Stationary count time series models

Christian H. Weiß

Department of Mathematics and Statistics, Helmut Schmidt University, Hamburg, Germany

Correspondence
Christian H. Weiß, Department of Mathematics and Statistics, Helmut Schmidt University, Hamburg, Germany. Email: weissc@hsu-hh.de

Abstract
During the last 20–30 years, there was a remarkable growth in interest on approaches for stationary count time series. We consider popular classes of models for such time series, including thinning-based models, conditional regression models, and Hidden-Markov models. We review and compare important members of these model families, having regard to stochastic properties such as the dispersion and autocorrelation structure. Our survey covers univariate and multivariate count data, as well as unbounded and bounded counts. We also discuss an illustrative data example. Besides this critical presentation of the current state-of-the-art, some existing challenges and opportunities for future research are identified.

This article is categorized under:
Statistical Models > Time Series Models
Data: Types and Structure > Time Series, Stochastic Processes, and Functional Data
Statistical Learning and Exploratory Methods of the Data Sciences > Modeling Methods
Statistical and Graphical Methods of Data Analysis > Modeling Methods and Algorithms

KEYWORDS
autocorrelation, counts, dispersion, Markov, modeling, nonlinear, process, regression, stationary, thinning, time series

1 | INTRODUCTION

The area of time series analysis has attracted a lot of interest in research and practice during the last 50 years. Many textbooks are meanwhile available on this topic, with the pioneering work being the famous book by Box and Jenkins (1970). However, most of these textbooks exclusively concentrate on real-valued time series, despite the fact that in applications, we are often faced with different types of data. Especially count time series are often monitored in practice, that is, time series consisting of discrete and quantitative observations from the set of non-negative integers, \( \mathbb{N}_0 = \{0, 1, \ldots\} \). Accordingly, models and methods for count time series were also covered by a large number of research articles during the last few decades, but they did not enter the general textbook literature yet. As a consequence, they are still not well communicated to a broader audience. As a step towards a greater visibility of this topic, this article presents a compact overview on the current state-of-the-art of models for stationary count time series. Certainly, there has also been research on nonstationary count time series or on other types of discrete-valued time series (see the book by...
Weiß (2018) for a more comprehensive survey), but the above focus offers an excellent starting point for discovering this fascinating and important area.

In what follows, we assume \( x_1, \ldots, x_T \) to be a univariate time series originating from a stationary count process \((X_t)_{t \in \mathbb{Z}} = \{\ldots, -1, 0, 1, \ldots\}\). In the case of multivariate counts, bold letters are used instead, that is, \( \mathbf{x}_t \text{ or } \mathbf{X}_t \), respectively. For identifying appropriate types of models for these data, we have to check the characteristics of the considered count time series. An important distinction is, for example, the one between bounded or unbounded counts: if there exists an upper limit \( n \in \mathbb{N} \) that can never be exceeded, then only models designed for the bounded range \([0, \ldots, n]\) are relevant. Otherwise, we are concerned with unbounded counts, having the full set \( \mathbb{N}_0 \) as their range. Next, we have to check marginal characteristics like the dispersion and zero probability, see Box 1 for a brief survey, and serial characteristics like the autocorrelation structure. Based on these findings, we define a set of candidate models for the subsequent statistical model fitting and diagnostics.

As an example, let us have a look at a count time series (length \( T = 242 \)) about the weekly number of sales of a soap product in a supermarket. The data are taken from MacDonald and Zucchini (2016), and a summary is provided by Figure 1. Since there is no specified upper bound for the number of sales, we treat the counts as being unbounded. They have the sample mean \( \approx 5.44 \), whereas the variance is about 2.8 times larger, taking the value \( \approx 15.40 \). So the counts exhibit a lot of overdispersion, which is also visible from the plot of the sample probability mass function (PMF) in Figure 1. The sample autocorrelation function (ACF) shows weak autocorrelation (the lag-1 value equals \( \hat{\rho}(1) \approx 0.392 \)), which quickly decreases with increasing time lag \( h \). Actually, the partial ACF (PACF) in Figure 1 has a significant value only at lag 1, indicating that the true data-generating process (DGP) might have a first-order autoregressive (AR) structure. So relevant candidate models for the sales count time series might be found within the class of conditionally linear models (Grunwald, Hyndman, Tedesco, & Tweedie, 2000), and the observed (P)ACF characteristics shall help us to identify them.

Possible models to be used as candidates for the model fitting of a stationary count time series might be found in the large family of thinning-based models, which is surveyed in Section 2. Another popular class are conditional regression models for count time series, see Section 3 for an overview. Furthermore, also the Hidden-Markov models (HMMs) discussed in Section 4 are commonly used for count time series. For each of these model families, important members and their stochastic properties are presented. In addition, common features across and differences between these families are discussed. In Sections 2–4, we also continue the data analysis of the above sales count time series, and briefly address some further data applications. Finally, Section 5 concludes and outlines current research directions.

**Box 1. Possible Properties of Count Distributions**

The “normal” distribution for a count random variable \( X \) having the unbounded range \( \mathbb{N}_0 \) is the Poisson (Poi) distribution. It is characterized, among others, by the so-called equidispersion property: its variance is equal to the mean, \( \mathbb{V}[X] = \mathbb{E}[X] \). Furthermore, its zero probability is related to the mean by the equation \( P(X = 0) = \exp(-\mathbb{E}[X]) \). If, in contrast, \( X \) has a variance being larger (smaller) than the mean, then \( X \) is said to exhibit overdispersion (underdispersion). Analogously, if \( P(X = 0) \) is larger (smaller) than \( \exp(-\mathbb{E}[X]) \), we refer to \( X \) as being zero-inflated (zero-deflated). A similar classification is done for bounded counts defined on \([0, \ldots, n]\), but then with respect to the binomial (Bin) distribution instead of the Poisson distribution. In particular, since the binomial variance satisfies \( n \mathbb{V}[X] = \mathbb{E}[X](n-\mathbb{E}[X]) \), a bounded-count random variable having \( \mathbb{V}[X] > \mathbb{E}[X](1-\mathbb{E}[X]/n) \) is said to exhibit extra-binomial variation.

Several count distributions have been proposed in the past to capture the above features. For example, a common choice for unbounded counts with overdispersion is the negative-binomial (NB) distribution, whereas the beta-binomial (BB) distribution is often used for bounded counts with extra-binomial variation, see Johnson, Kemp, and Kotz (2005) for a comprehensive overview. If one wants to capture underdispersion as well, then Conway-Maxwell extensions of the Poisson or binomial distribution, respectively, constitute a possible solution (Sellers, Swift, & Weems, 2017).

Such dispersion features are also observed for multivariate count data, and several multivariate extensions of parametric models for univariate counts have been developed, see the survey by Johnson, Kotz, and Balakrishnan (1997). Additional complexity might be caused by a sophisticated cross-correlation structure or by heterogeneous marginal properties. A possible solution is to use copulas to combine the different types of marginal distribution (Genest & Nešlehová, 2007; Joe, 1997).
In data applications with stationary count time series, one often observes an autocorrelation structure being similar to that of the ordinary autoregressive moving-average (ARMA) models (Box & Jenkins, 1970),

$$Y_t = \alpha_1 Y_{t-1} + \cdots + \alpha_p Y_{t-p} + \epsilon_t - \beta_1 \epsilon_{t-1} - \cdots - \beta_q \epsilon_{t-q},$$  

where the $Y_t$ and $\epsilon_t$ are real-valued random variables. Model (1), however, cannot be applied to count processes, because the multiplications in (1) do not preserve the integer range (“multiplication problem”). Thus, there has been a lot of research effort to modify the basic ARMA model (1) in such a way that the resulting model generates only count data, but still with an ARMA-like ACF. For the models to be discussed in this section, the idea is to directly substitute the multiplications in (1) by an integer-valued counterpart, a so-called “thinning operation.” The resulting models are then referred to as integer-valued ARMA (INARMA) models. There exists a large variety of possible thinning operations, see the recent survey by Scotto, Weiβ, and Gouveia (2015), but in what follows, we shall mainly concentrate on the binomial thinning operation “$^\alpha$” due to Steutel and van Harn (1979). For $\alpha \in [0; 1]$, and applied to the count random variable $X$, the newly generated count $\alpha \circ X$ arises from $X$ by binomial thinning if the distribution of $\alpha \circ X \mid X$ is the binomial distribution $\text{Bin}(X, \alpha)$. So in contrast to the ordinary multiplication “$\cdot$”, the thinning “$\circ$” constitutes a random operation that generates a count value from $\{0, \ldots, X\}$. But since both operations have the same mean, we can use binomial thinning as an integer-valued substitute of the multiplication. The name “thinning” expresses the shrinking effect of “$\circ$.” As discussed in section 2.1.1 in Weiβ (2018), we can interpret $\alpha \circ X$ as the random number of “survivors” from a “population” of size $X$, where each individual survives with the probability $\alpha$. This type of interpretation carries over to the INARMA models to be discussed now.

