Latent Gaussian Count Time Series

Yisu Jia
University of North Florida

Stefanos Kechagias
SAS Institute

James Livsey
United States Census Bureau

Robert Lund
University of California - Santa Cruz

Vladas Pipiras
University of North Carolina - Chapel Hill

June 7, 2021

Abstract

Keywords: Count Distributions; Hermite Expansions; Likelihood Estimation; Particle Filtering; Sequential Monte Carlo; State Space Models

*

†

1 Introduction

This paper develops the theory and methods for modeling a stationary discrete-valued time series by transforming a Gaussian process. Since the majority of discrete-valued time series involve integer counts supported on some subset of $0\ 1$, we isolate on this support set. Our methods are based on a copula-style transformation of a latent Gaussian stationary series and are able to produce any desired count marginal distribution. It is shown that the proposed model class produces the most exible pairwise correlation structures possible, including negatively dependent series. Model parameters are estimated via 1) a Gaussian pseudo-likelihood approach, developed from some new Hermite expansion techniques, which use only the mean and the autocovariance of the series, 2) an implied Yule-Walker moment estimation approch when the latent Gaussian process is an autoregression, and 3) a particle ltering (PF) / sequential Monte Carlo (SMC) approach that uses a state space model (SSM) representation of the transformation to approximate the true likelihood. Extensions to non-stationary settings, particularly those with covariates, are discussed.

The theory of stationary Gaussian time series is by now well developed. A central result is that a stationary Gaussian series $_{\in\mathbb{Z}}$ having the lag- autocovariance $( ) = \mathrm{Cov}(\ _{+}\ )$ exist if and only if is symmetric about lag zero and non-negative de nite (see Theorem 1.5.1 in [6]). However, such a result does not hold for stationary count series having a certain prescribed marginal distribution (e.g, Poisson). In principle, distributional existence issues are checked with Kolmogorov s consistency criterion (see Theorem 1.2.1 in [6]); in practice, one needs a speci ed joint distribution to check for consistency. Phrased another way, Kolmogorov s consistency criterion is not a constructive result and does not illuminate how to build stationary time series having a particular marginal distribution and correlation structure. Perhaps owing to this, count time series have been constructed

from a plethora of approaches over the years, as is next reviewed.

Drawing from the success of autoregressive moving-average (ARMA) models in describing stationary Gaussian series, early count authors constructed correlated count series from discrete ARMA (DARMA) and integer ARMA (INARMA) di erence equation methods. Focusing on the rst order autoregressive case for simplicity, a DAR(1) series $_{=1}$ with speci ed marginal distribution ( ) is obtained by generating $_1$ from ( ) and then at each subsequent time, either keeping the previous count value with probability or generating an independent copy of ( ) with probability 1 . INAR(1) series are built via the thinned AR(1) equation = $_{-1}$ + , where is an IID count-valued random sequence and is a thinning operator de ned by = B( ) for a binomial distribution B( ) with trials and success probability . DARMA methods were initially explored in [23], but were subsequently discarded by practitioners because their sample paths often remained constant for long periods, especially in highly correlated cases; INARMA series are still used today. In contrast to their Gaussian ARMA brethren, DARMA and INARMA models, and their extensions in [24], cannot produce negative autocorrelations.

The works [5] and [9] take a di erent approach, producing the desired count marginal distribution by combining IID copies of a correlated Bernoulli series built from a stationary renewal sequence. Explicit autocovariance functions when is made by binning (clipping) a stationary Gaussian sequence into zero-one categories are derived in [33]. While these models can have negative correlations, they do not necessarily produce the most negatively correlated count structures possible. Also, some important count marginal distributions, including generalized Poisson, are not easily built from these methods. The results here easily generate any desired count marginal distribution. Other count model classes studied include Gaussian processes rounded to their nearest integer [26], hierarchical

Bayesian count model approaches [2], and others (see [18] and [11] for recent reviews). Each approach has some drawbacks.

The models here impose a xed marginal distribution for the counts. This is in contrast to generalized ARMA methods (GLARMA), which typically posit conditional distributions in lieu of marginal distributions, with model parameters typically being random. As [1] shows in the Poisson case, once the randomness of the parameters is taken into account, the true marginal distribution of the series can be far from the posited conditional distribution. This said, the literature on GLARMA and other conditional models is extensive [3, 43]. See [16] for a recent review of GLARMA models.

A time series analyst generally needs four features in a count model: 1) general marginal distributions; 2) the most general correlation structures possible, both positive and negative; 3) the straight-forward accomodation of covariates; and 4) a well performing and computationally feasible likelihood inference approach. All previous count classes fail to accommodate one or more of these tenets. This paper s purpose is to introduce and study a count model class that, for the rst time, simultaneously achieves all four features. Our model employs a latent Gaussian process and a copula-style transformation. This type of construction has recently shown promise in spatial statistics [12, 21], multivariate modeling [39, 40], and regression [35], but the theory has yet to be developed for count series ([35, 30] provide some partial results). Our objectives here are several-fold. On a methodological level, it is shown, through some newly derived Hermite polynomial expansions, that accurate and e cient numerical quanti cation of the correlation structure of this count model class is feasible. Based on a result in [42], the class is shown to produce the most exible pairwise correlation structures possible, positive or negative (see Remark 2.2 below). Connections to both importance sampling schemes, where the popular GHK sampler in [35] is

adapted to our needs, and to the SSM and SMC literature, which allow natural extensions of the GHK sampler and likelihood evaluation, are made. The methods are tested on both synthetic and real data.

The works [35, 30] are perhaps the closest papers to this study. While the general latent Gaussian construct adopted is the same, our work di ers in that explicit autocovariance relations are developed via Hermite expansions, exibility and optimality issues of the model class are addressed, Gaussian pseudo-likelihood and implied least-squares parameter estimation approaches are developed, and both the importance sampling and SSM connections are explored in detail. Additional connections to [35, 30] and to the spatial count modeling papers [21, 22] are later made.

The rest of this paper proceeds as follows. The next section and Appendix A introduce our Gaussian transformation count model and establish its basic mathematical and statistical properties. Section 3 and Appendix B move to estimation, developing three techniques: a Gaussian pseudo-likelihood approach, implied Yule-Walker estimation, and PF/SMC methods. Section 4 and Appendix C present simulation results. Section 5 and Appendix D analyze soft drink sales counts at one location of the now defunct Dominicks Finer Foods retail chain. This series exhibits overdispersion, negative lag one autocorrelation, and dependence on a price reduction (sales) covariate, which illustrates the exibility of our approach. Section 6 concludes with comments and suggestions for future research.

2 Theory

We seek to construct a strictly stationary time series having marginal distributions from any family of count distributions supported in 0 1 , including the Binomial,

Poisson, mixture Poisson, negative binomial, generalized Poisson, and Conway-Maxwell-Poisson distributions. The later three distributions are over-dispersed (their variances are larger than their respective means), which is the case for many observed count time series.

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

$$= (\ \ ) \quad \text{where} \quad (\ ) = \ ^{-1}(\ (\ )) \qquad \mathbb{R} \tag{1}$$

and $(\ )$ is the CDF of a standard normal variable and $^{-1}(\ ) = \inf \ : \ (\ )$, $(0\ 1)$, is the generalized inverse (quantile function) of the CDF . The process $_{\in\mathbb{Z}}$ is standard Gaussian for each xed , but possibly correlated in time:

$$\mathbb{E}[\ \ ] = 0 \quad \mathbb{E}[\ ^2] = 1 \qquad (\ ) =: \text{Corr}(\quad _+\ ) = \mathbb{E}[\quad _+\ ] \tag{2}$$

This approach has been used in [39, 35, 21, 30] with good results. The autocovariance function (ACVF) of , denoted by $(\ )$, is the same as the autocorrelation function (ACF) due to standard normality and depends on another vector $\boldsymbol{\eta}$ of ACVF parameters.

As expanded on in Section 2.3, (1) can be viewed as a SSM:

$$\text{State equation} \ : \ (\quad _{-1} \qquad _1) \text{ governing latent Gaussian dynamics;}$$
$$\text{Observation equation} \ : \ \mathbb{P}(\quad = \quad) = 1 \ _k(\ ) \text{ with the set de ned below.}$$

Here, $(\ \ )$ is notation for an arbitrary conditional distribution.

