On an integer-valued stochastic intensity model for time series of counts

Aknouche, Abdelhakim and Dimitrakopoulos, Stefanos

USTHB and Qassim University, University of Leeds

1 January 2020

On an integer-valued stochastic intensity model for time series of counts

Abdelhakim Aknouche* and Stefanos Dimitrakopoulos[1]**

*Faculty of Mathematics, University of Science and Technology Houari Boumediene (Algeria) and Qassim University (Saudi Arabia)

**Economics Division, Leeds University, UK

Abstract

We propose a broad class of count time series models, the mixed Poisson integer-valued stochastic intensity models. The proposed specification encompasses a wide range of conditional distributions of counts. We study its probabilistic structure and design Markov chain Monte Carlo algorithms for two cases; the Poisson and the negative binomial distributions. The methodology is applied to simulated data as well as to various data sets. Model comparison using marginal likelihoods and forecast evaluation using point and density forecasts are also considered.

Keywords: Markov chain Monte Carlo, mixed Poisson, parameter-driven models

JEL CODE: C00, C10, C11, C13, C22

[1]Correspondence to: Stefanos Dimitrakopoulos, Economics Division, Leeds University Business School, Leeds University, UK, E-mail: s.dimitrakopoulos@leeds.ac.uk, Tel: +44(0) 113 3435637. **The present version of the paper is a substantial improvement of the MPRA paper (MPRA:91962) titled: "Integer-valued stochastic volatility".**

1 Introduction

Nowadays, time series count data models have a wide range of applications in many fields (finance, economics, environmental and social sciences). The analysis of this type of models is still an active area (Davis et al., 2016; Weiß, 2017), as numerous models and methods have been proposed to account for the main characteristics of count time series (such as, overdispersion, underdispersion, and excess of zeros).

Many count time series models are often related to the Poisson process with a given parametric intensity. Following the general terminology by Cox (1981), these models can be classified into observation-driven and parameter-driven, depending on whether the dependence structure of counts is induced by an observed or a latent process, respectively.

One way of introducing serial correlation in count time series is through a dynamic equation for the intensity parameter, which may evolve according to an observed or an unobserved process. Observation-driven models of counts with a dynamic specification for the intensity parameter are mainly represented by integer-valued generalized autoregressive conditional heteroscedastic processes ($INGARCH$). This paper deals with the theory and inference of parameter-driven models of counts with a dynamic specification for the intensity parameter.

As is well known, $INGARCH$ processes (Grunwald et al., 2000; Rydberg and Shephard, 2000; Ferland et al., 2006; Fokianos et al., 2009; Doukhan et al., 2012; Christou and Fokianos, 2014; Chen et al., 2016; Davis and Liu, 2016; Ahmad and Francq, 2016), are easier to interpret and estimate by maximum likelihood-type methods. They are also convenient for forecasting purposes, but it has been quite difficult to establish their stability properties; see Fokianos et al., (2009), Davis and Liu (2016) and Aknouche and Francq (2021).

In contrast, parameter-driven models with a dynamic equation for the intensity parameter[1] (Davis and Rodriguez-Yam, 2005; Frühwirth-Schnatter et al., 2006; Jung et al., 2006; Frühwirth-Schnatter et al, 2009; Barra et al, 2018; Sørensen, 2019), although they do not admit a weak $ARMA$ representation, are generally of a simple structure and offer a great deal of flexibility in representing dynamic dependence (Davis and Dunsmuir, 2016). However, their estimation by the maximum likelihood method is computationally very demanding, if not infeasible. In general, these models are estimated by Bayesian Markov chain Monte Carlo (MCMC) and Expectation–Maximization (EM)-type algorithms.

Focusing on such type of models, we propose the mixed Poisson integer-valued stochastic intensity models ($INSI$). This class of models encompasses a large number of conditional distributions of counts and is formulated by considering a mixed Poisson process (Mikosch, 2009), for which the logarithm of the latent conditional mean parameter (intensity) follows a first-order (drifted) autoregressive model, which in turn, is driven by independent and identically (not necessarily Gaussian) distributed innovations.

Although we focus on the mixed Poisson $INSI$ model, we show that the present framework can be easily generalized to account for larger classes of conditional distributions. Various INSI models can correspond to conditional distributions (e.g. the exponential family) that do not necessarily belong to the class of mixed Poisson INSI processes. Furthermore, since the INSI can be seen as an alternative to the INGARCH process, the present work is also related to observation-driven models.