2. INARMA models for unbounded counts

The first INARMA model was proposed by McKenzie (1985) and Al-Osh and Alzaid (1987): the INAR(1) model defined by the recursion $X_t = \alpha \circ X_{t-1} + \epsilon_t$ with $\alpha \in [0; 1)$, which constitutes an integer-valued counterpart to the ordinary first-order AR model (i.e., model (1) with $p = 1$ and $q = 0$). Here, the innovations $(\epsilon_t)_\mathbb{Z}$ are independent and identically distributed.

FIGURE 1  Sales count time series: plot against time $t$ in top row; plots of sample probability mass function against $x$ as well as (partial) autocorrelation function against lag $h$ in bottom row.
(i.i.d.) count random variables, and \( \epsilon_t \) is generated independently of \( \{X_{s:t} \} \). Since “\( \alpha \)” is a random operation, further model assumptions are necessary, namely that all thinning operations are performed independently of each other and of \( \{\epsilon_t\} \), and that the thinning operation at time \( t \) is independent of \( \{X_{s:t-1} \} \). The term “\( \alpha \circ X_{t-1} \)” can be interpreted as the number of survivors of generation \( t - 1 \), and “\( \epsilon_t \)” expresses the immigration to generation \( t \). Similarly, Al-Osh and Alzaid (1988) and McKenzie (1988) defined an integer-valued counterpart to the MA(1) model (so (1) with \( p = 0 \) and \( q = 1 \)), the INMA(1) model, by \( X_t = \epsilon_t + \beta \circ \epsilon_{t-1} \). Both types of first-order models have properties analogous to those of their (Gaussian) ARMA complement: Poi(\( \lambda \))-distributed innovations \( (\lambda > 0) \) lead to Poi(\( \mu \))-distributed observations \( (\mu = \lambda/(1 - \rho(1))) \), where the ACF \( \rho(h) \) is given by \( \rho(h) = a^h \) for the INAR(1) model, and where \( \rho(h) = 0 \) for \( h \geq 2 \) for the INMA(1) model.

Both the INAR(1) and the INMA(1) model can be embedded into higher-order AR-type or MA-type models, respectively. The INMA(\( q \)) model is defined by

\[
X_t = \beta_0 \circ \epsilon_t + \beta_1 \circ \epsilon_{t-1} + \cdots + \beta_q \circ \epsilon_{t-q}, \quad q \geq 1,
\]

where the \( q + 1 \) thinnings at time \( t \) (the time index \( t \) has been added below the operator “\( \circ \)” in (2) for clarification) are performed independently of each other. For the above INMA(1) model, we set \( \beta_0 = 1 \). But if \( q > 1 \), further model assumptions are necessary to uniquely define the INMA(\( q \)) model: the conditional distribution of \( (\beta_0 \circ \epsilon_t, \beta_1 \circ \epsilon_{t+1}, \epsilon_t, \ldots, \beta_q \circ \epsilon_{t+q}, \epsilon_t) \) given \( \epsilon_t \) has to be specified. This can be done in quite different ways, leading to different types of INMA (\( q \)) models, see Brännäs and Hall (2001) for details. For example, one may require conditional independence between the thinnings experienced by \( \epsilon_t \), as it was suggested by McKenzie (1988). But in any case, Poi-distributed innovations still go along with Poi-distributed observations, and also the ACF \( \rho(h) \) drops down to zero if \( h > q \).

For the INAR(\( p \)) model (\( p \geq 1 \)) defined by

\[
X_t = \alpha_1 \circ X_{t-1} + \cdots + \alpha_p \circ X_{t-p} + \epsilon_t, \quad \text{where} \quad \alpha := \sum_{j=1}^{p} \alpha_j \leq 1,
\]

the situation is more complex. If assuming the conditional distribution of \( (\alpha_1 \circ X_{t-1}, \ldots, \alpha_p \circ X_{t-p}, \epsilon_t) \) given \( X_t \) to be multinomial as suggested by Alzaid and Al-Osh (1990), Poi-distributed innovations cause Poi-distributed observations, but the ACF differs from the ACF of an ordinary AR(\( p \)) model. Therefore, the proposal of Du and Li (1991) is usually preferred in practice, because their assumption of conditional independence for the above thinnings implies that the INAR(\( p \))’s ACF can be computed by solving the well-known Yule–Walker equations

\[
\rho(h) = \sum_{i=1}^{p} \alpha_i \rho(|h-i|) \quad \text{for} \quad h \geq 1.
\]

For this ”DL-INAR(\( p \)) model,” however, the above Poi-preservation is lost. Generally, the observations’ variance \( \sigma^2 \) has to be computed from the innovations’ variance \( \sigma^2_\epsilon \) according to

\[
\sigma^2 \cdot \left( 1 - \sum_{i=1}^{p} \alpha_i \rho(i) \right) = \mu \sum_{j=1}^{p} \alpha_j (1 - \alpha_j) + \sigma^2_\epsilon,
\]

whereas we have the well-known relation \( \mu = \mu_\epsilon/(1 - \alpha_j) \) between the observations’ and innovations’ mean. The transition probabilities of the DL-INAR(\( p \)) process are computed as a convolution between \( p \) binomial distributions (caused by the thinnings) and the innovations’ distribution. For example, if \( p = 1 \) and \( \alpha_1 = \alpha \), we have

\[
P(X_t = k | X_{t-1} = l) = \sum_{j=0}^{\min(k,l)} \binom{l}{j} \alpha^j (1 - \alpha)^{l-j} \cdot P(\epsilon_t = k-j).
\]

The required PMF for \( \epsilon_t \) follows from the actual model assumption: for the Poi-INAR(1) model, we take the PMF of the Poi(\( \lambda \)) distribution, whereas the NB-INAR(1) model assumes the innovations to follow NB(\( n, \pi \)) (also see Box 1). The
latter approach allows to model an overdispersed and AR(1)-like count time series, which could be relevant for the sales count data introduced in Section 1; we shall continue this data example later in Section 2.4. Several further possibilities for choosing the innovations’ distribution have been discussed in the literature; especially members of the compound-Poisson family proved to be suitable for the INAR(1) model (Schweer & Weiβ, 2014).

Many modifications and generalizations of the basic INARMA models, especially of the widely-used INAR(1) model, have been proposed in the literature. A possible approach is to replace the binomial thinning operation by another type of thinning. Latour (1998) considered a family of generalized thinning operations, which includes the popular NB-thinning of Ristić, Bakouch, and Nastić (2009) as a special case. Another modification is to assume the thinning parameters to be random themselves (random coefficient thinning), see Zheng, Basawa, and Datta (2006), Zheng, Basawa, and Datta (2007), and Gomes and e Castro (2009) for details. For example, if the binomial thinning’s parameter is beta-distributed, then we obtain the BB-thinning operation, which allows to generate additional dispersion for the considered INARMA process. For the convolution-closed models proposed by Joe (1996), the thinning operations are defined such that the observations and innovations stem from the same distribution family. A broad subclass of these models has recently developed by Sellers, Peng, and Arab (2019), which relies on Conway–Maxwell extensions of the Poisson distribution (for the observations and innovations) and of the binomial distribution (for the thinning operation). See Scotto et al. (2015) and Weiβ (2018) for a comprehensive discussion of alternative thinning operations, and Joe (2019) for more background on convolution-closed models.