This model has alternative names in other literature. For example, [7] call this setup the normal to anything (NORTA) procedure in operations research, whereas [20] calls this

a translational model in mechanical engineering. Our goal is to give a reasonably complete analysis of the probabilistic and statistical properties of these models.

The construction in (1) ensures that the marginal CDF of is indeed ( ). Elaborating, the probability integral transformation theorem shows that ( ) has a uniform distribution over (0 1) for each ; a second application of the result justi es that has the marginal distribution ( ) for each . Moreover, temporal dependence in will induce temporal dependence in as quanti ed below. For notation, let ( ) = $\mathbb{E}[\ _{+}\ ]\ \ \mathbb{E}[\ _{+}\ ]\mathbb{E}[\ ]$ denote the ACVF of .

2.1 Relationship between autocovariances

The autocovariance functions of and can be related using Hermite expansions (see Chapter 5 of [37]). In particular, using the Hermite polynomials ( ) = (1) $^{2}\ ^{2}\frac{k}{k}(\ ^{-}\ ^{2}\ ^{2})$, $\mathbb{R}$ we can expand the ^{2} function as

$$( ) = \mathbb{E}[\ \ (\ _{0})] + \sum_{=1}^{\infty} \qquad ( ) \tag{3}$$

where the *Hermite coefficients* are given by

$$= \frac{1}{!}\int_{-\infty}^{\infty} \ ( )\ \ ( )\frac{^{-}\ ^{2}\ ^{2}}{\overline{2}} = \frac{1}{!}\mathbb{E}[\ \ (\ _{0})\ \ \ (\ _{0})] \tag{4}$$

for a standard normal variable $_{0}$. The relationship between ( ) and ( ) is key and is extracted from Chapter 5 of [37]:

$$( ) = \sum_{=1}^{\infty} \ ! \ ^{2}\ \ ( )\ =: \ (\ \ ( )) \tag{5}$$

where $(\ ) = \sum_{=1}^{\infty} \ !^{\ 2}$. For $= 0$, (5) yields $\mathrm{Var}(\ ) = \ (0) = \sum_{=1}^{\infty} \ !^{\ 2}$, which depends only on the marginal parameters in $\boldsymbol{\theta}$. Moreover, the ACF of is

$$(\ ) = \sum_{=1}^{\infty} \frac{!^{\ 2}}{(0)} \quad (\ ) =: \ (\quad (\ )) \tag{6}$$

where

$$(\ ) = \sum_{=1}^{\infty} \frac{!^{\ 2}}{(0)} \quad =: \sum_{=1}^{\infty} \tag{7}$$

and $= \ !^{\ 2} \quad (0)$. The function maps $[\ \ 1 \ 1]$ into (but not necessarily onto) $[\ \ 1 \ 1]$. For future reference, note that $(0) = 0$ and $(1) = \sum_{=1}^{\infty} \ = 1$. Using (3) and $\mathbb{E}[\quad (\ _0) \quad (\quad _0)] = (\ \ 1) \quad !1_{[\ =\ ]}$ gives $(\ \ 1) = \mathrm{Corr}(\ \ (\ _0) \quad (\quad _0))$; however, $(\ \ 1)$ is not necessarily 1 in general. As such, $(\ )$ starts at $(\ \ 1 \quad (\ \ 1))$, passes through $(0 \ 0)$, and connects to $(1 \ 1)$. Examples are given in Figure 2 of Appendix A.

We call the quantity $(\ )$ a *link* function, and the coe cients , 1, *link coefficients*. (Sometimes, slightly abusing terminology, we also use these terms for $(\ )$ and $^{\ 2} \ !$, respectively.) A key feature in (5) is that the e ects of the marginal CDF $(\ )$ and the ACVF $(\ )$ are decoupled in the sense that the correlation parameters in do not in uence the coe cients in (5) this is useful later in estimation.

Further properties and the numerical calculation of the link function and the Hermite coe cients are discussed in Appendix A. The computation of the Hermite coe cients, in particular, is feasible due to the following lemma, which is proved in Appendix A.

Lemma 2.1. *If* $\mathbb{E}[\quad]$ *for some* 1*, then the coefficients satisfy*

$$= \frac{1}{! \ \sqrt{2}} \sum_{=0}^{\infty} \ ^{-\ ^{-1}(\ _n)^2 \ 2} \quad _{-1}(\ ^{-1}(\quad)) \tag{8}$$

where $= \mathbb{P}[\quad\quad]$. *(When* $^{-1}(\quad) =$ *(that is,* $= 0$ *or* 1*), the summand* $^{-\ ^{-1}(\ _n)^2 \ 2} \quad _{-1}(\ ^{-1}(\quad))$ *is interpreted as zero.)*

Returning to the relationship between $\gamma_X(\cdot)$ and $\gamma_Y(\cdot)$, from (6), one can see that

$$\gamma_X(\cdot) \quad \gamma_Y(\cdot) \tag{9}$$

which implies that a positive $\gamma_X(\cdot)$ leads to a positive $\gamma_Y(\cdot)$. A negative $\gamma_X(\cdot)$ produces a negative $\gamma_Y(\cdot)$ since $\gamma_Y(\cdot)$ is, in fact, monotone increasing (see Proposition A.1 in Appendix A) and crosses zero at $= 0$ (the negativeness of $\gamma_Y(\cdot)$ when $\gamma_X(\cdot)$ 0 can also be deduced from the nondecreasing nature of via an inequality on page 20 of [41] for Gaussian variables).

Remark 2.1. The short- and long-range dependence properties of can be extracted from those of . Recall that a time series is short-range dependent (SRD) if $\sum_{=-\infty}^{\infty} \gamma_X(\cdot)$. According to one de nition, a series is long-range dependent (LRD) if $\gamma_X(\cdot) = \gamma_X(\cdot)^{2\ -1}$, where $(0\ 1\ 2)$ is the LRD parameter and is a slowly varying function at in nity [37]. The ACVF of such LRD series satis es $\sum_{=-\infty}^{\infty} \gamma_X(\cdot) =$. If is SRD, then so is by (9). On the other hand, if is LRD with parameter , then can be either LRD or SRD. The conclusion depends, in part, on the Hermite rank of $\gamma_X(\cdot)$, which is de ned as $= \min\ 1 : \ = 0$. Speci cally, if $(0\ (\ 1)\ 2\ )$, then is SRD; if $((\ 1)\ 2\ \ 1\ 2)$, then is LRD with parameter $(\ 1\ 2) + 1\ 2$ (see [37], Proposition 5.2.4).

The model in (1) admits the following structure: if and are independent, then so are and . It follows that if is stationary and -dependent, than both and must be th order moving-average time series. Unfortunately, no analogous autoregressive structure holds; in fact, if is a rst order autoregression, then may not be an autoregression of any order (this can be inferred from [28]).

Remark 2.2. The construction in (1) yields models with the most exible correlations possible for Corr($_1$ $_2$) for two variables $_1$ and $_2$ with the same marginal distribution . Indeed, let $_-$ = min Corr($_1$ $_2$) : $_1$ $_2$ and de ne $_+$ similarly with min replaced by max. Then, as shown in Theorem 2.5 of [42],

$$_+ = \text{Corr}(\ ^{-1}(\ )\ \ ^{-1}(\ )) = 1 \qquad _- = \text{Corr}(\ ^{-1}(\ )\ \ ^{-1}(1\ \ ))$$

where is a uniform random variable over (0 1). Since $\overset{\mathcal{D}}{=}$ ( ) and 1 $\overset{\mathcal{D}}{=}$ ( ) for a standard normal random variable , the maximum and minimum correlations $_+$ and $_-$ are indeed achieved with (1) when $_1 =$ $_2$ and $_1 =$ $_2$, respectively. The preceding statements are non-trivial for $_-$ only since $_+ = 1$ is attained whenever $_1 =$ $_2$. It is worthwhile to compare this to the discussion following (7). Finally, all correlations in ($_-$ $_+$) = ($_-$ 1) are achievable since ( ) in (7) is continuous in . The exibility of correlations for Gaussian copula models in the spatial context was also noted and studied in [21], especially in comparison to a class of hierarchical, e.g. Poisson, models.