[1]The literature on time series of counts has also put forward parameter-driven models, which do not consider a dynamic equation for the latent intensity parameter (Zeger, 1988; Davis et al., 1999, 2000; Hay and Pettitt, 2001; Davis and Wu, 2009). In this case, the parameter-driven models are constructed based on a particular conditional distribution of counts (Poisson, negative binomial, integer-valued exponential family), given some covariates and an intensity parameter.

We study the probabilistic path properties of the mixed Poisson $INSI$ model, such as ergodicity, mixing, covariance structure and existence of moments. Moreover, by construction, the proposed model leads to an intractable likelihood function, as it depends on high-dimensional integrals. Yet, conditionally on the intensity parameter, the likelihood function has a closed form and parameter estimation can be achieved by MCMC methods. To demonstrate that, we consider two conditional distributions that belong to the family of the mixed Poisson $INSI$ process; the Poisson $INSI$ model (P-$INSI$) and the negative binomial $INSI$ model (NB-$INSI$).

The difficult part in the construction of the algorithms is related to the efficient updating of the unobserved log-intensities in both models and of the dispersion parameter in the negative binomial case, the posteriors of which are of unknown form. For the log-intensities, we adopt the Fast Universal Self-Tuned Sampler (FUSS) of Martino et al, (2015), which is an efficient Metropolis-Hastings type method. For the dispersion parameter of the negative binomial $INSI$ model, we use the Metropolis-adjusted Langevin algorithm (MALA), introduced in Roberts and Tweedie (1996) and further studied in Roberts and Rosenthal (1998). Model selection is based on the computation of the marginal likelihood with the cross entropy method (Chan and Eisenstat, 2015). Forecast evaluation is based on the calculation of point and density forecasts as well as on optimal prediction pooling (Geweke and Amisano, 2011a).

We carry out a simulation study in order to evaluate the performance of our Bayesian methodology. To empirically illustrate its usefulness, we implement it to health data as well as to three high-frequency financial (transaction) data on the number of stock trades, which is characterized by pronounced correlation structure and overdispersion (Rydberg and Shephard 2000; Liesenfeld et al., 2006).

The paper is structured as follows. In section 2 we set up the proposed mixed Poisson $INSI$ model, examine its probabilistic properties and show how the modelling approach taken here can be generalized to account for other $INSI$-type models. In section 3 we describe the prior-posterior analysis for the two cases of the proposed specification (P-$INSI$ and NB-$INSI$), while in section 4 we conduct a simulation study. In section 5 we carry out our empirical analysis. Section 5 concludes. An Online Appendix accompanies this paper.

2 The mixed Poisson integer-valued stochastic intensity model

2.1 The set up

Consider the unknown real parameters ϕ_0 and ϕ_1 and an independently and identically distributed ($i.i.d$) latent sequence $\{e_t, t \in \mathbb{Z}\}$ with mean zero and unit variance. Let also $\{Z_t, t \in \mathbb{Z}\}$ be an $i.i.d$ sequence of positive random variables with unit mean and variance $\rho^2 \geq 0$ and $\{N_t(.), t \in \mathbb{Z}\}$ be an $i.i.d$ sequence of homogeneous Poisson processes with unit intensity. The sequences $\{e_t, t \in \mathbb{Z}\}$, $\{Z_t, t \in \mathbb{Z}\}$ and $\{N_t(.), t \in \mathbb{Z}\}$ are assumed to be independent.

A mixed Poisson integer-valued stochastic intensity ($INSI$) model is an observable integer-valued stochastic process $\{Y_t, t \in \mathbb{Z}\}$ given by the following equation

$$Y_t = N_t(Z_t \lambda_t), \tag{1}$$

where the logarithm of the intensity $\lambda_t > 0$ (latent mean process) follows a first-order autoregression[2]

[2] To simplify the analysis, we specify a latent AR(1) process, although generalizations to AR(p) processes are straightforward.

driven by ϕ_0, ϕ_1 and $\{e_t, t \in \mathbb{Z}\}$, that is,

$$\log(\lambda_t) = \phi_0 + \phi_1 \log(\lambda_{t-1}) + \sigma e_t, \quad t \in \mathbb{Z}, \tag{2}$$

with $\sigma > 0$. The family of processes represented by (1) is known as mixed Poisson process with mixing variable Z_t (Mikosch, 2009). It is important to mention that our proposed specification is different from that of Christou and Fokianos (2015) in two aspects. First, we introduce the latent process e_t and second the log-intensity here does not depend on previous counts. Model (3) should also not to be confused with similar specifications proposed by Zeger,1988; Chan and Ledolter, 1995; Davis et al., 2000. Depending on the law of Z_t, this class of models offers a wide range of conditional distributions for Y_t given λ_t. In the development of the proposed estimation methodology, two special distributions are considered.