### 2.2 Thinning-based models for bounded counts

The INARMA models discussed in the previous Section 2.1 can only be applied to time series consisting of unbounded counts. If a thinning-based model for bounded counts is required, with a given upper bound $n \in \mathbb{N}$, then modifications are required to avoid that $X_t > n$ happens. As an example, McKenzie (1985) proposed to modify the INAR(1) model

$$X_t = \alpha \cdot X_{t-1} + \beta \cdot (n - X_{t-1}),$$  \hfill (7)

where the thinning parameters $\alpha, \beta \in (0; 1)$ can be expressed as $\beta = \pi \cdot (1 - \rho)$ and $\alpha = \beta + \rho$ with $\pi \in (0; 1)$ and $\rho \in \{\frac{-\pi}{1-\pi}, \frac{1-\pi}{\pi}; 1\}$. Model (7) constitutes a finite Markov chain with transition probabilities

$$P(X_t = k | X_{t-1} = l) = \frac{\min[k, l]}{\max[l, 0, k + l - n]} \binom{l}{m} \binom{n - l}{k - m} \alpha^m (1 - \alpha)^{l - m} \beta^k (1 - \beta)^{n - l + m - k}. \hfill (8)$$

The parametrization by $(\pi, \rho)$ is useful regarding the basic stochastic properties of model (7): its stationary marginal distribution is Bin$(n, \pi)$, and the ACF equals $\rho(h) = \rho^h$. The parametrization by $(\alpha, \beta)$, in contrast, is more easy to interpret, with “$\alpha \cdot X_{t-1}$” again being the survival term, and “$\beta \cdot (n - X_{t-1})$” being the revival term.

Weiβ (2009b) extended the binomial AR(1) model (7) to a pth-order autoregression by using a random mixture similar to that of the discrete ARMA model discussed in Section 4.2. The resulting binomial AR(p) model satisfies the pth-order Yule–Walker equations (4). Model (7) was also modified to capture extra-binomial variation. Weiβ and Kim (2014) replaced the binomial thinning operations in (7) by beta-binomial ones, that is, the thinning parameters $\alpha, \beta$ are randomly drawn from a beta distribution, which causes additional (conditional) variance. Möller, Weiβ, Kim, and Sirchenko (2018) developed several extensions for zero-inflated bounded counts, with different types of zero patterns. Many further modifications have been proposed in the literature; Section 5 briefly discusses some of these recent developments, also regarding the other model families discussed in Sections 2–4.

Let us conclude this section by briefly referring to a data example. Weiβ and Kim (2014) considered a group of $n = 17$ countries in the Euro area and determined the number of countries with stable prices among them (per month for the period 2000–2006). A plot of the resulting bounded-counts time series in shown in Figure 2. From the corresponding sample ACF, an AR-like autocorrelation structure becomes clear. Furthermore, Weiβ and Kim (2014) showed that the data exhibit significant extra-binomial variation (about 1.52 times more variance than for a binomial model). Therefore, they decided to use the beta-binomial AR(1) model with parameter estimates $\hat{\alpha} \approx 0.72$ and $\hat{\beta} \approx 0.10$. 
Thus, a country with stable prices in month $t$ will also have stable prices in month $t + 1$ with probability 72\% (“survival”), whereas the probability for newly getting stable prices is only about 10\% (“revival”).

### 2.3 Multivariate INARMA models

While there has been a lot of research concerning univariate count time series during the last decades, there is still relatively little work on multivariate count time series $x_1, \ldots, x_T \in \mathbb{N}_0^d$ with $d \geq 2$. Within the thinning-based approaches, the first proposal is due to Franke and Rao (1993), who defined an integer-valued counterpart to the vector ARMA models by using matrix-binomial thinning as a substitute of matrix multiplication. More precisely, with $A \in [0; 1)^{d \times d}$, this type of multivariate thinning operation is defined as

$$
(A \circ X)_i = \sum_{j=1}^d a_{ij} X_j \quad \text{for} \quad i = 1, \ldots, d,
$$

(9)

where the $d^2$ univariate binomial thinning operations on the right-hand side of (9) are performed independently of each other. A more general family of matrix thinnings is considered by Latour (1997). Using such an operation (we still focus on the binomial type of matrix thinning for simplicity), a multivariate extension of the INAR(1) model is defined by

$$
X_t = A \circ X_{t-1} + \epsilon_t,
$$

(10)

where $(\epsilon_t)_\mathbb{Z}$ is an i.i.d. $d$-dimensional count process with finite mean $\mu_\epsilon$ and covariance matrix $\Sigma_\epsilon$ (Franke & Rao, 1993). A $p$-th-order extension, in analogy to the INAR($p$) model by Du and Li (1991), was proposed by Latour (1997). Model (10) has a unique stationary solution if the polynomial $\det(I - A z) \neq 0$ for $|z| \leq 1$ (Latour, 1997), where $I$ denotes the $d \times d$-identity matrix. In this case, the marginal mean equals $\mu = (I - A)^{-1} \mu_\epsilon$, and the autocovariance function satisfies the Yule–Walker equations

$$
\Gamma(k) = A^k \Gamma(0) = A^k \Sigma, \quad \Sigma = A \Sigma A^\top + \text{diag}(B) + \Sigma_\epsilon,
$$

(11)

where $B$ has the entries $b_{ij} = a_{ij}(1 - a_{ij})$. If the component processes $(X_{t,i})_\mathbb{Z}$ should constitute univariate INAR(1) processes on their own, then $A$ has to be chosen as a diagonal matrix, $A = \text{diag}(a_1, \ldots, a_d)$, such that the cross-correlation between $X_{t,i}$ and $X_{t,j}$ is solely caused by the innovations (Pedeli & Karlis, 2011). Such a diagonal-matrix bivariate INAR (1) model was used by Pedeli and Karlis (2011) for the bivariate count time series $x_1, \ldots, x_{365} \in \mathbb{N}_0^2$ plotted in Figure 3. It provides the daily number of road accidents in the Schiphol area (Netherlands) in 2001, where it is distinguished between daytime ($x_{1,t}$) and nighttime ($x_{2,t}$) accidents. Both component series are weakly autocorrelated, and they also exhibit a weak positive cross-correlation. For the innovations $\epsilon_t$, Pedeli and Karlis (2011) tried out bivariate extension of both the Poi- and the NB-distribution, and finally, they decided for the latter model in view of the apparent over-dispersion in the data.
To allow for a flexible cross-correlation structure (including negative cross-correlations), Karlis and Pedeli (2013) use copulas. Popović, Ristić, and Nastić (2016) constructed a bivariate model based on a random mixture of component-wise INAR(1) recursions; an analogous model with NB thinning instead of binomial thinning is due to Ristić, Nastić, Jayakumar, and Bakouch (2012). In both cases, the model can be rewritten similarly to (10), but using a special type of random coefficients for the thinning matrix $A$.

For the case of bivariate bounded counts (with possibly different bounds $n_1$, $n_2$ in the different components), Scotto, Weiß, Silva, and Pereira (2014) developed a counterpart to the binomial AR(1) model (7), which uses a thinning operation based on the so-called bivariate binomial distribution of type II (Johnson et al., 1997). If both components have the same upper bound $n_1 = n_2 = n$, the bivariate binomial AR(1) model by Ristić and Popović (2019) can be used as an alternative.

### 2.4 | Application to sales count data

Let us have a look back to the sales count data, see Figure 1 and the discussion in Section 1. Since the time series consists of univariate and unbounded counts, only the models presented in Section 2.1 are relevant for this example. The autocorrelation structure of the data appears AR(1)-like, so the INAR(1) model seems to be a reasonable candidate model. Furthermore, because of the strong overdispersion, an INAR(1) model with overdispersion should be used. Thus, we take the NB-INAR(1) model as a candidate, but also the Poi-INAR(1) model is included for illustration. The detailed estimation results as well as several checks for model adequacy can be found in Supplement S, Supporting information; here, only a brief summary of the main findings is given. Background information on the considered diagnostic tools can be found in Weiß (2018).

