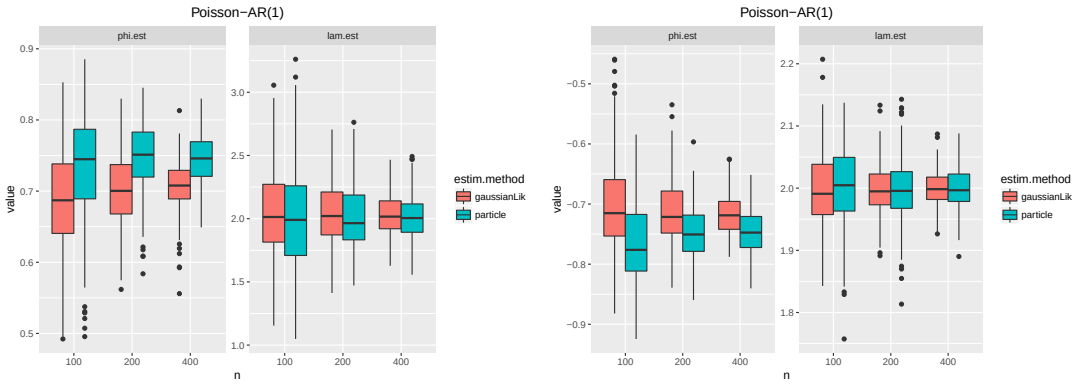

Figure 3.3: Estimates from simulated Poisson AR(1) series. In the left plots, the true value of parameters are $\lambda = 2$ and $\phi = 0.75$. In the right plots, the true value of parameters are $\lambda = 2$ and $\phi = -0.75$. All true parameter values are plotted as a horizontal red dashed line. Sample sizes of 100, 200, and 400 are indicated on the horizontal axis and estimation method are given in the legend.

3.4.2 Mixed Poisson AR(1)

This case reveals differences between our estimation methods. Here, the three-parameter mixture Poisson marginal distribution with parameters $\lambda_1 > 0, \lambda_2 > 0$, probability $p \in [0, 1]$, satisfying

$$P(X_t = k) = p\frac{e^{-\lambda_1}\lambda_1^k}{k!} + (1-p)\frac{e^{-\lambda_2}\lambda_2^k}{k!} \tag{3.68}$$

is considered. Again, 200 Gaussian series of lengths $T = 100, 200$ and 400 are simulated and transformed to a mixed Poisson AR(1) count time series that has marginal distributions as in (3.68). As in the last case, an AR(1) $\{Z_t\}$ is used with $\phi = \pm 0.25, \pm 0.75$ and the Gaussian likelihood and particle filtering estimation methods are applied.

Observe that Gaussian likelihood only uses the theoretical mean of the distribution of X_t and the covariance structure of $\{X_t\}$ to compute parameter estimates. Here, the probability that X_t is close to its mean value of $p\lambda_1 + (1-p)\lambda_2$ is small when λ_1 and λ_2 are far apart. Hence, one might expect Gaussian likelihood to do poorly. In contrast, the particle filtering likelihood approach should feel the entire joint distribution of the process, basing estimates on more than the first and second moments of the series.

Figure 3.4 shows box plots of 200 estimates for the model with $\lambda_1 = 2$, $\lambda_2 \in \{5, 10\}$, $p = 1/4$, and $\phi = 0.75$. To maintain parameter identifiability p was constrained to the interval $[0, 0.5]$. The results are as expected, with particle filtering likelihoods providing superior performance. As expected, the superiority of particle filtering is more pronounced as the difference between λ_1 and λ_2 gets larger.

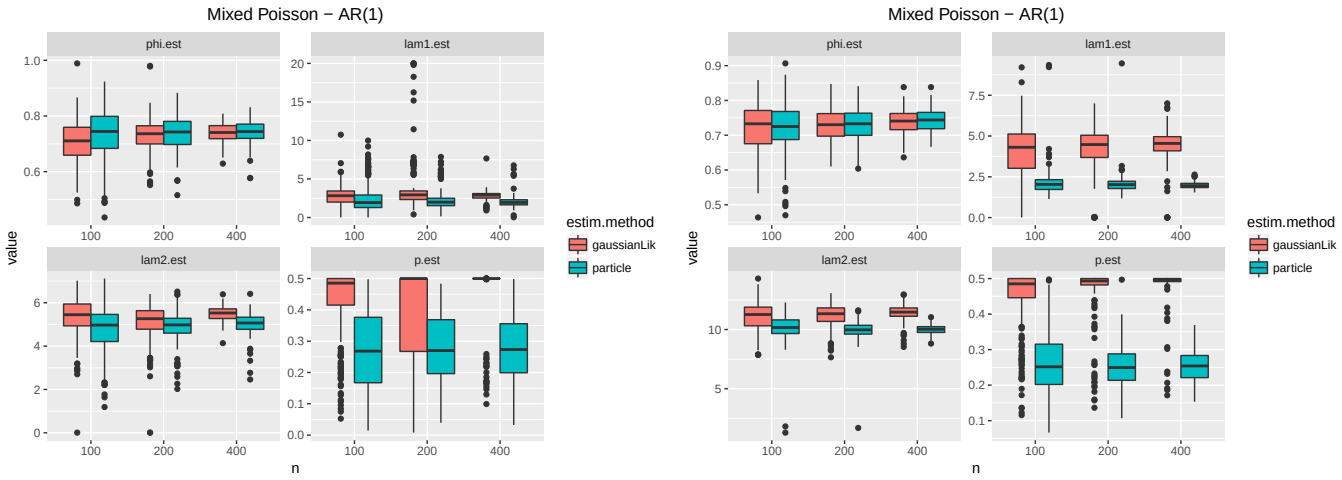

Figure 3.4: Estimates from simulated Mixed Poisson AR(1) series. In the left plots, the true parameter values are $\lambda_1 = 2$, $\lambda_2 = 5$, $\phi = 0.75$ and $p = 1/4$. In the right plots, the true parameter values are $\lambda_1 = 2$, $\lambda_2 = 10$, $\phi = 0.75$ and $p = 1/4$. True values are shown as a red horizontal dashed line. Sample size of 100, 200 and 400 are given on the horizontal axis and estimation method are indicated in the legend.