First, when Z_t is degenerate at 1 (i.e., $\rho^2 = 0$), the conditional distribution of $Y_t|\lambda_t$ is the Poisson distribution with intensity λ_t, namely,

$$Y_t|\lambda_t \sim \mathcal{P}(\lambda_t), \tag{3}$$

where $\mathcal{P}(\lambda)$ denotes the Poisson distribution with parameter λ. The model, given by (2) and (3), along with the normal distributional assumption that $e_t \overset{i.i.d}{\sim} N(0,1)$, is named the Poisson $INSI$ model (P-$INSI$). This model is characterized by conditional equidispersion, i.e., $E(Y_t|\lambda_t) = var(Y_t|\lambda_t) = \lambda_t$.

Second, when $Z_t \sim \mathcal{G}(\rho^{-2}, \rho^{-2})$ with $\rho^2 > 0$, the conditional distribution of model (1)-(2) reduces to the negative binomial distribution

$$Y_t|\lambda_t \sim \mathcal{NB}\left(\rho^{-2}, \tfrac{\rho^{-2}}{\rho^{-2}+\lambda_t}\right), \tag{4}$$

where $\mathcal{NB}(r,p)$ and $\mathcal{G}(a,b)$ denote the negative binomial distribution with parameters $r > 0$ and $p \in (0,1)$, and the gamma distribution with shape $a > 0$ and rate $b > 0$, respectively. The variance of the mixing sequence ρ^2 is called the dispersion parameter. We refer to the model, given by (2) and (4), along with the normal distributional assumption that $e_t \overset{i.i.d}{\sim} N(0,1)$, as the negative binomial $INSI$ model (NB-$INSI$). This model is characterized by conditional overdispersion, i.e., $var(Y_t|\lambda_t) = \lambda_t + p^2\lambda_t^2 > E(Y_t|\lambda_t) = \lambda_t$.

Other well-known conditional distributions of Y_t can be obtained, depending on the distribution of the mixing variable Z_t. For instance, if Z_t is distributed as an inverse-Gaussian, then $Y_t|\lambda_t$ follows the Poisson-inverse Gaussian model (Dean et al., 1989). Moreover, if the distribution of Z_t is log-normal, then the conditional distribution of Y_t is a Poisson-log-normal mixture (Hinde, 1982). The mixed Poisson $INSI$ model also includes the double Poisson distribution (Efron, 1986) that handles both underdispersion and overdispersion, the Poisson stopped-sum distribution (Feller, 1943) and the Tweedie-Poisson model (Jørgensen, 1997; Kokonendji et al., 2004).

The mixed Poisson $INSI$ model forms a particular class of unobserved conditional intensity models that are based on $\{N_t(.), t \in \mathbb{Z}\}$. Assuming stochastic processes other than $\{N_t(.), t \in \mathbb{Z}\}$ gives rise to different $INSI$-type models; see a remark of section 2.2. What is more, this paper deals with an alternative to the INGARCH processes, for which the intensity parameters depend only on the past process.

The $INGARCH$ model can not be written as a Multiplicative Error Model (MEM, Engle, 2002), but in spite of that it has an ARMA representation. The mixed Poisson $INSI$ model does not have a MEM structure. Furthermore, it is the conjugation between the non-MEM form and the log-intensity equation (in the $INGARCH$ there is no such equation) that makes the mixed Poisson $INSI$ model to not admit a weak ARMA representation. This means that studying the probabilistic structure (such as ergodicity, geometric ergodicity, etc.,) of these models could be tedious. It is much easier, though, to do that for the mixed Poisson $INSI$ model[3].