If using Akaike’s information criterion (AIC) or the Bayesian information criterion (BIC) for model selection, then the NB-INAR(1) model is clearly preferred. This preference is confirmed if looking at the fitted models variance (or at the variance of the respective Pearson residuals), because only the NB-INAR(1) model is able to capture the observed overdispersion reasonably well. The thinning parameter (“survival probability”) is estimated as $\hat{\alpha} \approx 0.254$, so the fitted model implies that about 25% of the soled soap products in week $t-1$ initiate a further purchase in week $t$. However, comparing the ACF of the fitted model to the sample ACF, it becomes clear that the NB-INAR(1) model is not able to fully explain the actual serial dependence structure (also the Pearson residuals exhibit rather large ACF values at lags 1–4). Therefore, we shall later continue our search for an adequate model once we discussed the next big class of models for stationary count time series.
3 | REGRESSION MODELS FOR COUNT TIME SERIES

Another popular approach for the modeling of stationary count time series is conditional regression models. Imposing appropriate parameter constraints, it is again possible to obtain an ARMA-like ACF structure. The resulting models are often referred to as integer-valued generalized autoregressive conditional heteroskedasticity (INGARCH) models, see Section 3.1. But also nonlinear forms of dependence are possible if defining a conditional regression model within the framework of generalized linear models (GLMs), see Section 3.2. Multivariate extensions of the INGARCH model are briefly surveyed in Section 3.3. For more comprehensive discussions of regression models for count time series, the reader is referred to Chapter 4 in the book by Kedem and Fokianos (2002), and to Chapters 4 and 5 in Weiß (2018).

3.1 | INGARCH models

The basic idea behind INGARCH models is to transfer the ARMA recursion (1) to the level of conditional means, \( M_t := E[X_t \mid X_{t-1}, \ldots] \). Since the mean of a count random variable is not integer-valued but just a positive real number, it is possible to find parameter constraints for the linear recursive scheme such that negative outcomes are avoided. At this point, one may certainly ask why an integer-valued ARMA-like model is referred to as the INGARCH model. To explain this apparent contradiction, let us recall that, for example, the GARCH(1, 1) model is defined by requiring the conditional variances \( \sigma_t^2 \) to follow \( \sigma_t^2 = \beta_0 + \alpha_1 Y_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \). The INGARCH(1, 1) model relies on the same recursive scheme but applied to the conditional means \( M_t \), that is, \( M_t = \beta_0 + \alpha_1 X_{t-1} + \beta_1 M_{t-1} \). So the data-generating mechanisms are analogous in this sense. Nevertheless, the name “INGARCH” is still controversially discussed and a couple of alternative proposals have been made in the literature (Weiß, 2018, Remark 4.1.2). We continue referring to these models as “INGARCH” in the sequel, because this name seems to be more often used than any of its competitors.

Following Ferland, Latour, and Oraichi (2006), the INGARCH(\( p, q \)) model with \( p \geq 1 \) and \( q \geq 0 \) is defined by the linear recursion

\[
M_t = \beta_0 + \sum_{i=1}^{p} \alpha_i X_{t-i} + \sum_{j=1}^{q} \beta_j M_{t-j}
\]  

(12)

with \( \beta_0 > 0 \) and \( \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q \geq 0 \). If \( q = 0 \), we refer to it as an INARCH(\( p \)) model. The INGARCH model is fully specified if the conditional distribution of \( X_t \) given \( X_{t-1}, X_{t-2}, \ldots \) has been fixed. The most simple choice is the conditional Poisson distribution \( \text{Poi}(M_t) \). The resulting Poi-INGARCH model is sometimes also referred to as a (linear) Poisson autoregressive model (Fokianos, Rahbek, & Tjøstheim, 2009). Ferland et al. (2006) showed that the condition \( \alpha_1 + \beta_1 + \sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1 \) guarantees the existence of a strictly stationary solution to (12) with finite first- and second-order moments. Its mean is given by \( \mu = \beta_0/(1 - \alpha - \beta) \), whereas the variance satisfies \( V[X_t] = \mu + V[M_t] \), that is, the observations are overdispersed (Heinen, 2003). Furthermore, see Weiß (2009a), the Poi-INGARCH model satisfies a set of Yule–Walker-like equations:

\[
\gamma(k) = \sum_{i=1}^{p} \alpha_i \gamma(k-i) + \min\{k-1, q\} \beta_j \gamma(k-j) + \sum_{j=1}^{q} \beta_j \gamma_M(j-k),
\]

\[
\gamma_M(l) = \sum_{i=1}^{\min\{l, p\}} \alpha_i \gamma_M(l-i) + \sum_{i=1}^{q} \alpha_i \gamma_M(l-i) + \sum_{j=1}^{q} \beta_j \gamma_M(l-j),
\]

(13)

where \( \gamma(h) := \text{Cov}[X_t, X_{t-h}] \), \( \gamma_M(h) := \text{Cov}[M_t, M_{t-h}] \), and \( k \geq 1, l \geq 0 \). Note that for \( q = 0 \), these equations reduce to the ordinary AR(\( p \))’s Yule–Walker equations, that is, a Poi-INGARCH(\( p \)) model satisfies (4). Thus, from a modeling perspective, it is a competitor to the DL-INAR(\( p \)) model (3). In fact, one might understand the Poi-INGARCH(\( p \)) model as a special type of “generalized-thinning INAR(\( p \)) model” (Section 2.1) using a Poi-thinning operation (Weiß, 2015, Section 2).

The INGARCH model can also be combined with non-Poisson conditional distributions. For example, additional overdispersion can be achieved by using the conditional distribution \( \text{NB}(n, \pi) \) with \( n_t = M_t \frac{\pi^r}{1-\pi} \), that is, \( \pi \in (0; 1) \) serves...
as an additional dispersion parameter (Xu, Xie, Goh, & Fu, 2012). As an illustration, such an NB-INARCH(1) model satisfies $\sigma^2 = \frac{1}{2} \cdot \mu/(1-\alpha^2)$, whereas for the Poi-INARCH(1) model, the extent of overdispersion is solely determined by the dependence parameter $\alpha$: $\sigma^2 = \mu/(1-\alpha^2)$. Many further proposals regarding the INGARCH’s conditional distribution can be found in the literature, also see Ahmad and Francq (2016). For example, another type of NB-INGARCH model was proposed by Zhu (2011), and both NB-INGARCH models can be embedded into the family of compound-Poisson INGARCH models developed by Gonçalves, Mendes-Lopes, and Silva (2015). Mixed Poisson models are studied by Christou and Fokianos (2014). If using a conditional binomial distribution instead, we obtain an INGARCH model for bounded counts (Weiβ & Pollett, 2014).

Let us conclude this section with a critical comparison of the INARMA and the INGARCH family. Both constitute an ARMA-like approach for stationary count time series, and their pure AR-members even have an identical autocorrelation structure. Nevertheless, there are a couple of differences between both families. First, while the INARMA family also includes pure MA-type models, this is not the case for the INGARCH family. Quite the contrary, the “MA-part” in (12) actually causes some kind of “long memory.” This is easily seen for the INGARCH(1, 1) model, where

$$M_t = \alpha_1 \sum_{k=1}^{t} \beta_1^{k-1} \cdot X_{t-k} + \beta_1^1 \cdot M_0 + \beta_0 \cdot \frac{1-\beta_1^1}{1-\beta_1^1},$$

that is, the current observation is affected by all previous observations (Fokianos, 2012). Because of this memory, the INGARCH(1, 1) model was successfully applied to transactions count data, see, for example, Christou and Fokianos (2014); Fokianos et al. (2009); Weiβ (2018). As a second main difference between INGARCH and INARMA, marginal properties are generally more easy to compute for INARMA models, whereas INGARCH models have more simple conditional distributions. For example, the transition probabilities of a Poi-INARCH(1) model, the extent of overdispersion is solely determined by the

$$\ln M_t = \beta_0 + \alpha_1 \ln(X_{t-1} + 1) + \beta_1 \ln M_{t-1},$$

which can be understood as a multiplicative type of INGARCH(1, 1) model, because $M_t = e^{\beta_0 \cdot (X_{t-1} + 1)^{\alpha_1} \cdot M_{t-1}^{\beta_1}}$. Thanks to the log-link, the model parameters $\beta_0$, $\alpha_1$, $\beta_1$ might also become negative while preserving a positive conditional mean $M_t$. But for the existence of a stationary solution, parameter restrictions are required (Fokianos & Tjøstheim, 2011): $|\alpha_1 + \beta_1| < 1$ if $\alpha_1$, $\beta_1$ have the same sign, and $\alpha_1^2 + \beta_1^2 < 1$ otherwise.