The preceding remark settles autocovariance exibility issues for stationary count series. Flexibility is a concern when the series is negatively correlated, an issue arising, for example, with hurricane counts in [33] and chemical process counts in [26]. Since any general count marginal distribution can also be achieved, the model class is quite general.

2.2 Covariates

There are situations where stationarity is not desired. Such scenarios can often be accommodated by simple variants of the above setup. For concreteness, consider a situation where a vector $\mathbf{M}$ of non-random covariates is available to explain the series at time . If one wants to have the marginal distribution $_{\boldsymbol{\theta}(\ )}(\ )$, where $\boldsymbol{\theta}(\ )$ is a vector-valued

function of containing marginal distribution parameters, then simply set

$$= \ ^{-1}_{\boldsymbol{\theta}(\ )}(\ (\ )) \tag{10}$$

and reason as before. We do not recommend modifying for the covariates as this may bring process existence issues into play.

Generalized linear models link functions (not to be confused with ( ) in (6) (7)) can be used when parametric support set bounds are encountered. For example, a Poisson regression with correlated errors can be formulated via a parameter vector $\boldsymbol{\beta}$ of regression coe cients with $\boldsymbol{\theta}(\ ) = \mathbb{E}[\ ] = \exp(\boldsymbol{\beta}'\boldsymbol{M}\ )$. Here, the exponential link guarantees that the Poisson parameter is positive. The above construct requires the covariates to be non-random; should covariates be random, marginal distributions may change from $_{\boldsymbol{\theta}(\ )}$.

2.3 Particle filtering and state space model connections

This subsection studies the implications of the latent structure of our model, especially as it relates to SSMs and importance sampling approaches. This will be used to construct PF/SMC approximations of various quantities, and in goodness-of- t assessments. Our main reference is [14]. As in that monograph, let $_{0:} = \ _0 = \ _0 \ = \ , \ _{0:} = \ _0 = \ _0 \ = \ $, and ( ) and ( ) denote joint and conditional probabilities (or their densities, depending on the context). For example, ($_{0:}$ $_{0:}$) denotes the conditional density of $_{0:}$ given $_{0:}$. Similarly, let $\mathbb{E}[\ _{0:}]$ denote conditional expectation given $_{0:}$ The SSM formulation starts by specifying ($_{+1}$ $_{0:}$) and ( ). While is often rst order Markov, implying that ($_{+1}$ $_{0:}$) = ($_{+1}$), this is not necessary.

To specify ($_{+1}$ $_{0:}$) in our stationary Gaussian case, we compute the best one-step-ahead linear prediction of $_{+1}$ from $_{0:}$ given by $_{+1} = \ _0 \ + \ + \ _0$. The coe cients

, 0 , can be computed recursively in from the ACF of via the classical Durbin-Levinson (DL) or the Innovations algorithm, for examples. As a convention, we take $_0 = 0$. Let $^2 = \mathbb{E}[(\quad)^2]$ be the corresponding unconditional mean squared prediction error. With this notation,

$$(\ _{+1} \ _{0:}) \overset{\mathcal{D}}{=} \mathcal{N}(\ _{+1} \ \ ^2_{+1}) \tag{11}$$

where $_{+1} = \ _0 + \ + \ _0$. Again, does not have to be Markovian (of any order). On the other hand, with (1),

$$(\quad) = \ _{(\ t)}(\ ) = \begin{cases} 1 & \text{if} \quad = \ (\ ) \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

where $(\ )$ is a unit point mass at . The equations in (11) and (12) constitute the SSM representation of (1).

In inference and related tasks for SSMs, the basic goal is to compute the conditional expectation $\mathbb{E}[\ (\ _{0:})\ _{0:}]$ for some function . This is often carried out through an importance sampling algorithm such as sequential importance sampling (SIS), which generates independent particle trajectories $_{0:}$ 1 , from a proposal distribution $(\ _{0:} \ _{0:})$ and approximates the conditional expectation as

$$\mathbb{E}[\ (\ _{0:})\ _{0:}] \quad \sum_{=1} \ (\ _{0:})\tilde{} =: \mathbb{E}[\ (\ _{0:})\ _{0:}] \tag{13}$$

where

$$\tilde{} = \frac{(\ _{0:})}{\sum_{=1} \ (\ _{0:})} \qquad (\ _{0:}) = \frac{(\ _{0:} \ _{0:})}{(\ _{0:} \ _{0:})} \tag{14}$$

are the (normalized) importance weights (see [14] and [32]). Furthermore, in SIS,

$$\tilde{} \quad \tilde{}_{-1} \ (\ _{0:}) \qquad (\ _{0:}) = \frac{(\quad)\ (\quad_{0:-1})}{(\quad_{0:-1} \ _{0:})} \tag{15}$$

(see (1.6) in [14], which is adapted to a possibly non-Markov setting by replacing $(\quad_{-1})$ with $(\quad_{0:-1})$). The two probability terms in the numerator of $(\ _{0:})$ in (15) constitute the SSM, whereas the denominator relates to the proposal distribution.

We suggest the following proposal distribution and the resulting SIS algorithm for our model. Take

$$(\quad_{0:-1}\quad_{0:}) \overset{\mathcal{D}}{=} \mathcal{N}_{\ x_t}(\quad^2) \tag{16}$$

where $\mathcal{N}$ denotes a normal distribution restricted to the set , and

$$= \quad : \ ^{-1}(\quad_{-1}) \qquad ^{-1}(\quad) \tag{17}$$

The role of stems from the fact

$$= \quad (\ ) \tag{18}$$

(i.e., the count value is obtained if and only if ; see the expression (A.2) for $(\ )$). In particular, for generated from the proposal distribution (16), the term $(\quad)$ in the incremental weight $(\ _{0:})$ of (15) is always set to unity. The rest of the incremental weights are calculated as

$$(\ _{0:}) = \frac{(\quad_{0:-1})}{(\quad_{0:-1}\quad_{0:})} = \frac{^{-\frac{(z_t-\hat{z}_t)^2}{2r_t^2}}\ (2\quad^2)^{1\ 2}}{^{-\frac{(z_t-\hat{z}_t)^2}{2r_t^2}}\ [(2\quad^2)^{1\ 2}\quad \mathbb{P}(\quad(\quad^2)\quad_{\ t})]}$$

$$= \mathbb{P}(\mathcal{N}(\quad^2)\quad_{\ t}) = \left(\frac{^{-1}(\quad_{\ t})}{\quad}\right)\quad\left(\frac{^{-1}(\quad_{\ t-1})}{\quad}\right) =: \quad(\ ) \tag{19}$$

The choice of the proposal distribution is largely motivated by $\mathbb{P}(\quad=\quad) = 1\ _k(\quad)$ and the explicit form in (19) for the incremental weights $(\ _{0:})$. Optimality considerations are mentioned in Remark B.3.

The following steps summarize our SIS algorithm.

Sequential Importance Sampling (SIS): For 1 , where represents the number of particles, initialize the weight $_0 = 1$ and the latent series $_0$ by

$$_0 \overset{\mathcal{D}}{=} \mathcal{N}_{x_0}(0\ 1) \tag{20}$$

Then, recursively over $= 1$, perform the following steps:

1: Compute with the DL or other algorithm using the previously generated values of $_0$ $_{-1}$.

2: Update the series and the importance weight via

$$\overset{\mathcal{D}}{=} \mathcal{N}_{x_t}(\quad ^2) \quad = \quad _{-1} \quad (\quad) \tag{21}$$

where ( ) is de ned in (19).

Remark 2.3. For 1 , the constructed path $_{=0}$ is one of the independent particles used to approximate the conditional expectation in (13). Equation (21) ensures that for each , the path $_{=0}$ obeys the restriction ( ) = and matches the temporal structure of . These two properties show that $_{=0}$ is a realization of the latent Gaussian stationary series producing = for all . Finally, we note where the model parameters enter into the SIS algorithm. The marginal distribution parameters $\boldsymbol{\theta}$ enter through the form of in (19), whereas the temporal dependence parameters $\boldsymbol{\eta}$ enter through the one-step-ahead prediction coe cients 0 , in the calculation of in Step 1 of the algorithm, and through the prediction error .