3.4.3 Negative Binomial MA(1)

The prior subsections were based on the unbquitious equidispersed Poisson distribution. Here, the negative Binomial distribution (NB) with parameters $r > 0$ and $p \in (0, 1)$, is considered. The marginal law of X_t is

$$P(X_t = k) = \frac{\Gamma(k + r)}{\Gamma(r)k!} p^r (1 - p)^k, \quad k \in \{0, 1, \ldots\}. \tag{3.69}$$

For the ACVF, a moving average (MA) of order one is utilized. Series of lengths $T = 100, 200$ and 400 from the difference equation

$$Z_t = \epsilon_t + \theta \epsilon_{t-1}, \tag{3.70}$$

where $\{\epsilon_t\}$ is an i.i.d. zero mean Gaussian series with the variance set to $(1 + \theta^2)^{-1}$.

The parameter values $\theta \in \{\pm 0.25, \pm 0.75\}$ and $p \in \{0.2, 0.5\}$ were considered. Again, simulation procedures were run 200 times for each combination of parameters. Figure 3.5 shows box plots of our results for the models with $\theta = \pm 0.75$, $r = 3$, and $p = 0.2$. The results again seem reasonable, with the possible exception of a large variance in estimation of θ for the smaller sample size $T = 100$ for GL. When the sample size increases to $T = 400$, Gaussian likelihood boundary issues dissipate and sampling variability becomes appreciably smaller.

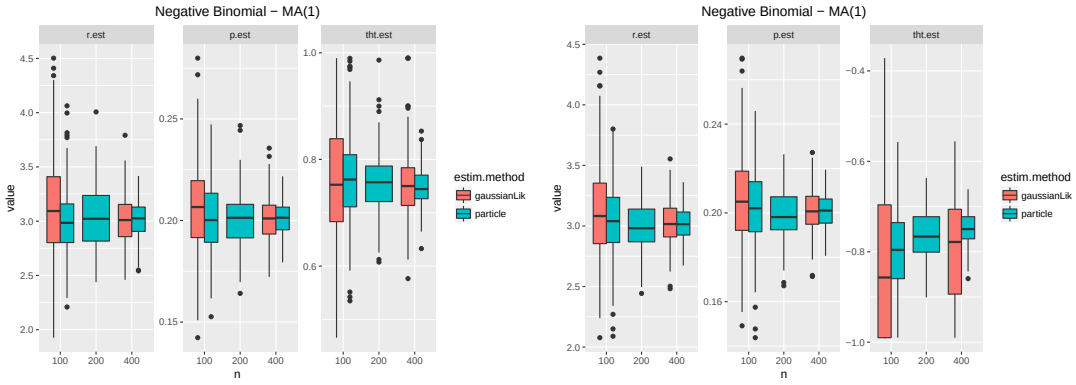

Figure 3.5: Estimates from simulated Negative Binomial MA(1) series. In the left plots, the true value are $r = 3$, $p = 0.2$ and $\theta = 0.75$. In the right plots, the true value are $r = 3$, $p = 0.2$ and $\theta = -0.75$. True values are shown as red horizontal dashed line. Sample sizes of 100, 200 and 400 are given on the horizontal axis and estimation method are indicated in the legend.

3.5 An Application

This section analyzes a count series containing the number of major league baseball games where no hitters were pitched on an annual basis from 1893 — 2017 ($T = 125$). The data are over-dispersed, with a sample mean of 2.12 and a sample variance of 3.40. The counts and its sample autocorrelations and partial autocorrelations are plotted in Figure 3.6. Two covariates are available

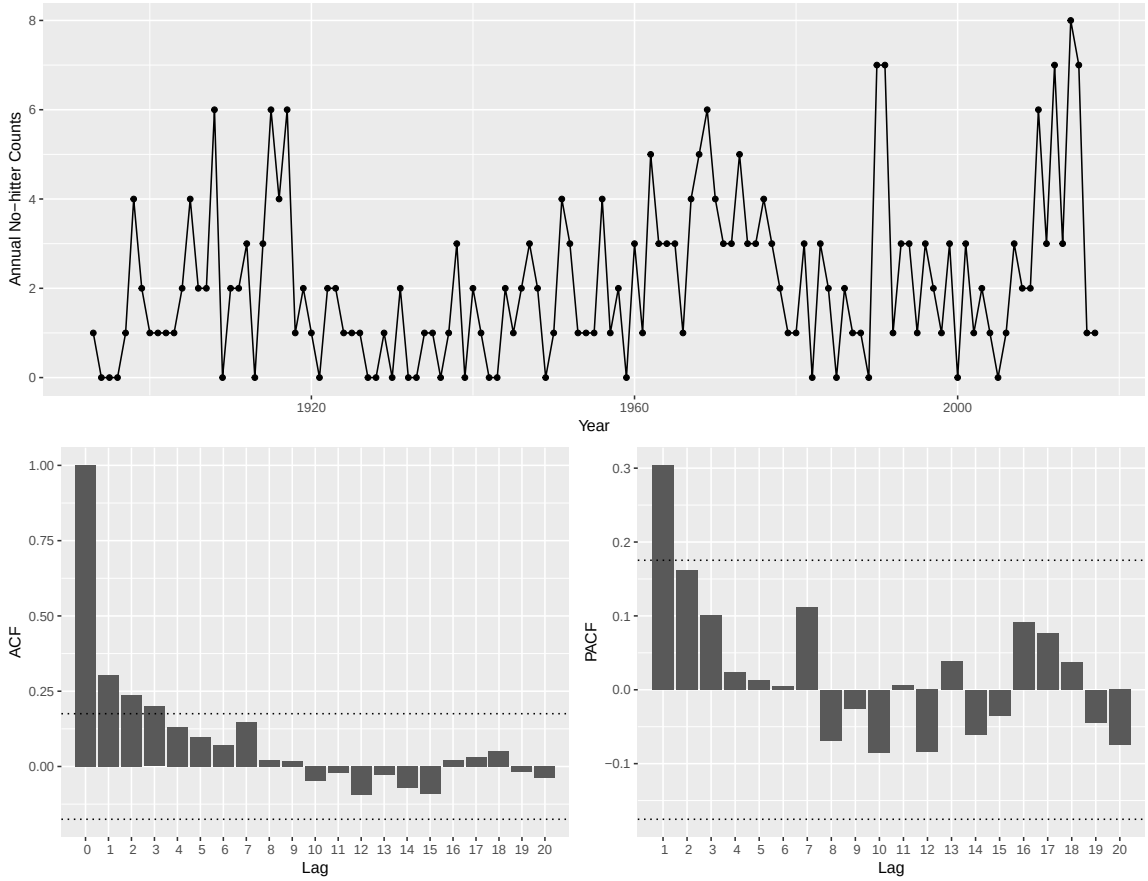

Figure 3.6: The number of no-hitters pitched by season from 1893 to 2017 and its sample autocorrelation and partial autocorrelations. Pointwise 95% confidence intervals are displayed.