Another ARMA-like but nonlinear conditional regression model is the generalized ARMA (GARMA) model considered by Benjamin, Rigby, and Stasinopoulos (2003), having a model recursion of the form

$$\ln M_t = \beta_0 + \sum_{i=1}^{p} \alpha_i \cdot A(X_{t-i}) + \sum_{j=1}^{q} \beta_j \cdot M(X_{t-j}, M_{t-i}),$$

where $A$ and $M$ are functions representing the AR- and MA-terms, respectively. Also the generalized linear ARMA (GLARMA) model proposed by Davis, Dunsmuir, and Streett (2003) has to be mentioned in this context, which first defines the residual-like quantities $e_i = (X_i - M_i)/M_i^\lambda$ with $\lambda \in (0; 1]$. Then the model recursion is given by

3.2 Nonlinear conditional regression models

Within the framework of GLMs, it is easily possible to also obtain a nonlinear autocorrelation structure. This is achieved by introducing an additional link function on the left-hand side of (12), preferably one having the full sets of reals as its range, because this may allow to weaken the necessary parameter constraints. For time series consisting of unbounded counts, the log-link is often used, whereas one commonly applies the logit-link to the normalized conditional mean $M_t/n$ of bounded counts. As an example, Fokianos and Tjøstheim (2011) consider the log-linear Poisson autoregressive model defined by

$$\ln M_t = \beta_0 + \alpha_1 \ln(X_{t-1} + 1) + \beta_1 \ln M_{t-1},$$

where $A$ and $M$ are functions representing the AR- and MA-terms, respectively. Also the generalized linear ARMA (GLARMA) model proposed by Davis, Dunsmuir, and Streett (2003) has to be mentioned in this context, which first defines the residual-like quantities $e_i = (X_i - M_i)/M_i^\lambda$ with $\lambda \in (0; 1]$. Then the model recursion is given by
While all the previous models are observation-driven, a parameter-driven Poisson regression model is the one by Zeger (1988). It is based on a latent process \((e_i)_i\), which is positive real-valued and weakly stationary with mean 1, variance \(\sigma^2\) and ACF \(\rho(h)\). The count \(X_t\) at time \(t\) is then generated from the state-dependent distribution \(\text{Poi}(\mu e_t)\) with parameter \(\mu > 0\), and conditioned on \((e_i)\), \((X_i)\) is assumed to consist of independent counts. Thus, the observations’ mean equals \(\mu\), the variance equals \(\mu + \mu^2 \sigma^2\) (overdispersion), and the ACF is given by \(\rho(h) = \rho_e(h) / \left(1 + (\mu \sigma^2)^{-1}\right)\) (being the ACF of the latent process with an additional damping factor).

As a final remark, although we focus on stationary count processes here, it should be mentioned that the regression models discussed in the present section are easily extended to nonstationary processes by additionally including covariate information. Details are provided in the above references as well as in the books by Kedem and Fokianos (2002); Weiß (2018). As an example, both Benjamin et al. (2003) and Zeger (1988) analyzed a time series of monthly counts of poliomyelitis cases in the United States. They applied a NB-GARMA(0, 2) model and a parameter-driven Poisson regression model with latent AR(1) process, respectively. But in both cases, also a linear trend and harmonic oscillation (with annual and semiannual cycles) had to be included to obtain a satisfactory fit.

### 3.3 Multivariate INGARCH models

In analogy to the INARMA case (recall the discussion in Section 2.3), there is rather little work on multivariate extensions of the INGARCH model. A multivariate INGARCH(1, 1) model for unbounded counts assumes the conditional mean \(M_t = E[X_t | X_{t-1},...]\) to satisfy the linear recursion

\[
M_t = b_0 + A X_{t-1} + B M_{t-1},
\]

where \(b_0 \in (0; \infty)^d\), and where \(A, B\) are non-negative \(d \times d\)-matrices. This linear scheme can also be extended to higher model orders, see Heinen and Rengifo (2003). The count \(X_t\) at time \(t\) is then generated according to the selected conditional distribution. Cui and Zhu (2018), for example, consider a special type of bivariate Poisson distribution, which also allows for negative cross-correlation. Heinen and Rengifo (2003) propose to use (among others) copulas for defining a multivariate extension of the conditional Poisson distribution (also recall the analogous approach by Karlis and Pedeli (2013) concerning multivariate INARMA models). Such a copula approach is also considered by Fokianos, Støve, Tjøstheim, and Doukhan (2020) for model (17) and, in addition, for a log-linear type of multivariate INGARCH(1, 1) model, defined in analogy to (14):

\[
\ln M_t = b_0 + A \ln(X_{t-1} + 1) + B \ln M_{t-1}.
\]

Here, the logarithm is applied component-wise, and \(1\) denotes the vector of ones. Conditions for the existence of a stationary solution to (17) and (18) are derived by Cui and Zhu (2018) and Fokianos et al. (2020). Both papers also present applications to bivariate count time series regarding the number of transactions of certain stocks traded at the New York Stock Exchange. For model (17), the stationary mean of \(X_t\) is given by \((I - A - B)^{-1} b_0\), and formulae for variance and autocovariance are provided by Heinen and Rengifo (2003). A bivariate INARCH(1) model for bounded counts, using a conditional bivariate binomial distribution of type II, was proposed by Scotto et al. (2014).

### 3.4 Application to sales count data

Let us return to the sales count time series with its AR(1)-like autocorrelation structure (Section 1). In Section 2.4, it turned out that the INAR(1) model does not constitute an approach for adequately describing the data. Therefore, in this section, we shall consider its counterpart within the INGARCH family instead, that is, the INARCH(1) model. Note that, in contrast to the Poi-INAR(1) model, the Poi-INARCH(1) model exhibits some degree of overdispersion in its marginal distribution, so it might turn out as a reasonable choice for the data. But especially the NB-INARCH(1) model
(in the sense of Xu et al. (2012)) appears as a plausible candidate. Detailed results regarding model estimation and adequacy checks can again be found in Supplement S.

For both models, ML estimation leads to significant estimates, where both information criteria, AIC and BIC, result in a clear preference for the NB-INARCH(1) model. Although the fitted Poi-INARCH(1) model exhibits some overdispersion, it can neither explain the sample variance, nor the conditional dispersion structure as evaluated by the Pearson residuals (variance much larger than one) and the PIT histogram (clearly U-shaped). The NB-INARCH(1) model, in contrast, is even favored over the NB-INAR(1) model by the information criteria. This preference appears plausible in view of the fact that NB-INARCH(1) model is better able to capture the actual autocorrelation structure. But while the Pearson residuals do not show significant ACF values anymore, they still have relatively large ACF values at lags 2–4. A possible solution could be to increase the AR order \( p \); in Section 4.3, we shall consider another type of candidate model instead, namely a Poisson HMM.

4 \textbf{FURTHER MODELS FOR COUNT TIME SERIES}

Although INARMA and INGARCH models are probably the most widely used approaches for stationary count time series, a couple of further models have been proposed in the literature. Two of them are briefly presented in the sequel: the HMM in Section 4.1, and the discrete ARMA model in Section 4.2.