2.2 The probabilistic structure of the mixed Poisson $INSI$ model

The conditional mean and conditional variance of the mixed Poisson $INSI$ model are given, respectively, by (see, for example, Christou and Fokianos, 2015 ; Fokianos, 2016)

$$E\left(Y_t|\lambda_t\right) = \lambda_t \tag{5}$$

$$Var\left(Y_t|\lambda_t\right) = \lambda_t + \lambda_t^2\rho^2, \tag{6}$$

Under the following condition

$$|\phi_1| < 1, \tag{7}$$

expression (2) admits a unique strictly stationary and ergodic solution given by

$$\lambda_t = \exp\left(\frac{\phi_0}{1-\phi_1} + \sum_{j=0}^{\infty}\phi_1^j\sigma e_{t-j}\right),\ t \in \mathbb{Z}, \tag{8}$$

where the series in (8) converges almost surely and in mean square. The following result shows that (7) is a necessary and sufficient condition for strict stationarity and ergodicity of $\{Y_t, t \in \mathbb{Z}\}$.

Theorem 1. *The process $\{Y_t, t \in \mathbb{Z}\}$, defined by (1)- (2), is strictly stationary and ergodic if and only if (7) holds. Moreover, for all $t \in \mathbb{Z}$,*

$$Y_t = N_t\left(Z_t \exp\left\{\frac{\phi_0}{1-\phi_1} + \sum_{j=0}^{\infty}\phi_1^j\sigma e_{t-j}\right\}\right). \tag{9}$$

Proof. *Appendix.* □

Other properties, such as strong mixing are obvious.

Theorem 2. *Assume that e_t has an a.s. positive density on $\mathbb{R}$. Under the condition $|\phi_1| < 1$, the process $\{Y_t, t \in \mathbb{Z}\}$, defined by (1)- (2), is β-mixing.*

Proof. *Appendix.* □

Given the form of the stationary solution in (9), we can derive its moment properties. Let $\Delta_{tj} = \exp\left(\phi_1^j\sigma e_{t-j}\right)$, $j \in \mathbb{N}, t \in \mathbb{Z}$, and assume that the following conditions hold

$$E\left(\prod_{j=0}^{\infty}\Delta_{tj}\right) = \prod_{j=0}^{\infty}E\left(\Delta_{tj}\right), \tag{10}$$

[3] The same strategy is used by the literature on $INGARCH$ models (Ferland et al., 2006; Fokianos et al., 2009; Weiß, 2009 Zhu, 2010; Christou and Fokianos, 2014; Davis and Liu, 2016; Aknouche and Francq, 2021; Aknouche and Demmouche, 2019; Silva and Barreto-Souza, 2019).

$$\prod_{j=0}^{\infty} E\left(\Delta_{tj}\right) < \infty. \tag{11}$$

The equality in (10) is not always satisfied for any independent sequence $\{\Delta_{tj}, j \in \mathbb{N}, t \in \mathbb{Z}\}$ and one can exhibit examples of independent sequences for which (10) is not fulfilled; see Aknouche (2017). Nevertheless, by the *dominated convergence theorem*, a sufficient condition for (10) to hold is that

$$\prod_{j=0}^{n} \Delta_{tj} \leq W_t, \ a.s. \text{ for all } n \in \mathbb{N}, \tag{12}$$

for some integrable random variable W_t.

The mean, the variance and the autocovariances of the mixed Poisson integer-valued stochastic intensity model are given as follows.

Proposition 1. *Under (7) and (10)-(12), the mean of the process $\{Y_t, t \in \mathbb{Z}\}$, defined by (1)-(2), is given by*

$$E\left(Y_t\right) = \exp\left(\frac{\phi_0}{1-\phi_1}\right) \prod_{j=0}^{\infty} E\left(\Delta_{tj}\right). \tag{13}$$

If, in addition, $e_t \sim N(0,1)$, then (11) reduces to (7) and $E(Y_t)$ is explicitly given by

$$E\left(Y_t\right) = \exp\left(\frac{\phi_0}{1-\phi_1} + \frac{\sigma^2}{2\left(1-\phi_1^2\right)}\right). \tag{14}$$

Proof. *Appendix.* □

To calculate the variance of the mixed Poisson INSI consider the following modifications of expressions (10) and (11)

$$E\left(\prod_{j=0}^{\infty} \Delta_{tj}^2\right) = \prod_{j=0}^{\infty} E\left(\Delta_{tj}^2\right), \tag{15}$$

$$\prod_{j=0}^{\infty} E\left(\Delta_{tj}^2\right) < \infty. \tag{16}$$

As for expression (10), a sufficient condition for (15) to hold is that

$$\prod_{j=0}^{n} \Delta_{tj}^2 \leq V_t, \ a.s. \text{ for all } n \in \mathbb{N}, \tag{17}$$

for some integrable random variable V_t.