to help describe the counts: 1) the total number of games played in each year (denoted as M_1) and the height of the pitching mound (denoted as M_2). The total number of games played in a season has increased over the years as more teams have entered the league over time. Also, baseball seasons have lengthed, with a team currently playing 162 games each season. Of course, one expects more no hitters when more games are played. The height of the pitching mound has varied over the years and could also be a significant factor. Higher pitching mounds tend to give pitchers more of an

advantage; this said, hitters tend to adjust to changes over time. In what follows, two over-dispersed count marginal distributions, the negative binomial and generalized Poisson distributes, are fitted to the data.

The overdispersed generalized Poisson distribution with parameters $\eta \in (0,1)$ fixed in time and λ_t varying with time will be one considered model for the counts. We use

$$\lambda_t = \exp\left(\beta_0 + \beta_1 M_{1,t} + \beta_2 M_{2,t}\right), \tag{3.71}$$

where a log link is used to keep λ_t positive. One can also let η depend on the covariates, but we will not need to do this as evidenced by the residual diagnostics below.

The negative binomial distribution with parameters $r > 0$ and $p \in (0,1)$ will also be considered for this count series. In this fit, we again use the log link

$$r_t = \exp\left(\beta_0 + \beta_1 M_{1,t} + \beta_2 M_{2,t}\right)$$

and p is kept fixed in time t.

We now explore some low order ARMA model fits for $\{Z_t\}$, specifically the AR(1), AR(2), ARMA(1,1), MA(1), and MA(2) models. The classical AIC and/or BIC model order selection statistics will be used to select the best fitting model. Table 3.1 shows Gaussian likelihood (GL) and particle filtering (PF) estimates for each ARMA model and both the generalized Poisson and negative binomial distributions.

Marginal Distribution	Model	AR(1)	AR(2)	ARMA(1,1)	MA(1)	MA(2)
Generalized Poisson	$-2\log(\hat{L}_{GL})$	477.5984	476.2281	476.3462	479.1375	478.1373
	AIC	487.5984	488.2281	488.3462	489.1375	490.1373
	BIC	488.0829	488.8029	488.9277	489.6221	490.7188
Negative Binomial	$-2\log(\hat{L}_{GL})$	480.4010	477.1121	480.3373	481.2713	480.9248
	AIC	490.4010	489.1121	492.3373	491.2713	492.9248
	BIC	491.3012	489.6936	492.9187	491.7559	493.5063

Table 3.1: Optimized log likelihood with the AIC/BIC for different latent Gaussian structures.

The standard errors quoted above were obtained in the usual way from the Hessian matrix at the estimated model parameters. Both the AIC and BIC statistics prefer the generalized Poisson fit with AR(1) $\{Z_t\}$. This conclusion holds for both Gaussian likelihood and particle filtering estimates.

In the AR(1) fit shown in table 3.2, the standard errors indicate that the β_0 and β_2 parameters are not significantly different from zero; the parameter β_1 appears significantly positive. We remark that asymptotic normality of the parameter estimators has not been proven (this would take us far from our salient points), but there is no reason not to expect it. The implication here is that the number of no-hitters pitched increase with the number of games played, but past changes in the mound height have not influenced the counts greatly.

Parameters	ϕ	β_0	β_1	β_2	η
GL Estimates	0.2665	-1.1496	0.7583	0.0338	0.1679
GL Standard Errors	0.0658	0.9069	0.2173	0.0436	0.0480

Table 3.2: Estimates and standard errors of the full model with $\{Z_t\}$ being AR(1).

As a tuning step, the AR(1) model for $\{Z_t\}$ was refit assuming that $\beta_0 = \beta_2 = 0$. Table 3.3 shows estimates of the model parameters and their standard errors. It is perhaps worth commenting that the above application entails a negative binomial and generalized Poisson regression with correlated errors.

Parameters	ϕ	β_1	η
GL Estimates	0.2456	0.4059	0.1212
GL Standard Errors	0.0621	0.0480	0.0416

Table 3.3: Estimates and standard errors of the reduced model with $\{Z_t\}$ being AR(1).

Let us now consider residual diagnostics to assess the AR(1) generalized Poisson model fit. Residuals can be obtained by estimating the latent $\{Z_t\}$ process from the observed $\{X_t\}$ and the estimated model parameters. Since a specific count X_t can be backtracked to an interval of Z_t values, an estimated value of the latent Gaussian process at time t is defined as the average value

$$\hat{Z}_t = \mathbb{E}(Z_t|X_t = k) = \frac{1}{\sqrt{2\pi}} \frac{\exp(-\Phi^{-1}(C_{k-1})^2/2) - \exp(-\Phi^{-1}(C_k)^2/2)}{C_k - C_{k-1}}. \tag{3.72}$$

The estimated time t ARMA residual is then simply $\hat{Z}_t - P(\hat{Z}_t|\hat{Z}_1, \ldots, \hat{Z}_{t-1})$, where $P(X|A_1, \ldots, A_k)$ denotes the best (minimal mean squared error) prediction of X from linear combinations of $A_1, \ldots, A_k$. These are computed from the fitted ARMA model coefficients in the usual time series manner. In the AR(1) case, one has $P(\hat{Z}_t|\hat{Z}_1, \ldots, \hat{Z}_{t-1}) = \hat{\phi}_1 \hat{Z}_{t-1}$ for $t \geq 2$ and $P(\hat{Z}_1|\cdot) = 0$.