4.1 \textbf{HMM for counts}

HMM is a parameter-driven time series model, which can be applied to count processes, but also to processes with any other kind of range, even to categorical processes. HMMs date back to Baum and Petrie (1966), who referred to these models as “probabilistic functions of Markov chains.” Nowadays, the standard reference on HMMs is the comprehensive book by Zucchini, MacDonald, and Langrock (2016), while compact surveys are provided by MacDonald and Zucchini (2016) and Weiß (2018). A HMM is a bivariate process \((X_t, Q_t)_{t \in \mathbb{N}}\), where the \(X_t\) are the actual (count) observations, and the \(Q_t\) are the hidden states. The latter are assumed to be categorical with some numerically coded range \(Q = \{0, \ldots, d_Q\}\) where \(d_Q \in \mathbb{N}\). \((X_t, Q_t)_{t \in \mathbb{N}}\) is said to follow a HMM if the following three properties hold for all \(q, r \in \mathbb{Q}\) and all \(t \in \mathbb{N}_0\):

1. observation equation \(P(X_t \mid X_{t-1}, \ldots, Q_t, \ldots) = P(X_t \mid Q_t)\), with time-homogeneous state-dependent distributions \(P(X_t = x \mid Q_t = q) = p(x \mid q)\);
2. state equation \(P(Q_t \mid X_{t-1}, \ldots, Q_{t-1}, \ldots) = P(Q_t \mid Q_{t-1}, \ldots)\);
3. Markov assumption \(P(Q_t = q \mid Q_{t-1}, \ldots) = P(Q_t = q \mid Q_{t-1})\), with time-homogeneous state transition probabilities \(P(Q_t = q \mid Q_{t-1} = r) = a_{q \mid r}\).

Thus, HMMs are special types of parameter-driven (generalized) state-space models (Brockwell & Davis, 2016). For a stationary HMM, we have to refine the third requirement by an additional stationarity assumption.

A stationary HMM has \(d_Q (d_Q + 1)\) parameters related to the transition probabilities of the hidden states, which are summarized through the transition matrix \(A = (a_{q \mid r})_{q,r}\). Further parameters are implied by the \(d_Q + 1\) state-dependent distributions. These are typically taken to be parametric count models, for example, state-dependent Poisson models for the case of unbounded counts (thus referred to as a Poi-HMM), or state-dependent binomial models for time series of bounded counts. In both cases, we end up with altogether \((d_Q + 1)^2\) model parameters, which increases quadratically in the number of hidden states.

The stationary marginal distribution is a probabilistic mixture of the state-dependent distributions, for example, a Poisson mixture for a Poi-HMM. Here, the mixture weights are given by the stationary marginal probabilities \(\pi_q = P(X_t = q)\) of the hidden states. For a Poi-HMM, with \(X_t \mid Q_t = q \sim \text{Poi}(\lambda_q)\), we thus have the observations’ mean \(\mu = \sum_{q \in \mathbb{Q}} \pi_q \lambda_q\) and variance \(\mu + \left(\sum_{q \in \mathbb{Q}} \pi_q \lambda_q^2\right) - \mu^2\) (overdispersion). The autocovariance function is given by

\[
\gamma(k) = \sum_{q,r \in \mathbb{Q}} (A^k)_{q,r} \pi_q \lambda_q \lambda_r - \mu^2.
\]
Joint probabilities and likelihood function are computed efficiently by using the forward algorithm, and a likelihood-based identification of the sequence of hidden states \( q_1, ..., q_T \), corresponding to the given time series data \( x_1, ..., x_T \), is done by using the Viterbi algorithm; detailed descriptions are provided by Zucchini et al. (2016) and Weiß (2018).

### 4.2 Discrete ARMA models for counts

One of the first approaches for generating discrete time-series data with an ARMA-like dependence structure are the discrete ARMA models proposed by Jacobs and Lewis (1983). They define two types of discrete ARMA models relying on a random mixture of past observations and innovations; in what follows, we concentrate on the one referred to as “NDARMA model” by Jacobs and Lewis (1983). Let the observations \((X_t)_{t \in \mathbb{Z}}\) and the innovations \((\epsilon_t)_{t \in \mathbb{Z}}\) be count processes, where \((\epsilon_t)_{t \in \mathbb{Z}}\) is i.i.d. with \( P(\epsilon_t = i) = \pi_i \), and where \( \epsilon_t \) is independent of \((X_s)_{s < t}\). Furthermore, we have to assume a process of i.i.d. multinomial vectors

\[
\left( \alpha_{t,1}, ..., \alpha_{t,p}, \beta_{t,0}, ..., \beta_{t,q} \right) \sim \text{Mult}(1; \phi_1, ..., \phi_p, \varphi_0, ..., \varphi_q) \quad \text{with } p + q \geq 1,
\]

which are independent of \((\epsilon_t)_{t \in \mathbb{Z}}\) and of \((X_s)_{s < t}\). These multinomial vectors become one in exactly one of their components (and zero otherwise), and they enable us to define a random mixture as follows: \((X_t)_{t \in \mathbb{Z}}\) is said to follow a discrete ARMA model if

\[
X_t = \alpha_{t,1} \cdot X_{t-1} + \cdots + \alpha_{t,p} \cdot X_{t-p} + \beta_{t,0} \cdot \epsilon_t + \cdots + \beta_{t,q} \cdot \epsilon_{t-q}
\]

holds. As a consequence, the upcoming observation \(X_t\) is generated by simply selecting one of either the past observations \(X_{t-1}, ..., X_{t-p}\) (with probabilities \(\phi_1, ..., \phi_p\), or the available innovations \(\epsilon_t, ..., \epsilon_{t-q}\) (with probabilities \(\varphi_0, ..., \varphi_q\). Therefore, the stationary marginal distribution of the observation \(X_t\) is identical to the one of the innovations \(\epsilon_t\). Furthermore, \((X_t)_{t \in \mathbb{Z}}\) has an ARMA-like autocorrelation structure: the autocovariance function satisfies

\[
\gamma(k) = \sum_{j=1}^{p} \phi_j \cdot \gamma(|k-j|) + \sigma^2 \sum_{i=0}^{q-k} \varphi_{i+k} \cdot r(i) \quad \text{for } k \geq 1,
\]

where \(r(i) = 0\) for \(i < 0\), and \(r(i) = \sum_{j=\max\{0,i-p\}}^{i-1} \phi_{j} \cdot r(j) + \phi_{j} \cdot 1(0 \leq i \leq q)\) otherwise. Here, \(1(\cdot)\) denotes the indicator function. Note that \(\sigma^2 = \sigma^2_{\epsilon}\) holds for discrete ARMA models, so (21) applies to the ACF after having removed the factor “\(\sigma^2\)”.

In view of the simple data-generating mechanism and the Yule–Walker equations (21), the discrete ARMA models appear as an attractive approach at the first glance. However, their sample paths are characterized by long runs of count values, interrupted by sudden jumps, a pattern that is hardly found in applications. As a possible solution to make these models more practice-oriented, Gouveia, Möller, Weiß, and Scotto (2018) suggest to generate more variation within successive counts by adding a variation operator to the model recursion (20). More precisely, Gouveia et al. (2018) consider the case of bounded counts and define the binomial variation operator by \(bv_n(X) \mid X \sim \text{Bin}(n, X/n)\). Then they modify (20) as

\[
X_t = \sum_{i=1}^{p} \alpha_{t,i} \cdot bv_n(X_{t-i}) + \beta_{t,0} \cdot \epsilon_t + \sum_{j=1}^{q} \beta_{t,j} \cdot bv_n(\epsilon_{t-j}),
\]

and they refer to the resulting model as the bvARMA model. Despite this modification, the Yule–Walker equations (21) still hold true for the bvARMA model, and observations and innovations still have the same mean. But other distributional properties differ between \(X_t\) and \(\epsilon_t\); in particular, we generally have \(\sigma^2 \neq \sigma^2_{\epsilon}\) for the bvARMA model, see Gouveia et al. (2018) for details. They applied the bvARMA model to a large data base of count time series regarding the number of rainy days per week (thus \(n = 7\)) at meteorological stations throughout Europe and Russia.
4.3 Application to sales count data