To compute the model likelihood, several known formulas applicable in the (general) SIS setting are needed. The relation

$$\frac{(\ _{0:} \quad _{0:})}{(\ _{0:} \quad _{0:})} = \prod_{=0} \quad (\ _{0:}) \frac{(\ _0)}{(\ _{0:})}$$

produces

$$\mathbb{E}[\quad (\ _{0:}\ )\ \ _{0:}\ ]\ (\ _{0}) = \mathbb{E}[\ (\ _{0:}\ )\ \ _{0:}\ ]\ (\ _{0:}\ )$$

In particular (with $(\ )\quad 1$),

$$\mathbb{E}[\ \ ]\ (\ _{0}) = \ (\ _{0:}\ ) \tag{22}$$

To conduct prediction, we use Equation (1.2) in [14] to get

$$\mathbb{E}[\ (\ \ _{+1})\ \ _{0:}\ ] = \mathbb{E}\Big[\mathbb{E}[\ (\ \ (\ \ _{+1}))\ \ _{0:}\ ]\ \ _{0:}\Big] =: \mathbb{E}[\quad _{+1}(\ \ _{+1})\ \ _{0:}\ ] \tag{23}$$

where

$$_{+1}(\ ) = \mathbb{E}\big[\ (\ \ (\mathcal{N}(\quad ^{2}_{+1})))\big] = \int_{\mathbb{R}}\ (\ \ (\ _{+1}))\frac{1}{\sqrt{2\ \ ^{2}_{+1}}}\ ^{-\frac{(z_{t+1}-z)^2}{2r^2_{t+1}}}\quad _{+1} \tag{24}$$

since $_{+1}\ \ _{0:} \overset{\mathcal{D}}{=} \mathcal{N}(\ _{+1}\ \ ^{2}_{+1})$. In view of (23) and (13), the following prediction approximation arises:

$$\mathbb{E}[\ (\ \ _{+1})\ \ _{0:}\ ]\quad \sum_{=1} \text{——}\quad _{+1}(\ \ _{+1}) =: \mathbb{E}[\ (\ \ _{+1})\ \ _{0:}\ ]\qquad = \sum_{=1} \tag{25}$$

Appendix B further connects our model and algorithm to the popular GHK sampler, hidden Markov models (HMMs), and PF and SMC techniques.

The SIS algorithm has a fundamental weakness called weight degeneracy : as the algorithm propagates through an increasing number of iterations, a large number of the normalized weights become negligible. As a result. only a few particles contribute in the likelihood approximation. Following the developments in the SMC (see [14], [31] and [8]) and HMM literatures (Sections 10.4.1 and 10.4.2 in [13]), we modify the SIS algorithm by adding a resampling step (all future simulations and computations use resampling). **Sequential Importance Sampling with Resampling (SISR):** Proceed as in the SIS algorithm, but modify Step 2 and add a resampling Step 3 as follows:

2: Modify Step 2 of the SIS by setting

$$\tilde{} \overset{\mathcal{D}}{=} \mathcal{N}_{x_t}(\quad ^2) \quad \tilde{} = \quad _{-1} \quad (\quad) \quad \tilde{} \quad = \sum_{=1} \tilde{} \tag{26}$$

3: For each particle 1 , draw, conditionally and independently given ( ) 1 and $\tilde{}$, a multinomial trial 0 1 for each and with the success probabilities $\tilde{}$ $\tilde{}$ and set = $\tilde{}^{\,i}_t$ and = 1.

While the resampling step removes particles with low weights, mitigating degeneracy issues, it introduces additional estimator variance. We follow standard practice and resample only when the variance of the weights exceeds a certain threshold, quanti ed by the so-called *effective sample size* de ned as ESS( ) = $(\sum_{=1}($ $)^2)^{-1}$, and the resampling step is executed when ESS( ) 2 as in [15]. See also Section 2.5.3 in [31] for a justi cation of the ESS based on the Delta method.

3 Inference

The model in (1) contains the parameters $\boldsymbol{\theta}$ in the marginal count distribution and $\boldsymbol{\eta}$ in the dependence structure of . This section addresses inference questions, including parameter estimation and goodness-of- t assessment. Three methods are presented for parameter estimation: Gaussian pseudo-likelihood, implied Yule-Walker moment methods, and full likelihood. Gaussian pseudo-likelihood estimators, a time series staple, pretend that the series is Gaussian and maximize its Gaussian-based likelihood. These estimators only involve the mean and covariance structure of the series, are easy to compute, and will provide a comparative basis for likelihood estimators. They can also be used as initial

guesses in gradient step-and-search likelihood optimizations. Implied Yule-Walker techniques are moment based estimators applicable to the commonly encountered case where

is a causal autoregression. Likelihood estimators, the statistical gold standard and the generally preferred estimation technique, are based on the PF and SMC methods of the last section. Finally, we will not delve into a detailed statistical inference for the aforementioned methods: while consistency and asymptotic normality are expected in some of the examined cases (e.g. likelihood estimation with an autoregressive), a rigorous theoretical treatment is beyond the scope of this paper.

3.1 Gaussian pseudo-likelihood estimation

As in Section 2.3, we work with observations for the times 0 and set $\mathbf{X} = (\ _0 \quad)'$. Denote the likelihood of the model in (1) by

$$\mathcal{L}\ (\boldsymbol{\theta}\ \boldsymbol{\eta}) = \mathbb{P}(\ _0 = \ _0 \ \ _1 = \ _1 \quad = \ ) = (\ _{0:}\ ) \tag{27}$$

While this likelihood is a multivariate normal probability, it is di cult to calculate or approximate when is large. For most count model classes, true likelihood estimation is di cult to conduct as joint distributions are generally intractable [11]. While Section 3.3 below devises a well performing PF/SMC likelihood approximation (see also [39]), we rst consider a simple Gaussian pseudo-likelihood (GL) approach. In a pseudo GL approach, parameters are estimated via

$$(\boldsymbol{\theta}\ \boldsymbol{\eta}) = \underset{\boldsymbol{\theta}\ \boldsymbol{\eta}}{\operatorname{argmax}} \frac{^{-\frac{1}{2}(\mathbf{X}-\boldsymbol{\mu}_{\boldsymbol{\theta}})'\ \ _T(\boldsymbol{\theta}\ \boldsymbol{\eta})^{-1}(\mathbf{X}-\boldsymbol{\mu}_{\boldsymbol{\theta}})}}{(2\ )^{(\ +1)\ 2}\quad (\boldsymbol{\theta}\ \boldsymbol{\eta})^{\ 1\ 2}} \tag{28}$$

where $\boldsymbol{\mu}_{\boldsymbol{\theta}} = (\ _{\boldsymbol{\theta}} \quad _{\boldsymbol{\theta}})'$ is a $(\ +1)$-dimensional constant mean vector. These estimators maximize the series likelihood assuming the data are Gaussian, each component having

mean $\mu_{\boldsymbol{\theta}}$, and all components having covariance matrix $\Sigma(\boldsymbol{\theta}, \boldsymbol{\eta}) = (\gamma(i-j))_{i,j=0}^{n-1}$. Time series analysts have been maximizing Gaussian pseudo likelihoods for decades, regardless of the series' marginal distribution, with often satisfactory performance. The next section and Appendix C present a case where this approach works reasonably well, and one where it does not. For large n, the pseudo GL approach is equivalent to least squares estimation, where the sum of squares $\sum_{i=0}^{n-1}(X_i - \mathbb{E}[X_i \mid X_0, \ldots, X_{i-1}])^2$ is minimized (see Chapter 8 in [6]). The covariance structure of $\mathbf{X}$ was efficiently computed in Section 2; the mean $\mu_{\boldsymbol{\theta}}$ is usually explicitly obtained from the marginal distribution posited. Numerical optimization of (28) yields a Hessian matrix that can be inverted to obtain standard errors for the model parameters. These standard errors can be asymptotically corrected for distributional misspecification via the sandwich methods of [19].