Figure 3.7 shows an analysis of the AR(1) residuals from the generalized Poisson model with $\beta_0 = \beta_2 = 0$. The sample autocorrelations and partial autocorrelation do not show any departures

from a white noise model. It is stressed that normality of these residuals is needed. A QQ plot of the residuals is presented in the northeast plot of the figure and suggests a good fit, some misfit in the lower quantiles aside. The p-value for the Shapiro-Wilks test for normality is 0.4012, which is quite reasonable. Overall, it would seem that the model fits the data well.

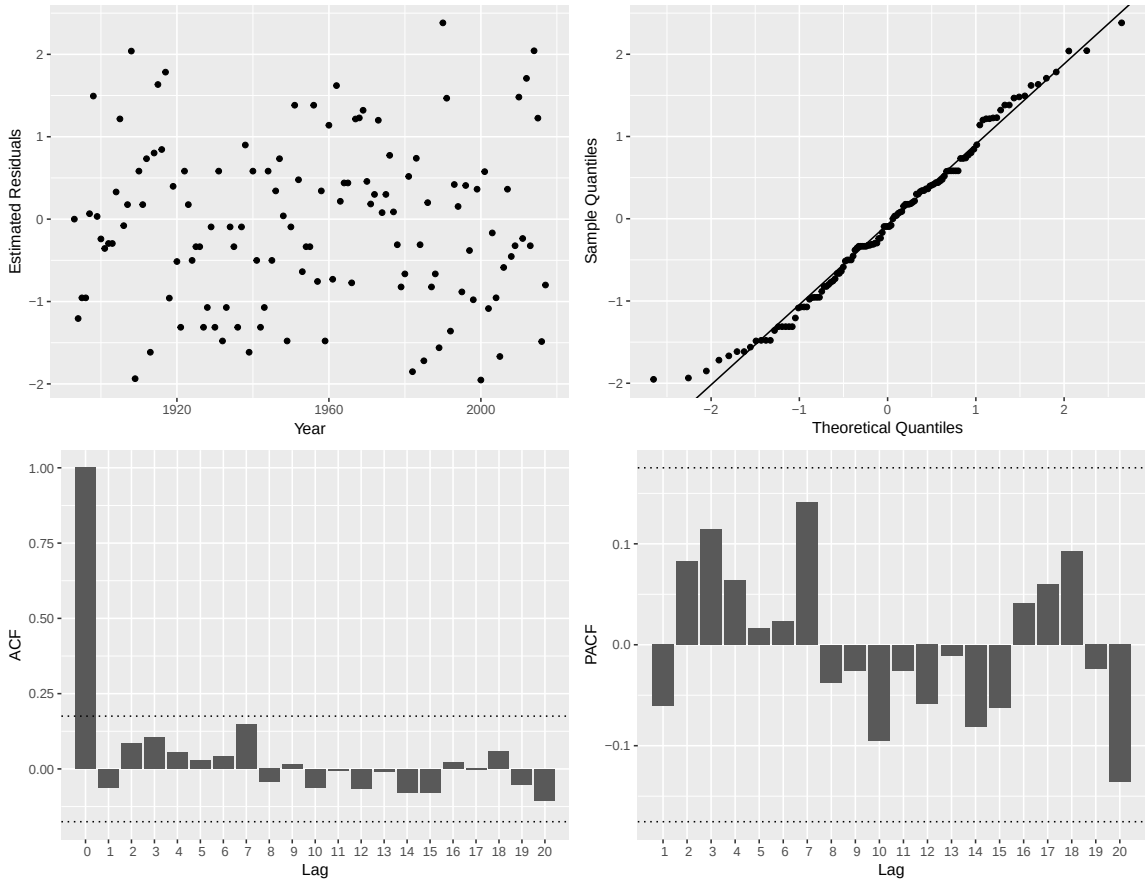

Figure 3.7: The upper left is the plot of the estimated residuals against time. The upper right is the QQ plot the estimated residuals. The two plots on the bottom are the sample ACF and sample PACF plot of the estimated residuals.

Appendix A

Appendices

A.1 Proof of Theorem 2.2.1 in Chapter 2

Equation (2.4) gives

$$\mathbb{E}(z^{X_t}) = \mathbb{E}\left[\mathbb{E}\left(z^{\sum_{i=1}^{M_t} B_{t,i}}|M_t\right)\right] = \mathbb{E}\left[(1-p_B+p_Bu)^{M_t}\right] = \psi_M(1-p_B+p_Bz).$$

Substituting $u = 1 - p_B + p_Bz$ yields the result in part a). For part b), simple computations show that $\mathbb{E}[X_t] = p_B\mu_M$ and $\text{var}(X_t) = p_B^2(\sigma_M^2 - \mu_M) + \mu_M p_B$. Hence,

$$D_X = \frac{p_B^2\sigma_M^2 + \mu_M p_B(1-p_B)}{\mu_M p_B} = p_B D_M + 1 - p_B$$

as claimed.

To prove part c), apply the law of total covariance to get

$$\begin{aligned}\text{cov}(X_t, X_{t+h}) &= \text{cov}\left(\sum_{i=1}^{M_t} B_{t,i}, \sum_{j=1}^{M_{t+h}} B_{t+h,j}\right) \\ &= \mathbb{E}\left[\text{cov}\left(\sum_{i=1}^{M_t} B_{t,i}, \sum_{j=1}^{M_{t+h}} B_{t+h,j}\middle| M_t, M_{t+h}\right)\right] + \text{cov}\left(\mathbb{E}\left[\sum_{i=1}^{M_t} B_{t,i}\middle| M_t, M_{t+h}\right], \mathbb{E}\left[\sum_{j=1}^{M_{t+h}} B_{t+h,j}\middle| M_t, M_{t+h}\right]\right)\end{aligned} \tag{A.1}$$

For any positive integers m_t, m_{t+h}, $\text{cov}\left(\sum_{i=1}^{m_t} B_{t,i}, \sum_{j=1}^{m_{t+h}} B_{t+h,j}\right) = \min(m_t, m_{t+h})\gamma_B(h)$ and $\mathbb{E}\left[\sum_{i=1}^{m_t} B_{t,i}\right] = m_t p_B$. Plugging this result in (A.1) gives

$$\text{cov}(X_t, X_{t+h}) = \mathrm{E}\left[\min(M_t, M_{t+h})\right]\gamma_B(h) + p_B^2 \text{cov}(M_t, M_{t+h}).$$

When $h \neq 0$, M_t and M_{t+h} are independent and $\text{cov}(M_t, M_{t+h}) = 0$, implying that $\gamma_X(h) = \kappa\gamma_B(h)$, where $\kappa = \mathbb{E}\left[\min(M_t, M_{t+h})\right]$. When $h = 0$, extracting the first two moments from the probability generating function gives $\gamma_X(0) = \gamma_B(0)\mu_M + p_B^2\sigma_M^2$ as claimed.