Let us return one last time to the sales count time series. In Section 3.4, we concluded that the three-parameter NB-INARCH(1) model constitutes a good fit to the data (although having some deficiencies). From the models discussed here in Section 4, only the HMM appears as a reasonable candidate model. Actually, in MacDonald and Zucchini (2016), a 3-state Poi-HMM was recommended for the sales counts. Therefore, we shall now conclude our data analysis by also fitting such a type of HMM to the data, and by comparing its performance to that of the NB-INARCH(1) model. The estimation results in Supplement S give a somewhat ambiguous picture. On the one hand, the fitted HMM performs very well with respect to the considered marginal and serial properties. In particular, it closely matches the sample ACF at lags 1–4, which goes along with nearly uncorrelated Pearson residuals. Also the PIT histogram has a nearly uniform shape, so the conditional dispersion is well captured by the HMM. There are some deviations in the marginal distribution (e.g., the zero probability differs notably from the zero frequency), but also here, we mainly have a close agreement. On the other hand, this excellent fit is partially offset by the disadvantage of having a large number of model parameters: nine parameters for the 3-state Poi-HMM versus three parameters for the NB-INARCH(1) model. Accordingly, the BIC indeed prefers the NB-INARCH(1) model, whereas the AIC selects the Poi-HMM. As another consequence, the fitted Poi-HMM has several nonsignificant estimates. If one finally decides for the fitted 3-state Poi-HMM, then the actual sales counts are understood as stemming from three possible states expressing some kind of “general sales activity” (parameter-driven data generation); the result of a global decoding of the sales count time series is shown Supplement S.6. In contrast, if selecting the NB-INARCH(1) model, then the current count is explained based on the previous count, so in an observation-driven way.

5 SUMMARY AND PERSPECTIVE

We reviewed important model classes for stationary count time series. These include the widely used INARMA, INGARCH, and HMMs, which were also applied to an illustrative data example. Essential properties of the presented models were summarized, and a critical comparison of the different model classes was also provided. Furthermore, first indications about current research directions were given, namely an increasing interest in multivariate count time series models as well as in more flexible distributional settings. We now continue in these regards with a discussion of further existing challenges and arising opportunities for future research in this field, without any claim to be exhaustive.

A frequently expressed criticism refers to the inability of many count time series models to handle negative autocorrelations. This concerns the INARMA and INGARCH models for unbounded counts, whereas the bounded-counts models such as binomial AR or binomial INARCH are more flexible in these respects. Also the models developed by Cui and Lund (2009) and Livsey, Lund, Kechagias, and Pipiras (2018), which use a superpositioning of binary processes, can capture negative autocorrelations. Log-linear models allow for negative ACF values, but these do not satisfy a simple Yule–Walker scheme anymore. So developing models with an approximately linear autocorrelation structure but allowing also for negative autocorrelations appears to be a promising direction for future research.

Another recent trend is to equip count time series models with a nonlinear dependence structure. This might be done by using conditional regression models as in Section 3.2, or by assuming the model parameters to depend on previous observations (Weiβ, 2015; Weiβ & Pollett, 2014). Especially models with a self-exciting threshold mechanism have been studied intensively during the last years, among others by Monteiro, Scotto, and Pereira (2012), Wang, Liu, Yao, Davis, and Li (2014), and Möller, Silva, Weiβ, Scotto, and Pereira (2016). In contrast, approaches with a less abrupt regime switching have been widely ignored yet; an exception is the hysteretic INGARCH model by Truong, Chen, and Sriboonchitta (2017), where the threshold value is replaced by a buffering region. Also the development of count time series models with a long memory appears as an important research direction; first contributions are Cui and Lund (2009), Quoreshi (2014), and Segnon and Stapper (2019).

Although the present article concentrated on the modeling of stationary count time series, also their statistical analysis is a field of utmost importance for practice. In fact, there already exists a large literature on approaches for parameter estimation and model diagnostics concerning count time series, see Weiβ (2018) for references, but also several opportunities remain (e.g., with respect to nonparametric or semiparametric statistical inference). Finally, after having fitted a count time series model and having checked its adequacy, the model is commonly used for forecasting or change-point detection (process monitoring). Also these fields continue to attract a lot of attention by researchers.
ACKNOWLEDGMENTS
The author thanks the two anonymous referees for providing valuable comments that have helped strengthen this manuscript.

CONFLICT OF INTEREST
The author has declared no conflicts of interest for this article.

ORCID
Christian H. Weiβ https://orcid.org/0000-0001-8739-6631

RELATED WIREs ARTICLES
Autoregressive processes
Integer-valued time series

REFERENCES
Ahmad, A., & Francq, C. (2016). Poisson QMLE of count time series models. *Journal of Time Series Analysis*, 37(3), 291–314.
Al-Osh, M., & Alzaid, A. (1987). First-order integer-valued autoregressive (INAR(1)) process. *Journal of Time Series Analysis*, 8(3), 261–275.
Al-Osh, M., & Alzaid, A. (1988). Integer-valued moving average (INMA) process. *Statistical Papers*, 29(1), 281–300.
Alzaid, A., & Al-Osh, M. (1990). An integer-valued $p$th-order autoregressive structure (INAR($p$)) process. *Journal of Applied Probability*, 27(2), 314–324.
Baum, L., & Petrie, T. (1966). Statistical inference for probabilistic functions of finite state Markov chains. *Annals of Mathematical Statistics*, 37(6), 1554–1563.
Benjamin, M., Rigby, R., & Stasinopoulos, D. (2003). Generalized autoregressive moving average models. *Journal of the American Statistical Association*, 98(1), 214–223.
Box, G., & Jenkins, G. (1970). *Time series analysis: Forecasting and control* (1st ed.). San Francisco, CA: Holden-Day.
Brännäs, K., & Hall, A. (2001). Estimation in integer-valued moving average models. *Applied Stochastic Models in Business and Industry*, 17(3), 277–291.
Brockwell, P., & Davis, R. (2016). *Introduction to time series and forecasting* (3rd ed.). Switzerland: Springer International Publishing.
Christou, V., & Fokianos, K. (2014). Quasi-likelihood inference for negative binomial time series models. *Journal of Time Series Analysis*, 35(1), 55–78.
Cui, Y., & Lund, R. (2009). A new look at time series of counts. *Biometrika*, 96(4), 781–792.
Cui, Y., & Zhu, F. (2018). A new bivariate integer-valued GARCH model allowing for negative cross-correlation. *TEST*, 27(2), 428–452.
Davis, R., Dunsmuir, W., & Streett, S. (2003). Observation-driven models for Poisson counts. *Biometrika*, 90(4), 777–790.
Du, J.-G., & Li, Y. (1991). The integer-valued autoregressive (INAR($p$)) model. *Journal of Time Series Analysis*, 12(2), 129–142.
Ferland, R., Latour, A., & Oraichi, D. (2006). Integer-valued GARCH processes. *Journal of Time Series Analysis*, 27(6), 923–942.
Fokianos, K. (2012). Count time series models. In T. Rao, S. Rao, & C. Rao (Eds.), *Handbook of statistics: Time series analysis – Methods and applications* (pp. 315–347). Amsterdam: Elsevier B.V.
Fokianos, K., Rahbek, A., & Tjøstheim, D. (2009). Poisson autoregression. *Journal of the American Statistical Association*, 104(4), 1430–1439.
Fokianos, K., Steve, B., Tjostheim, D., & Doukhan, P. (2020). Multivariate count autoregression. *Bernoulli*, 26(1), 471–499.
Fokianos, K., & Tjøstheim, D. (2011). Log-linear Poisson autoregression. *Journal of Multivariate Analysis*, 102, 563–578.
Franke, J., & Rao, T. S. (1993). *Multivariate first-order integer-valued autoregression* (Tech. Rep. No. 95). Germany: University of Kaiserslautern.
Genest, C., & Nešlehová, J. (2007). A primer on copulas for count data. *Astin Bulletin*, 37(2), 475–515.
Gomes, D., & e Castro, L. C. (2009). Generalized integer-valued random coefficient for a first order structure autoregressive (RCINAR) process. *Journal of Statistical Planning and Inference*, 139(12), 4088–4097.
Gonçalves, E., Mendes-Lopes, N., & Silva, F. (2015). Infinitely divisible distributions in integer-valued GARCH models. *Journal of Time Series Analysis*, 36(4), 503–527.
Gouveia, S., Möller, T., Weiß, C., & Scotto, M. (2018). A full ARMA model for counts with bounded support and its application to rainy-days time series. *Stochastic Environmental Research and Risk Assessment*, 32(9), 2495–2514.
Grunwald, G., Hyndman, R., Tedesco, L., & Tweedie, R. (2000). Non-Gaussian conditional linear AR(1) models. *Australian and New Zealand Journal of Statistics*, 42(4), 479–495.
Heinen, A. (2003). *Modelling time series count data: an autoregressive conditional Poisson model* (CORE Discussion Paper No. 2003/62). Belgium: University of Louvain.
Heinen, A., & Rengifo, E. (2003). *Multivariate modelling of time series count data: an autoregressive conditional Poisson model* (CORE Discussion Paper No. 2003/25). Belgium: University of Louvain.
Jacobs, P., & Lewis, P. (1983). Stationary discrete autoregressive-moving average time series generated by mixtures. *Journal of Time Series Analysis*, 4(1), 19–36.
Joe, H. (1996). Time series models with univariate margins in the convolution-closed infinitely divisible class. Journal of Applied Probability, 33(3), 664–677.