3.2 Implied Yule-Walker estimation for latent AR models

Suppose that $\mathbf{X}$ follows the causal AR(p) model $X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t$, where $\mathbf{Z}$ consists of IID $\mathcal{N}(0, \sigma^2)$ variables. Here, σ^2 depends on the autoregressive coefficients $\phi_1, \ldots, \phi_p$ in a way that induces $\mathbb{E}[X_t^2] = 1$. The Yule-Walker equations are

$$\boldsymbol{\phi} = \boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma} \tag{29}$$

where $\boldsymbol{\Gamma} = (\gamma(i-j))_{i,j=1}^{p}$, $\boldsymbol{\gamma} = (\gamma(1), \ldots, \gamma(p))'$, and $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)'$. From (6), note that

$$\gamma(h) = \sigma^{-1}(\gamma_Z(h)) \tag{30}$$

the inverse being justified via the strictly increasing nature of $\gamma(h)$ in $\gamma_Z(h)$.

Equations (29) and (30) suggest the following estimation procedure. First, estimate the CDF parameter $\boldsymbol{\theta}$ directly from the counts; standard methods (e.g. method of moments) are

typically available for this task. The estimated parameter $\boldsymbol{\theta}$ de nes an estimated link $\quad(\ )$ through its estimated power series coe cients. From a numerical power series reversion procedure, one can now e ciently construct the inverse estimator $\quad^{-1}(\ )$.

Next, in view of (30) and (29), set

$$(\ ) = \quad^{-1}(\quad(\ )) \qquad \boldsymbol{\phi} = \boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma} \tag{31}$$

where $\quad(\ )$ is the lag- sample autocorrelation of $\qquad$, and $\boldsymbol{\Gamma}$ and $\boldsymbol{\gamma}$ are de ned analogously to the above using $\quad(\ )$ in place of $\quad(\ )$.

3.3 Particle filtering and sequential Monte Carlo likelihoods

Using (23) and its notation, the true likelihood in (27) is

$$\mathcal{L}\ (\boldsymbol{\theta}\ \boldsymbol{\eta}) = \ (\ _0)\prod_{=1}\ (\qquad_{0:\ -1}) = \ (\ _0)\prod_{=1}\mathbb{E}[1_{\{\ _s\}}(\quad)\ _{0:\ -1}] = \ (\ _0)\prod_{=1}\mathbb{E}[\quad(\quad)\ _{0:\ -1}] \tag{32}$$

where (23) was used with $\ _{1_{\{x_s\}}}(\ ) = \quad(\ )$ and $\quad(\ )$ is de ned and numerically computed akin to (19). The particle approximation of the likelihood is then

$$\mathcal{L}\ (\boldsymbol{\theta}\ \boldsymbol{\eta}) = \ (\ _0)\prod_{=1}\mathbb{E}[\quad(\quad)\ _{0:\ -1}]; \tag{33}$$

this uses the notation in (13) and supposes that the particles are generated by one of the methods in Section 2.3. The approximate PF maximum likelihood estimates satisfy

$$(\boldsymbol{\theta}\ \boldsymbol{\eta}) = \underset{\boldsymbol{\theta}\ \boldsymbol{\eta}}{\operatorname{argmax}}\,\mathcal{L}\ (\boldsymbol{\theta}\ \boldsymbol{\eta}) \tag{34}$$

Remark 3.1. With the SIS algorithm, (33) reduces to

$$\mathcal{L}\ (\boldsymbol{\theta}\ \boldsymbol{\eta}) = \ (\ _0)\frac{1}{}\sum_{=1} \tag{35}$$

which is consistent with (22). The work [35] also essentially implements (35). In contrast to [35], our approach includes a resampling step in the likelihood approximations, considers other estimation approaches (pseudo GL and implied Yule-Walker), and provides model diagnostic tools more speci c to count series (the PIT histograms in Section 3.4 below).

To optimize the estimate $\mathcal{L}\ (\boldsymbol{\theta}\ \boldsymbol{\eta})$, we employ a large number of particles (growing linearly with) and common random number (CRN) techniques, a standard practice that serves to smooth $\mathcal{L}\ (\boldsymbol{\theta}\ \boldsymbol{\eta})$ somewhat by expressing its random quantities through parameter-dependent transformations of uniform random variables that remain constant for likelihood evaluations across distinct parameters. While the CRN procedure works well in SIS, it fails to ward against discontinuous $\mathcal{L}\ (\boldsymbol{\theta}\ \boldsymbol{\eta})$ in our preferred SISR algorithm. An elegant solution to this issue for univariate state processes is proposed in [34]: rst re-order the (real-valued) particles and then replace the discontinuous resampling CDF with a piecewise linear approximation. More recent and well performing (but less straightforward) approaches such as the sequential quasi Monte Carlo and the SMC^2 algorithm are reviewed in detail in Chapters 13, 14, and 18 of [8] (see also the Chapter 19 references on controlled sequential Monte Carlo methods). We do not pursue these issuees further here.

In our numerical implementations, gradient-free algorithms from the R package optimx [36] are used, which follows standard practices in optimizing noisy objective functions. These routines allow for boundary constraints and performed well in modest computing times for our sample sizes. On the other hand, we found less success with the more popular gradient-based quasi-Newton algorithm L-BFGS-B (gradients were computed via nite di erences) as convergence instabilities and high-variance estimates were encountered. However, promising recent developments for optimizing noisy objectives in [4] and [38] were not explored. A comprehensive investigation of these approaches and of the rich gradient-

based SMC inference literature for our framework as in [27] is deferred to future work.

3.4 Model diagnostics

The goodness-of- t of count models is commonly assessed through probability integral transform (PIT) histograms and related tools [10, 29]. These are based on the predictive distributions of , de ned at time by

$$( ) = \mathbb{P}(\quad _0 = \ _0 \quad _{-1} = \ _{-1}) = \mathbb{P}(\quad _{0:\ -1}) \qquad 0 \ 1 \tag{36}$$

This quantity can be estimated through the PF/SMC methods in Section 2.3 as

$$( ) = \sum_{=0} \mathbb{E}[1_{\{ \}}(\ ) \ _{0:\ -1}] = \sum_{=0} \mathbb{E}[\ _{1_{\{\ell\}}} (\ ) \ _{0:\ -1}] \tag{37}$$

which uses (24) and (25) and supposes that the particles are generated by the SIS, SISR, or other algorithms. Similar to $_{1_{\{x_s\}}}( ) = \ ( )$, note that $_{1_{\{x\}}}( ) = \tilde{} \ ( )$, where

$$\tilde{} \ ( ) = \left(\frac{ ^{-1}(\ )}{ }\right) \ \left(\frac{ ^{-1}(\ _{-1})}{ }\right) \tag{38}$$

and $\tilde{}_{\ t}( ) = \ ( )$.

The (non-randomized) sample mean PIT is de ned as

$$\bar{}( ) = \frac{1}{ + 1} \sum_{=0} \ (\ ) \qquad [0 \ 1] \tag{39}$$

where

$$(\ ) = \begin{cases} 0 & \text{if} \quad (\ 1) \\ \frac{ - \ _t(\ -1)}{ _t(\ )- \ _t(\ -1)} & \text{if} \quad (\ 1) \\ 1 & \text{if} \quad (\ ) \end{cases} \qquad ( ) \tag{40}$$

which is estimated by replacing by in practice. The PIT histogram with bins is de ned as a histogram with the height ¯( ) ¯((1)) for bin 1 .

Another possibility considers model residuals based on

$$= \mathbb{E}[=] = \frac{\exp(\quad ^{-1}(\quad _{t-1})^2 \ 2) \quad \exp(\quad ^{-1}(\quad _t)^2 \ 2)}{\overline{2 \ }(\quad _t \quad _{t-1})} \tag{41}$$

which is the estimated mean of the latent Gaussian process at time given only (not the entire past), where (41) follows by direct calculations for the model (1) (using the estimated parameters $\boldsymbol{\theta}$ of the marginal distribution in the s). For a tted underlying time series model with parameter $\boldsymbol{\eta}$, the residuals are then de ned as the standard time series residuals of this model tted to the series , after centering by the sample mean.