For part d), when $h \neq 0$, conditioning on M_t and M_{t+h} gives

$$\mathbb{P}(X_t = x_t, X_{t+h} = x_{t+h}) = \sum_{m_t=0}^{\infty}\sum_{m_{t+h}=0}^{\infty} H_{m_t,m_{t+h}}(x_t, x_{t+h}) f_M(m_t) f_M(m_{t+h}),$$

where $H_{m_t,m_{t+h}}(x_t, x_{t+h}) := \mathbb{P}\left(S_t = x_t, S_{t+h} = x_{t+h}\right)$ and $S_t = \sum_{i=1}^{m_t} B_{t,i}$, with m_t fixed. This joint probability can be calculated by conditioning on the value of X_t:

$$H_{m_t,m_{t+h}}(x_t, x_{t+h}) = \mathbb{P}\left(S_{t+h} = x_{t+h} | S_t = x_t\right)\mathbb{P}(S_t = x_t). \tag{A.2}$$

It is easy to see that S_t has a binomial distribution with m_t trials and success probability p_B. The conditional probability above changes form with two cases.

When $m_t < m_{t+h}$, the conditional probability in (A.2) represents the sum of three independent binomial distributions: one with x_t trials and success probability $p_{1,1}$, one with $m_t - x_t$ trials and success probability $p_{0,1}$, and one with $m_{t+h} - m_t$ trials and success probability p_B. Hence,

$$\mathbb{P}\left(S_{t+h} = x_{t+h} | S_t = x_t\right) = \sum_{b=0}^{\min(m_t - x_t, x_{t+h})}\sum_{a=0}^{\min(x_t, x_{t+h}-b)} T_1 T_2 T_3,$$

where $T_1 = B(x_t, a, p_{1,1}), T_2 = B(m_t - x_t, b, p_{0,1})$, and $T_3 = B(m_{t+h} - m_t, x_{t+h} - a - b, p_B)$. Here, $B(n, k, p) := \binom{n}{k}p^k(1-p)^{n-k}$ denotes the binomial probability mass function.

The case where $m_t \geq m_{t+h}$ is more complicated. Further conditioning on $\sum_{i=1}^{m_{t+h}} B_{t,i}$ (this

is neither S_t nor S_{t+h}) gives

$$\mathbb{P}\left(S_{t+h}=x_{t+h}|S_t=x_t\right)=\sum_{k=0}^{\infty}\mathbb{P}\left(S_{t+h}=x_{t+h}|S_t=x_t,\sum_{i=1}^{m_{t+h}}B_{t,i}=k\right)\mathbb{P}\left(\sum_{i=1}^{m_{t+h}}B_{t,i}=k\middle|S_t=x_t\right). \tag{A.3}$$

Conditional on $S_t=x_t$, $\sum_{i=1}^{m_{t+h}}B_{t,i}$ has a hypergeometric distribution with m_{t+h} draws from a population of containing x_t Type I items and m_t-x_t Type II items:

$$\mathbb{P}\left(\sum_{i=1}^{m_{t+h}}B_{t,i}=k\middle|S_t=x_t\right)=\frac{\binom{x_t}{k}\binom{m_t-x_t}{m_{t+h}-k}}{\binom{m_t}{m_{t+h}}},$$

for $k\in\{\max(0,m_{t+h}-m_t+x_t),\cdots,\min(x_t,m_{t+h})\}$.

The distribution of S_{t+h} conditioned on $S_t=x_t$ and $\sum_{i=1}^{m_{t+h}}B_{t,i}=k$ has a form that is the sum of two independent binomial distributions: one with k trials and success probability $p_{1,1}$, and the other with $m_{t+h}-k$ trials and success probability $p_{0,1}$. This gives

$$\mathbb{P}\left(S_{t+h}=x_{t+h}\middle|S_t=x_t,\sum_{i=1}^{m_{t+h}}B_{t,i}=k\right)=\sum_{a=0}^{\min(k,x_{t+h})}B(m_{t+h}-k,x_{t+h}-a,p_{0,1})B(k,a,p_{1,1}).$$

This identifies the two terms in (A.3). Plugging these back into (A.2) identifies $H_{m_t,m_{t+h}}(x_t,x_{t+h})$.

To prove part e), the renewal case is established in [20]. In the clipped Gaussian case, if $\rho_Z(h)\to 0$ as $h\to\infty$, then the result follows from (2.3) and the limit comparison test since $\lim_{x\to 0}\sin^{-1}(x)/x=1$. Should $\rho_Z(h)\nrightarrow 0$, then $\{Z_t\}$ must have long memory and there are an infinite number of lags h where $|\rho_Z(h)|>\delta$ for some $\delta>0$. For these lags, we also must have $|\sin^{-1}(\rho_Z(h))|>\delta^*$ for some $\delta^*>0$ by properties of the inverse sin function. It follows from (2.3) that $\{X_t\}$ must also have long memory. This proves part e).