Joe, H. (1997). Multivariate models and dependence concepts (1st ed.). Boca Raton, FL: Chapman & Hall/CRC Press.

Joe, H. (2019). Likelihood inference for generalized integer autoregressive time series models. Econometrics, 7(4), 43.

Johnson, N., Kemp, A., & Kotz, S. (2005). Univariate discrete distributions (3rd ed.). Hoboken, NJ: John Wiley & Sons, Inc.

Johnson, N., Kotz, S., & Balakrishnan, N. (1997). Discrete multivariate distributions (1st ed.). Hoboken, NJ: John Wiley & Sons, Inc.

Karlis, D., & Pedeli, X. (2013). Flexible bivariate INAR(1) processes using copulas. Communications in Statistics—Theory and Methods, 42(4), 723–740.

Kedem, B., & Fokianos, K. (2002). Regression models for time series analysis (1st ed.). Hoboken, NJ: John Wiley & Sons, Inc.

Latour, A. (1997). The multivariate GINAR(p) process. Advances in Applied Probability, 29(1), 228–248.

Latour, A. (1998). Existence and stochastic structure of a non-negative integer-valued autoregressive process. Journal of Time Series Analysis, 19(4), 439–455.

Livsey, J., Lund, R., Kechagias, S., & Pipiras, V. (2018). Multivariate integer-valued time series with flexible autocovariances and their applications to major hurricane counts. Annals of Applied Statistics, 12(1), 408–431.

MacKenzie, E. (1985). Some ARMA models for dependent sequences of Poisson counts. Advances in Applied Probability, 17(4), 645–650.

MacKenzie, E. (1988). Some ARMA models for dependent sequences of Poisson counts. Advances in Applied Probability, 20(4), 822–835.

Möller, T., Silva, M., Weiß, C., Scotto, M., & Pereira, I. (2016). Self-exciting threshold binomial autoregressive processes. ASTA Advances in Statistical Analysis, 100(4), 369–400.

Möller, T., Weiß, C., Kim, H.-Y., & Sirchenko, A. (2018). Modeling zero inflation in count data time series with bounded support. Methodology and Computing in Applied Probability, 20(2), 589–609.

Monteiro, M., Scotto, M., & Pereira, I. (2012). Integer-valued self-exciting threshold autoregressive processes. Communications in Statistics—Theory and Methods, 41(15), 2717–2737.

Pedeli, X., & Karlis, D. (2011). A bivariate INAR(1) process with application. Statistical Modelling, 11(4), 325–349.

Popović, B., Ristić, M., & Nastić, A. (2016). A geometric bivariate time series with different marginal parameters. Statistical Papers, 57(3), 731–753.

Quareshi, A. (2014). A long-memory integer-valued time series model, INARFIMA, for financial application. Quantitative Finance, 14(12), 2225–2235.

Ristić, M., Bakouch, H., & Nastić, A. (2009). A new geometric first-order integer-valued autoregressive (NGINAR(1)) process. Journal of Statistical Planning and Inference, 139(7), 2218–2226.

Ristić, M., Nastić, A., Jayakumar, K., & Bakouch, H. (2012). A bivariate INAR(1) time series model with geometric marginals. Applied Mathematics Letters, 25(3), 481–485.

Ristić, M., & Popović, B. (2019). A new bivariate binomial time series model. Markov Processes and Related Fields, 25(2), 1–26.

Schweer, S., & Weiß, C. (2014). Compound Poisson INAR(1) processes: Stochastic properties and testing for overdispersion. Computational Statistics and Data Analysis, 77, 267–284.

Scotto, M., Weiß, C., & Gouveia, S. (2015). Thinning-based models in the analysis of integer-valued time series: A review. Statistical Modelling, 15(6), 590–618.

Scotto, M., Weiß, C., Silva, M., & Pereira, I. (2014). Bivariate binomial autoregressive models. Journal of Multivariate Analysis, 125, 233–251.

Segnon, M., & Stapper, M. (2019). Long memory conditional heteroscedasticity in count data (Tech. Rep. No. 82). Germany: University of Münster.

Sellers, K., Peng, S., & Arab, A. (2019). A flexible univariate autoregressive time-series model for dispersed count data. Journal of Time Series Analysis (in press). https://doi.org/10.1111/jtsa.12516.

Sellers, K., Swift, A., & Weems, K. (2017). A flexible distribution class for count data. Journal of Statistical Distributions and Applications, 4(22). https://doi.org/10.1186/s40488-017-0077-0.

Steutel, F., & van Harn, K. (1979). Discrete analogues of self-decomposability and stability. Annals of Probability, 7(5), 3893–3899.

Truong, B.-C., Chen, C., & Sriboonchitta, S. (2017). Hysteretic Poisson INGARCH model for integer-valued time series. Statistical Modelling, 17(6), 401–422.

Wang, C., Liu, H., Yao, J.-F., Davis, R., & Li, W. (2014). Self-excited threshold Poisson autoregression. Journal of the American Statistical Association, 109(2), 777–787.

Weiß, C. (2009a). Modelling time series of counts with overdispersion. Statistical Methods and Applications, 18(4), 507–519.

Weiß, C. (2009b). A new class of autoregressive models for time series of binomial counts. Communications in Statistics—Theory and Methods, 38(4), 447–460.

Weiß, C. (2015). A Poisson INAR(1) model with serially dependent innovations. Metrika, 78(7), 829–851.

Weiß, C. (2018). An introduction to discrete-valued time series (1st ed.). Chichester, England: John Wiley & Sons, Inc.

Weiß, C., & Kim, H.-Y. (2014). Diagnosing and modeling extra-binomial variation for time-dependent counts. Applied Stochastic Models in Business and Industry, 30(5), 588–608.

Weiß, C., & Pollett, P. (2014). Binomial autoregressive processes with density dependent thinning. Journal of Time Series Analysis, 35(2), 115–132.
Xu, H.-Y., Xie, M., Goh, T., & Fu, X. (2012). A model for integer-valued time series with conditional overdispersion. *Computational Statistics and Data Analysis, 56*(12), 4229–4242.

Zeger, S. (1988). A regression model for time series of counts. *Biometrika, 75*(4), 621–629.

Zheng, H., Basawa, I., & Datta, S. (2006). Inference for pth-order random coefficient integer-valued autoregressive processes. *Journal of Time Series Analysis, 27*(3), 411–440.

Zheng, H., Basawa, I., & Datta, S. (2007). First-order random coefficient integer-valued autoregressive processes. *Journal of Statistical Planning and Inference, 137*(1), 212–229.

Zhu, F. (2011). A negative binomial integer-valued GARCH model. *Journal of Time Series Analysis, 32*(1), 54–67.

Zucchini, W., MacDonald, I., & Langrock, R. (2016). *Hidden Markov models for time series: An introduction using R* (2nd ed.). London, England: Chapman & Hall/CRC.

**SUPPORTING INFORMATION**

Additional supporting information may be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Weiß CH. Stationary count time series models. *WIREs Comput Stat.* 2021;13:e1502.  
https://doi.org/10.1002/wics.1502