3.5 Nonstationarity and covariates

As discussed in Section 2.2, covariates can be accommodated by allowing a time-varying parameter $\boldsymbol{\theta}$ in the marginal distribution. With covariates, $\boldsymbol{\theta}$ at time is denoted by $\boldsymbol{\theta}( )$. The GL and PF/SMC procedures are modi ed for $\boldsymbol{\theta}( )$ as follows.

For the GL procedure, the covariance Cov($_1$ $_2$) = Cov($_{\boldsymbol{\theta}(_1)}(_1)$ $_{\boldsymbol{\theta}(_2)}(_2)$) is needed, where is subscripted to signify dependence on $\boldsymbol{\theta}( )$. But as in (5),

$$\text{Cov}(\quad _1 \quad _2) = \text{Cov}(\quad _{\boldsymbol{\theta}(_1)}(\quad _1) \quad _{\boldsymbol{\theta}(_2)}(\quad _2)) = \sum_{=1}^{\infty} \ ! \ _{\boldsymbol{\theta}(_1)} \quad _{\boldsymbol{\theta}(_2)} \quad (_1 \quad _2) \tag{42}$$

where again, the subscript $\boldsymbol{\theta}( )$ is added to the s to indicate dependence on . Numerically, evaluating (42) is akin to the task in (5); in particular, both calculations are based on the Hermite coe cients .

For the PF/SMC approach, the modi cation is somewhat simpler: one just needs to replace $\boldsymbol{\theta}$ by $\boldsymbol{\theta}( )$ at time when generating the underlying particles. For example, for the

SIS algorithm, $\boldsymbol{\theta}(\ )$ enters only through the s in (19), (20), and (21). This is because the covariates enter only through $\boldsymbol{\theta}$, the parameter controlling marginal distributions.

4 A simulation study

To evaluate our estimation methods, a simulation study considering several marginal distributions and dependence structures was conducted. Here, the classic Poisson count distribution $\mathcal{P}$ is examined (mixed Poisson and negative binomial simulations are presented in Appendix C), with taken from the ARMA( ) class. All simulation cases are replicated 200 times for three distinct series lengths: $= 100$ 200, and 400. For notation, estimates of a parameter from Gaussian pseudo-likelihood (GL), implied Yule-Walker (IYW), and PF/SMC methods are denoted by , , and , respectively.

We now consider the classical case where has a Poisson marginal distribution for each with mean 0. To obtain , the AR(1) process $= {}_{-1} + (1 \quad {}^{2})^{1\ 2}$, was simulated and transformed via (1) with $= \mathcal{P}$; $\mathbb{E}[\ {}^{2}]$ 1 was induced by taking Var( ) 1. Twelve parameter schemes resulting from all combinations of 2 5 10 and 0 25 0 75 were considered.

Figure 1 displays box plots of the parameter estimates when $= 2$. In estimating , all methods perform reasonably well. When the lag-one correlation in (and hence also that in) is negative (right panel), , , and have smaller variability than the positively correlated case (left panel note the di erent y-axis scales on the panels). This is expected: the mean of is , and the variability of the sample mean, one good estimator of the mean for a stationary series, is smaller for negatively correlated series than for positively correlated ones. Note that is biased toward zero for both negatively and

positively correlated series, whereas $\hat{\phi}_{IYW}$ and $\hat{\phi}_{PF}$ only show bias when ϕ is positive for the sample sizes $T = 100$ and $T = 200$. Overall, the PF/SMC estimates were the least biased. All estimates of ϕ have roughly similar variances. Simulations with $\lambda = 5$ and $\lambda = 10$ produced analogous results with smaller values of λ yielding less variable estimates. This is again expected as the variance of the Poisson distribution is also λ. Graphics of these box plots are omitted for brevity's sake.

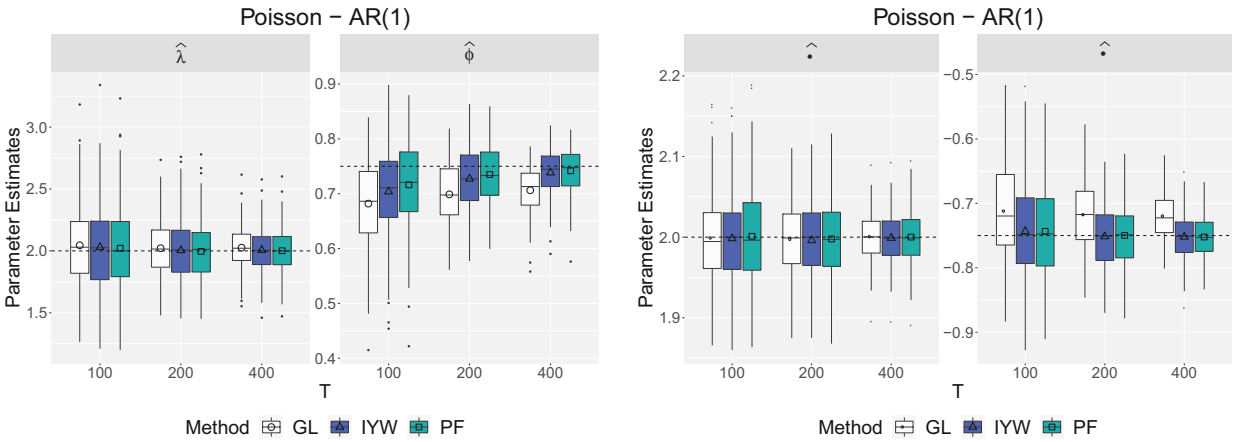

Figure 1: *Gaussian likelihood, implied Yule-Walker, and PF/SMC parameter estimates for 200 synthetic Poisson–AR(1) series with lengths* $T = 100, 200,$ *and* 400. *The true parameter values (indicated by horizontal dashed lines) are* $\lambda = 2$ *and* $\phi = 0.75$ *(left panel), and* $\lambda = 2$ *and* $\phi = -0.75$ *(right panel).*

5 An application

This section applies our methods to a weekly count series of product sales at Dominicks Finer Foods, a now defunct U.S. grocery chain that operated in Chicago, IL and adjacent areas

from 1918 - 2013. Soft drink sales of an unnamed brand from a single store will be analyzed over a two-year span commencing on September 10, 1989. The series is plotted in Figure 2 (leftmost plot) and is part of a large and well-studied retail dataset, publicly available at https://www.chicagobooth.edu/research/kilts/datasets/dominicks (Source: The James M. Kilts Center for Marketing, University of Chicago).[1]. Our goal here is not an in-depth retail

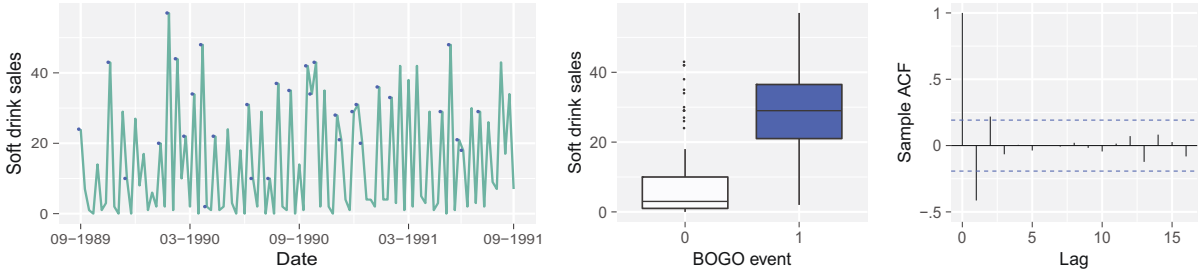

Figure 2: *Left: Weekly sales of a soft drink product sold at a single store of the grocery store Dominick's Finer Foods from 09-10-1989 to 09-03-1991. The dots indicate the weekly sales were at least one "Buy one and get one free" (BOGO) sales promotion event took place. Middle: Boxplots of sales grouped by the BOGO covariate (0: weekly sales with no BOGO event, 1: weekly sales with at least one BOGO day during the week). Right: Sample ACF of the series with 95% pointwise bands for zero correlation.*

analysis, but to illustrate our methods with a real world example of an overdispersed time series of small counts that has negative autocorrelation and dependence on a covariate.