Proof of (2.6): To rig up a type of regeneration epoch, condition on the minimum of $A_{t,1}$ and $A_{t+h,1}$ to get

$$\psi_{i,j}(h)=\sum_{\ell=1}^{\infty}\mathbb{E}\left[M_t^{(i)}M_{t+h}^{(j)}\middle|\min(A_{t,1},A_{t+h,1})=\ell\right]\mathbb{P}(\min(A_{t,1},A_{t+h,1})=\ell) \tag{A.4}$$

When $A_{t,1} = A_{t+h,1} = \ell$, due to the memoryless property of the geometric distribution, $M_t^{(i)}$ is distributionally equivalent to $M_t^{(i-1)} + \ell$ and $M_{t+h}^{(j)}$ is equal in distribution to $M_{t+h}^{(j-1)} + \ell$. Similarly, when $\ell = A_{t,1} < A_{t+h,1}$, $M_t^{(i)}$ is equal in distribution to $M_t^{(i-1)} + \ell$ and $M_{t+h}^{(j)}$ is equal to $M_{t+h}^{(j)} + \ell$ in distribution. When $M_{t,1} > M_{t+h,1} = \ell$, $M_t^{(i)}$ equal to $M_t^{(i)} + \ell$ in distribution and $M_{t+h}^{(j)}$ is equal to $M_{t+h}^{(j-1)} + \ell$ in distribution. Using these in (A.4) and simplifying gives

$$\begin{aligned}\psi_{i,j}(h) = \psi_{i-1,j-1}(h)p_1 + \psi_{i,j-1}(h)p_2 + \psi_{i-1,j}(h)p_3 + \frac{i+j-1-p_1}{p_B}\mathbb{E}\left[\min(A_{t,1}, A_{t+h,1})\right] \\ + \operatorname{var}(\min(A_{t,1}, A_{t+h,1})) + \mathbb{E}^2\left[\min(A_{t,1}, A_{t+h,1})\right]\end{aligned} \tag{A.5}$$

where $i, j \in \{1, 2, \ldots\}$. Here, the p_is are

$$\begin{aligned}p_1 &= \mathbb{P}(A_{t,1} = A_{t+h,1} = \ell | \min(A_{t,1}, A_{t+h,1}) = \ell) = \frac{p_B p_{1,1}}{1-(1-p_B)p_{0,0}}, \\ p_2 &= \mathbb{P}(A_{t,1} > A_{t+h,1} = \ell | \min(A_{t,1}, A_{t+h,1}) = \ell) = \frac{(1-p_B)p_{0,1}}{1-(1-p_B)p_{0,0}}, \\ p_3 &= \mathbb{P}(\ell = A_{t,1} < A_{t+h,1} | \min(A_{t,1}, A_{t+h,1}) = \ell) = \frac{p_B p_{1,0}}{1-(1-p_B)p_{0,0}}.\end{aligned}$$

To verify the expression in (2.7), notice that $M_{t+h}^{(1)}$ conditioned on $M_t^{(1)} = k$ has the distributional form:

ℓ	$P(M_{t+h}^{(1)} = \ell \mid M_t^{(1)} = k)$
1	$p_{0,1}$
$\vdots$	$\vdots$
$k-1$	$p_{0,0}^{k-2} p_{0,1}$
k	$p_{0,0}^{k-1} p_{1,1}$
$k+1$	$p_B(p_{0,0}^{k-1} p_{1,0})$
$k+2$	$(1-p_B)p_B(p_{0,0}^{k-1} p_{1,0})$
$\vdots$	$\vdots$

Using these in the law of total expectation gives

$$\psi_{1,1}(h) = \frac{1}{p_B p_{0,1}} - \frac{p_B p_{0,0}}{[1-(1-p_B)p_{0,0}]^2 p_{0,1}} + \frac{p_{1,0}}{[1-(1-p_B)p_{0,0}]^2}.$$

Finally, we derive an explicit form for $\mathbb{E}\left[\min(A_{t,1}, A_{t+h,1})\right]$ and $\mathrm{var}\left(\min(A_{t,1}, A_{t+h,1})\right)$. The tail distribution of $\min(A_{t,1}, A_{t+h,1})$ satisfies

$$\mathbb{P}(\min(A_{t,1}, A_{t+h,1}) > \ell) = P\left(B_{t,1} = 0, B_{t+h,1} = 0, \cdots, B_{t,\ell} = 0, B_{t+h,\ell} = 0\right) = \left[(1-p_B)p_{0,0}\right]^{\ell},$$

which is an ordinary geometric distribution. Hence, $\mathbb{E}[\min(A_{t,1}, A_{t+h,1})] = [1-(1-p_B)p_{0,0}]^{-1}$ and $\mathrm{var}(\min(A_{t,1}, A_{t+h,1}) = (1-p_B)p_{0,0}/[1-(1-p_B)p_{0,0}]^2$. Plugging these into (A.5) verifies the claim.

A.2 Proofs of results in Chapter 3

We gather here the proofs of the results given in this work. Lemma 3.2.1 will follow from the following more general result.

Lemma A.2.1. *For bounded functions h_1, h_2, we have*

$$\mathbb{E}_X h_1(Z_0, \ldots, Z_t) = \frac{\int_{z_s \in A(x_s), s=0,\ldots,t} h_1(z_0, \ldots, z_t) e^{-\frac{1}{2}\sum_{s=0}^{t}(z_s - \widehat{z}_s)^2/r_s^2} dz_0 \ldots dz_t}{\int_{z_s \in A(x_s), s=0,\ldots,t} e^{-\frac{1}{2}\sum_{s=0}^{t}(z_s - \widehat{z}_s)^2/r_s^2} dz_0 \ldots dz_t} \tag{A.6}$$

and

$$\mathbb{E}_X h_2(Z_0, \ldots, Z_{t+1}) = \frac{\int_{z_s \in A(x_s), s=0,\ldots,t} h_2(z_0, \ldots, z_{t+1}) e^{-\frac{1}{2}\sum_{s=0}^{t+1}(z_s - \widehat{z}_s)^2/r_s^2} dz_0 \ldots dz_t dz_{t+1}}{\int_{z_s \in A(x_s), s=0,\ldots,t} e^{-\frac{1}{2}\sum_{s=0}^{t+1}(z_s - \widehat{z}_s)^2/r_s^2} dz_0 \ldots dz_t}, \tag{A.7}$$

where $\mathbb{E}_X$ refers to the expectation conditioned on $X_0 = x_0, \ldots, X_t = x_t$. Moreover, for bounded h_3,

$$\mathbb{E}h_3(X_0, \ldots, X_t) = \int_{\mathbb{R}^{t+1}} h_3(G(z_0), \ldots, G(z_t)) \frac{e^{-\frac{1}{2}\sum_{s=0}^{t}(z_s - \widehat{z}_s)^2/r_s^2}}{(2\pi)^{(t+1)/2} r_0 \ldots r_t} dz_0 \ldots dz_t. \tag{A.8}$$

Proof: We shall use the following expression which is well-known in the time series literature:

$$\mathbb{E}h(Z_0, \ldots, Z_t) = \int_{\mathbb{R}^{t+1}} h(z_0, \ldots, z_t) \frac{e^{-\frac{1}{2}\sum_{s=0}^{t}(z_s - \widehat{z}_s)^2/r_s^2}}{(2\pi)^{(t+1)/2} r_0 \ldots r_t} dz_0 \ldots dz_t \tag{A.9}$$

(e.g. Brockwell and Davis [6], Section 8.6). For the relation (A.6), note by (3.30) that

$$\mathbb{E}g(X_0, \ldots, X_t) h_1(Z_0, \ldots, Z_t) = \mathbb{E}g(G(Z_0), \ldots, G(Z_t)) h_1(Z_0, \ldots, Z_t)$$

$$= \sum_{i_0,\ldots,i_t \in \mathbb{N}_0} \mathbb{E} g(G(Z_0),\ldots,G(Z_t)) h_1(Z_0,\ldots,Z_t) 1_{\{G(Z_s)=i_s, s=0,\ldots,t\}}$$

$$= \sum_{i_0,\ldots,i_t \in \mathbb{N}_0} g(i_0,\ldots,i_t) \mathbb{E} h_1(Z_0,\ldots,Z_t) 1_{\{Z_s \in A(i_s), s=0,\ldots,t\}}$$

$$= \sum_{i_0,\ldots,i_t \in \mathbb{N}_0} g(i_0,\ldots,i_t) \frac{\mathbb{E} h_1(Z_0,\ldots,Z_t) 1_{\{Z_s \in A(i_s), s=0,\ldots,t\}}}{\mathbb{E} 1_{\{Z_s \in A(i_s), s=0,\ldots,t\}}} \mathbb{P}(X_s = i_s, s=0,\ldots,t).$$

This implies that

$$\mathbb{E}_X h_1(Z_0,\ldots,Z_t) = \frac{\mathbb{E} h_1(Z_0,\ldots,Z_t) 1_{\{Z_s \in A(x_s), s=0,\ldots,t\}}}{\mathbb{E} 1_{\{Z_s \in A(x_s), s=0,\ldots,t\}}},$$

which can be expressed as in (A.6) by using (A.9). The relation (A.7) can be proved similarly. The relation (A.8) follows from (A.9) since $h_3(X_0,\ldots,X_t) = h_3(G(Z_0),\ldots,G(Z_t))$. □

Remark A.2.1. The relation (A.8) implies, in particular, that

$$\mathbb{P}(X_0 = x_0,\ldots,X_t = x_t) = \int_{A(x_s), s=0,\ldots,t} \frac{e^{-\frac{1}{2}\sum_{s=0}^t (z_s - \widehat{z}_s)^2 / r_s^2}}{(2\pi)^{(t+1)/2} r_0 \ldots r_t} dz_0 \ldots dz_t. \tag{A.10}$$

We next give a proof of Lemma A.2.1.

Proof of Lemma A.2.1: The relation (3.31) follows from (A.6) since $\widehat{Z}_{t+1} = \widehat{z}_{t+1}(Z_0,\ldots,Z_t)$. Similarly, the relation (3.32) follows from (A.7) since $f(X_{t+1}) = f(G(Z_{t+1}))$. □

We next give a proof of Proposition 3.2.1.

Proof of Proposition 3.2.1: We drop the superscript i for notational simplicity. Note that

$$\mathbb{E}_X w_t f(\widehat{Z}_{t+1}) = \mathbb{E}_X w_{t-1} w_t(\widehat{Z}_t) f(\widehat{z}_{t+1}(Z_0,\ldots,Z_t)) = \mathbb{E}_X w_{t-1} w_t(\widehat{Z}_t) f(\widehat{z}_{t+1}(Z_0,\ldots,\widehat{Z}_t + r_t\epsilon_t))$$

$$= \mathbb{E}_X \Big[\mathbb{E}_X [w_{t-1} w_t(\widehat{Z}_t) f(\widehat{z}_{t+1}(Z_0,\ldots,\widehat{Z}_t + r_t\epsilon_t)) | Z_0,\ldots,Z_{t-1}] \Big]$$

$$= \mathbb{E}_X \Big[w_{t-1} w_t(\widehat{Z}_t) \frac{\int_{A(x_t)} f(\widehat{z}_{t+1}(Z_0,\ldots,z_t)) \frac{e^{-\frac{1}{2r_t}(z_t - \widehat{Z}_t)^2}}{\sqrt{2\pi r_t^2}} dz_t}{\int_{A(x_t)} \frac{e^{-\frac{1}{2r_t}(z_t - \widehat{z}_t)^2}}{\sqrt{2\pi r_t^2}} dz_t} \Big]$$

$$= \mathbb{E}_X \Big[w_{t-1} \int_{A(x_t)} f(\widehat{z}_{t+1}(Z_0,\ldots,z_t)) \frac{e^{-\frac{1}{2r_t}(z_t - \widehat{Z}_t)^2}}{\sqrt{2\pi r_t^2}} dz_t \Big],$$

where we used the definition of ϵ_t in Step 2 of the SIS algorithm, and that of $w_t(z)$ in (3.41).

Repeating a similar argument would lead next to the expression

$$\mathbb{E}_X\Big[w_{t-2}\int_{A(x_{t-1})}\int_{A(x_t)} f(\widehat{z}_{t+1}(Z_0,\dots,z_{t-1},z_t))\frac{e^{-\frac{1}{2r_{t-1}}(z_{t-1}-\widehat{Z}_{t-1})^2-\frac{1}{2r_t}(z_t-\widehat{z}_t(Z_0,\dots,z_{t-1}))^2}}{\sqrt{2\pi r_{t-1}^2}\sqrt{2\pi r_t^2}}dz_{t-1}dz_t\Big].$$

Further similar iterations yield

$$\mathbb{E}_X w_t f(\widehat{Z}_{t+1}) = \frac{\int_{z_s\in A(x_s),s=0,\dots,t} f(\widehat{z}_{t+1})\frac{e^{-\frac{1}{2}\sum_{s=0}^t (z_s-\widehat{z}_s)^2/r_s^2}}{(2\pi)^{(t+1)/2}r_0\dots r_t}dz_0\dots dz_t}{\int_{z_0\in A(x_0)}\frac{e^{-z_0^2/2}}{(2\pi)^{1/2}r_0}dz_0}$$

(the term in the denominator does not cancel out since $w_0 = 1$), and the proposition follows in view of (3.31). □

Bibliography

[1] M. Al-Osh and A.A. Alzaid. Integer-valued moving averages (INMA). *Statistical Papers*, 29:281–300, 1988.