The covariate we use is a zero-one "buy one get one free" (BOGO for short) sales promotion event S_t, $S_t = 1$ implying that the BOGO promotion was offered at least one day during week t. The dots in the left plot of Figure 2 signify that the week had at

[1]In the dataset manual, the series in Figure 2 (leftmost plot) is the sales of the product with universal product code (UPC) 4640055081 from store 81.

least one BOGO day. The middle plot shows the soft drinks sales distribution grouped by , visually suggesting that a BOGO event increases soft drink sales. The rightmost plot shows the sample ACF of the series and reveals negative dependence at lag one. The lag one sample autocorrelation of the residuals after a linear regression of the series on the BOGO covariate is also negative, but comparatively smaller in magnitude.

To model overdispersion, negative binomial and generalized Poisson marginal distributions will be considered. Although similar, these two distributions can yield di erent conclusions [25]. Following standard generalized linear modeling practice, both distributions are parametrized via the series mean (although our setup allows covariates to enter through other parameters as well). More speci cally, for the negative binomial marginal, the standard pair ( ) used in Appendix C is now mapped to the parameter pair ( ), where = (1) is the mean of the process and = 1 is the overdispersion parameter. Similarly, the generalized Poisson distribution of Appendix C is parametrized through the pair ( ) as in [17], relation (2.4). In this parametrization, is the mean of the series, whereas the sign of controls the type of dispersion, with positive values indicating overdispersion. To incorporate the BOGO covariate into the model, the mean of the series is allowed to depend on time through the typical GLM log-link = exp($_0$ + $_1$), while the parameters and are kept xed in time .

An exploratory examination of the sample ACFs and PACFs of the series along with diagnostic plots of residuals obtained by tting all ARMA( ) models with 5 suggest an AR(3) model as a suitable choice for . Table 1 in Appendix D shows the AICc and BIC for both marginal distributions obtained via PF/SMC and GL methods (we omit IYW results for simplicity). The AR(3) model was selected by AICc and BIC in both ts. Interestingly, both the sample ACF and PACF of the series show one large non-zero value

at lag one, but relatively smaller values at other lags (except perhaps the lag two value, which barely exceeds the 95% 1 96 $\overline{}$ dashed con dence threshold for zero correlation).

We also considered a white noise latent series (labeled as WN Table 1 in Appendix D), which renders our model a standard GLM. The PF/SMC WN estimates from both distributions (omitted here for brevity) closely agree with parameter estimates obtained from exact generalized linear model ts (using, for example, functions from the R package MASS). As expected, the WN model yielded the highest AICc and BIC values among all considered dependence structures, thus con rming the need for a model with temporal dependence.

Table 1 shows parameter estimates and standard errors from tting a negative binomial-AR(3). (Table 2 in Appendix D is for a generalized Poisson-AR(3) model.) All marginal distributions and estimation methods yielded $_1$ 0. Although a formal asymptotic theory is beyond the scope of our presentation here, asymptotic normality is expected. Assuming this, the PF/SMC standard errors (the ones believed most trustworthy) suggest that all parameters are signi cantly non-zero at level 95%. The ndings suggest the negative binomial distribution is preferred over the generalized Poisson, that the correlation in the series at lag one is negative, and that a BOGO event indeed increases sales.

	1 2 3 0 1

Table 1:

Turning to residual diagnostics, the plots in Figure 3 for the negative binomial-AR(3)

fit suggest that the model has captured both the marginal distribution and the dependence structure. The residuals here were computed using (41).

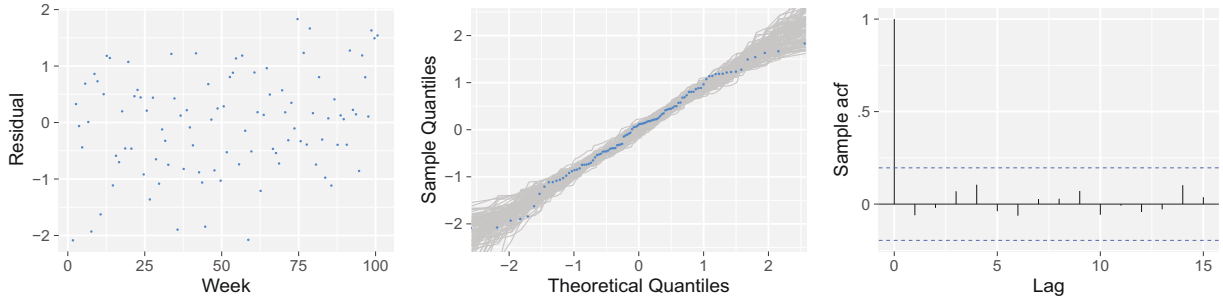

Figure 3: *The leftmost plot displays the estimated residuals against time. The middle graph is a QQ plot for normality of the estimated residuals. The shaded region in the QQ plot shows100 realizations from a normal distribution with size, mean and standard deviation matching the residual sample counterparts. The right plot displays the sample autocorrelations of the estimated residuals.*

We next assess the predictive ability of the two fits via the non-randomized histograms shown in Figure 4 and discussed in detail in Section 3.4. We selected ten bins at the points $h/10, h = 1, \ldots, 10$ as is typical in the literature. The negative binomial PIT plot suggests a satisfactory predictive ability with most bar heights being close to 0.1 (1 over the number of bins). In comparison, the generalized Poisson fit deviates more from the uniform distribution, with somewhat more pronounced peaks and valleys. We remind the reader here that PIT plots are known to be sensitive for smaller series lengths. Quantifying this uncertainty (for each bin) through a statistical test is beyond the scope of this paper. Nevertheless, we gauged the variability of the uniform distribution's bin heights through a small experiment. Specifically, 500 synthetic realizations of sample size $T = 104$ were generated and the percentiles of all bin heights were collected. The 5th and 95th percentiles

ranged in the intervals (0 048 0 058) and (0 145 0 154) respectively, suggesting that the peaks and valleys of the negative binomial PIT plot (which are within these percentiles) are mild; that is, uniformity is plausible and the marginal distribution ts seems adequate.

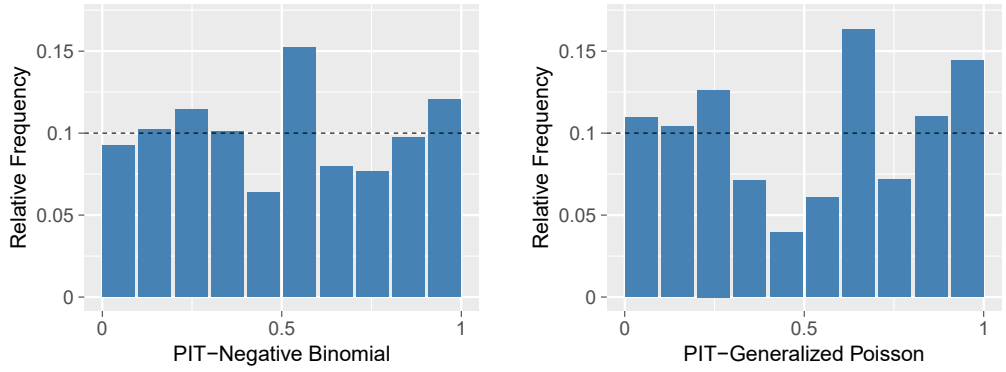

Figure 4:

6 Conclusions and comments

This paper developed the theory and methods for a stationary count time series model made from a latent Gaussian process. By using Hermite expansions, a very general model class was devised. In particular, the autocorrelations in the series can be positive or negative, and in a pairwise sense, span the range of all achievable correlations. The series can have any marginal distribution desired, thereby improving classical DARMA and INARMA count time series methods. On inferential levels, autocovariances of the model were extracted from Hermite expansions, allowing for Gaussian pseudo-likelihood and implied Yule-Walker inference procedures. A PF/SMC likelihood approach was also developed and produced estimators that were demonstrated to outperform the Gaussian

pseudo-likelihood and implied Yule-Walker estimators in most cases. These results complement the importance sampling methods for copula likelihoods in [39]. The methods were used in a simulation study and were applied in a regression analysis of a count series of weekly grocery sales that exhibited overdispersion, a negative lag one correlation, and dependence on a buy one get one free covariate. Model fits and predictive abilities of the methods were illustrated with generalized Poisson and negative binomial marginal distributions.

While the paper provides a reasonably complete treatment for count time series models, additional research is needed. Some statistical issues, like asymptotic normality of parameter estimators, were not addressed here. PF/SMC algorithms that optimize model likelihoods, which can be unwieldy, also merit further exploration. The paper only considers univariate methods. Multivariate count time series models akin to those in [40] could be developed by replacing the univariate with a multivariate Gaussian process $\mathbf{Z}$, whose components have a standard normal marginal distribution, but are cross-correlated for each fixed . The details for such a construction would proceed akin to the methods developed here. Also, while the count case is considered here, the same methods will produce stationary time series having any general prescribed continuous distribution. Finally, the same methods should prove useful in constructing spatial and spatio-temporal processes having any prescribed marginal distribution. While [12, 21] recently addressed this issue in the spatial setting, additional work is needed, including exploring spatial Markov properties and likelihood evaluation techniques. To the best of our knowledge, no comprehensive analogous work has been conducted for space-time count modeling to date.

References

[1] Asmussen, S. (2014). Modeling and performance of bonus-malus systems: stationarity versus age-correction. *Risks 2*, 49 73.

[2] Belyaev, M., E. Burnaev, and Y. Kapushev (2015). Gaussian process regression for structured data sets. In A. Gammerman, V. Vovk, and H. Papadopoulos (Eds.), *Statistical Learning and Data Sciences: Third International Symposium, SLDS 2015*. Switzerland: Springer International Publishing.

[3] Benjamin, M. A., R. A. Rigby, and D. M. Stasinopoulos (2003). Generalized autoregressive moving average models. *Journal of the American Statistical Association 98*, 214 223.

[4] Berahas, A. S., R. H. Byrd, and J. Nocedal (2019). Derivative-free optimization of noisy functions via quasi-Newton methods. *SIAM Journal on Optimization 29*, 965 993.

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

[6] Brockwell, P. J. and R. A. Davis (1991). *Time Series: Theory and Methods* (Second ed.). New York City: Springer-Verlag.

[7] Chen, H. (2001). Initialization of NORTA: Generation of random vectors with speci ed marginals and correlations. *Inform Journal on Computing 13*, 312 331.

[8] Chopin, N. and O. Papaspiliopoulos (2020). *An Introduction to Sequential Monte Carlo Methods*. New York City: Springer.

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

[10] Czado, C., T. Gneiting, and L. Held (2009). Predictive model assessment for count data. *Biometrics 65*, 1254 1261.

[11] Davis, R. A., S. H. Holan, R. B. Lund, and N. Ravishanker (Eds.) (2016). *Handbook of Discrete-Valued Time Series*. Boca Raton, Florida, USA: CRC Press.

[12] De Oliveira, V. (2016). Hierarchical Poisson models for spatial count data. *Journal of Multivariate Analysis 122*, 393 408.

[13] Douc, R., E. Moulines, and D. S. Sto er (2014). *Nonlinear Time Series: Theory, Methods, and Applications with R Examples*. Boca Raton, Florida, USA: CRC Press.

[14] Doucet, A., N. De Freitas, and N. Gordon (2001). An introduction to sequential Monte Carlo methods. In *Sequential Monte Carlo Methods in Practice*, pp. 3 14. New York City: Springer.

[15] Doucet, A. and A. M. Johansen (2009). A tutorial on particle ltering and smoothing: Fifteen years later. *Handbook of Nonlinear Filtering 12*, 656 704.

[16] Dumsmuir, W. T. M. (2016). Generalized linear autoregressivee moving-averge models. In *Handbook of Discrete-valued Time Series*, pp. 51 76. Boca Raton, Florida, USA: CRC Press.

[17] Famoye, F. (1993). Restricted generalized Poisson regression model. *Communications in Statistics-Theory and Methods 22*, 1335 1354.

[18] Fokianos, K. (2012). Count time series models. In *Handbook of Statistics*, Volume 30, pp. 315 347. Amsterdam: Elsevier.

[19] Freedman, D. (2006). On the so-called Huber sandwich estimator and robust standard errors . *The American Statistician 60*, 299 302.

[20] Grigoriu, M. (2007). Multivariate distributions with speci ed marginals: applications to wind engineering. *Journal of Engineering Mechanics 133*, 174 184.

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

[22] Han, Z. and V. De Oliveira (2020). Maximum likelihood estimation of Gaussian copula models for geostatistical count data. *Communications in Statistics - Simulation and Computation 49*, 1957 1981.

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

[24] Joe, H. (1996). Time series models with univariate margins in the convolution-closed in nitely divisible class. *Journal of Applied Probability 33*, 664 677.

[25] Joe, H. and R. Zhu (2005). Generalized Poisson distribution: the property of mixture of Poisson and comparison with negative binomial distribution. *Biometrical Journal: Journal of Mathematical Methods in Biosciences 47*, 219 229.

[26] Kachour, M. and J. F. Yao (2009). First order rounded integer valued autoregressive (RINAR(1)) processes. *Journal of Time Series Analysis 30*, 417 448.

[27] Kantas, N., A. Doucet, S. S. Singh, J. Maciejowski, and N. Chopin (2015). On particle methods for parameter estimation in state-space models. *Statistical Science 30*, 328 351.

[28] Kedem, B. (1980). Estimation of the parameters in stationary autoregressive processes after hard limiting. *Journal of the American Statistical Association 75*, 146 153.

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

[30] Lennon, H. (2016). *Gaussian copula modelling for integer-valued time series.* Ph. D. thesis, The University of Manchester.

[31] Liu, J. S. (2008). *Monte Carlo Strategies in Scientific Computing.* New York City: Springer Science & Business Media.

[32] Liu, J. S. and R. Chen (1998). Sequential Monte Carlo methods for dynamic systems. *Journal of the American Statistical Association 93*, 1032 1044.

[33] Livsey, J., R. B. Lund, S. Kechagias, and V. Pipiras (2018). Multivariate integer-valued time series with exible autocovariances and their application to major hurricane counts. *Annals of Applied Statistics 12*, 408 431.

[34] Malik, S. and M. K. Pitt (2011). Particle lters for continuous likelihood evaluation and maximisation. *Journal of Econometrics 165*, 190 209.

[35] Masarotto, G. and C. Varin (2012). Gaussian copula marginal regression. *Electronic Journal of Statistics 6*, 1517 1549.

[36] Nash, J. C. and R. Varadhan (2011). Unifying optimization algorithms to aid software system users: optimx for R. *Journal of Statistical Software 43*, 1 14.

[37] Pipiras, V. and M. S. Taqqu (2017). *Long-Range Dependence and Self-Similarity*, Volume 45. Boca Raton, Florida, USA: Cambridge University Press.

[38] Shi, H.-J. M., M. Q. Xuan, F. Oztoprak, and J. Nocedal (2021). On the numerical performance of derivative-free optimization methods based on nite-di erence approximations. *arXiv preprint arXiv:2102.09762*.

[39] Smith, M. S. and M. A. Khaled (2012). Estimation of copula models with discrete margins via Bayesian data augmentation. *Journal of the American Statistical Association 107*, 290 303.

[40] Song, P., M. Li, and P. Zhang (2013). Vector generalized linear models: a Gaussian copula approach. In P. Jaworski, F. Durante, and W. Hardle (Eds.), *Copulae in Mathematical and Quantitavie Finance*. Heidelberg, Germany: Springer.

[41] Tong, Y. L. (1990). *The Multivariate Normal Distribution*. New York City: Springer-Verlag.

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

[43] Zheng, T., H. Xiao, and R. Chen (2015). Generalized ARMA models with martingale di erence errors. *Journal of Econometrics 189*, 492 506.