[2] L. Beghin and C. Macci. Fractional discrete processes: compound and mixed poisson representations. *Journal of Applied Probability*, 51:19–36, 2014.

[3] P.A. Blight. Time series formed from the superposition of discrete renewal processes. *Journal of Applied Probability*, 26:189–195, 1989.

[4] G. Box and G. Jenkins. *Time Series Analysis: Forecasting and Control*. First edition, 1970.

[5] Braun R.K. Braun, J.V. and H.G. Muller. Multiple changepoint fitting via quasilikelihood, with application to dna sequence segmentation. *Biometrika*, 87:301–314, 2000.

[6] P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods*. Springer Series in Statistics. Springer-Verlag, New York, second edition, 1991.

[7] Y. Cui and R.B. Lund. A new look at time series of counts. *Biometrika*, 96:781–792, 2009.

[8] Y. Cui and R.B. Lund. Inference for binomial AR(1) models. *Statistics and Probability Letters*, 80:1985–1990, 2010.

[9] C. Czado, T. Gneiting, and L. Held. Predictive model assessment for count data. *Biometrics*, 65:1254–1261, 2009.

[10] R. Douc, E. Moulines, and D. S. Stoffer. *Nonlinear Time Series*. Chapman & Hall/CRC Texts in Statistical Science Series. Chapman & Hall/CRC, Boca Raton, FL, 2014. Theory, methods, and applications with R examples.

[11] Z. Han and V. De Oliveria. On the correlation structure of Gaussian copula models for geostatistical count data. *Australian & New Zealand Journal of Statistics*, 58:47–69, 2016.

[12] C.R. Heathcote. Complete exponential convergence and related topics. *Journal of Applied Probability*, 4:217–256, 1967.

[13] J.M. Hilbe. *Negative Binomial Regression*. Second edition edition.

[14] P.A. Jacobs and P.A.W. Lewis. Discrete time series generated by mixtures i: Correlational and runs properties. *Journal of the Royal Statistical Society*, 40:94–105, 1978a.

[15] P.A. Jacobs and P.A.W. Lewis. Discrete time series generated by mixtures ii: Asymptotic properties. *Journal of the Royal Statistical Society*, 40:222–228, 1978b.

[16] H. Joe. Time series models with univariate margins in the convolution-closed infinitely divisible class. *Journal of Applied Probability*, 33:664–677, 1996.

[17] M. Kachour and J.F. Yao. Firstorder rounded integervalued autoregressive (RINAR(1)) process. *Journal of Time Series Analysis*, 30:417–448, 2009.

[18] S. Kolassa. Evaluating predictive count data distributions in retail sales forecasting. *International Journal of Forecasting*, 32:788–803, 2016.

[19] J. Livsey, R. Lund, S. Kechagias, and V. Pipiras. Multivariate integer-valued time series with flexible autocovariances and their application to major hurricane counts. *Annals of Applied Statistics*, Forthcoming.

[20] R.B. Lund, S. Holan, and J.A. Livsey. Long memory discrete-valued time series. In R.A. Davis, S.H. Holan, R.B. Lund, and N. Ravishanker, editors, *Handbook of Discrete-Valued Time Series*. CRC Press, New York City, NY, 2015.

[21] R.B. Lund and J. Livsey. Renewal-based count time series. In R.A. Davis, S.H. Holan, R.B. Lund, and N. Ravishanker, editors, *Handbook of Discrete-Valued Time Series*. CRC Press, New York City, NY, 2015.

[22] E. McKenzie. Some simple models for discrete variate time series. *Water Resources Bulletin*, 21:645–650, 1985.

[23] E. McKenzie. Autoregressive-moving average processes with negative-binomial and geometric marginal distributions. *Advances in Applied Probability*, 18:679–705, 1986.

[24] E. McKenzie. Some ARMA models for dependent sequences of poisson counts. *Advances in Applied Probability*, 20:822–835, 1988.

[25] C.T. Ng, H. Joe, D. Karlis, and J. Liu. Composite likelihood for time series models with a latent autoregressive process. *Statistica Sinica*, 21:279–305, 2011.

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

[27] V. Pipiras and M. S. Taqqu. *Long-Range Dependence and Self-Similarity*, volume 45. Cambridge University Press, 2017.

[28] C. Rose and M.D. Smith. The multivariate normal distribution. *The Mathematica Journal*, 6:32–37, 1996.

[29] F.W. Steutel and K. Van Harn. Discrete analogues of self-decomposability and stability. *Annals of Probability*, 7:893–899, 1979.

[30] J.H. Van Vleck and D. Middleton. The spectrum of clipped noise. *Proceedings of the IEEE*, 52:2–19, 1966.

[31] J.M. Ver Hoef and P.L. Boveng. Quasi-poisson vs. negative binomial regression: how should we model overdispersed count data? *Ecological Society of America*, 88:2766–2772, 2007.

[32] C.H. Weiβ. Controlling correlated processes with binomial marginals. *Journal of Applied Statistics*, 36:399–414, 2009.

[33] C.H. Weiβ. A new class of autoregressive models for time series of binomial counts. *Communications in Statistics theory and Methods*, 38:447–460, 2009.

[34] G.C. White and R.E. Bennetts. Analysis of frequency count data using the negative binomial distribution. *Ecological Society of America*, 77:2549–2557, 1996.

[35] W. Whitt. Bivariate distributions with given marginals. *The Annals of Statistics*, 4(6):1280–1289, 1976.

[36] P. Whittle. Hypothesis testing in time series analysis. *Journal of the American Statistical Association*, 49:197–200, 1954.