Proposition 2. *Under (7) and (15) -(17), the variance of the process $\{Y_t, t \in \mathbb{Z}\}$, defined by (1) -(2), is given by*

$$var\left(Y_t\right) = \exp\left(\frac{\phi_0}{1-\phi_1}\right) \prod_{j=0}^{\infty} E\left(\Delta_{tj}\right) + \exp\left(\frac{2\phi_0}{1-\phi_1}\right)\left[\left(\rho^2+1\right) \prod_{j=0}^{\infty} E\left(\Delta_{tj}^2\right) - \prod_{j=0}^{\infty}\left[E\left(\Delta_{tj}\right)\right]^2\right]. \tag{18}$$

If, in addition, $e_t \sim N(0,1)$*, then (16) reduces to (7) and* $var(Y_t)$ *is explicitly given by*

$$var(Y_t) = \exp\left(\frac{\phi_0}{1-\phi_1} + \frac{\sigma^2}{2\left(1-\phi_1^2\right)}\right) + \exp\left(\frac{2\phi_0}{1-\phi_1}\right)\left[\left(\rho^2+1\right)\exp\left(\frac{2\sigma^2}{1-\phi_1^2}\right) - \exp\left(\frac{\sigma^2}{1-\phi_1^2}\right)\right]. \tag{19}$$

Proof. *Appendix.* □

The Poisson $INSI$ model is conditionally equidispersed but unconditionally overdispersed as

$$\begin{aligned} var(Y_t) &= \exp\left(\tfrac{\phi_0}{1-\phi_1} + \tfrac{\sigma^2}{2\left(1-\phi_1^2\right)}\right) + \exp\left(\tfrac{2\sigma^2}{1-\phi_1^2} + \tfrac{2\phi_0}{1-\phi_1}\right) - \exp\left(\tfrac{\sigma^2}{1-\phi_1^2} + \tfrac{2\phi_0}{1-\phi_1}\right) \\ &= E(Y_t) + \exp\left(\tfrac{2\sigma^2}{1-\phi_1^2} + \tfrac{2\phi_0}{1-\phi_1}\right) - \exp\left(\tfrac{\sigma^2}{1-\phi_1^2} + \tfrac{2\phi_0}{1-\phi_1}\right) \\ &> E(Y_t) \end{aligned}$$

The negative binomial $INSI$ model is conditionally overdispersed, so it is clear that it is also unconditionally overdispersed. However, it is important to note that overdispersion implied by the negative binomial case is more pronounced than the one implied by the Poisson case, and this is what we have emphasized on.

Let $\gamma_h = E(Y_tY_{t-h}) - E(Y_t)E(Y_{t-h})$ be the autocovariance function of the process $\{Y_t, t \in \mathbb{Z}\}$. The expression of γ_h is quite complicated for the negative binomial $INSI$ model and we restrict our attention to the Poisson $INSI$ model. Assume that

$$E\left[\prod_{j=0}^{\infty} \exp\left\{\left(\phi_1^h+1\right)\phi_1^j\sigma e_{t-h-j}\right\}\right] = \prod_{j=0}^{\infty} E\left[\exp\left\{\left(\phi_1^h+1\right)\phi_1^j\sigma e_{t-h-j}\right\}\right], \tag{20}$$

$$\prod_{j=0}^{\infty} E\left[\exp\left\{\left(\phi_1^h+1\right)\phi_1^j\sigma e_{t-h-j}\right\}\right] < \infty. \tag{21}$$

Proposition 3. *Under (3) and (20)-(21), the autocovariance of the process* $\{Y_t, t \in \mathbb{Z}\}$ *is given, for* $h > 0$ *, by*

$$\gamma_h = \exp\left(\frac{2\phi_0}{1-\phi_1}\right)\left(\prod_{j=0}^{h-1} E(\Delta_{tj}) \prod_{j=0}^{\infty} E\left[\exp\left\{\left(\phi_1^h+1\right)\phi_1^j\sigma e_{t-h-j}\right\}\right] - \prod_{j=0}^{\infty}\left[E(\Delta_{tj})\right]^2\right). \tag{22}$$

If, in addition, $e_t \sim N(0,1)$*, then*

$$\gamma_h = \exp\left(\tfrac{2\phi_0}{1-\phi_1}\right)\left(\exp\left(\tfrac{\sigma^2}{2}\tfrac{1-\phi_1^{2h}}{1-\phi_1^2} + \tfrac{\left(\phi_1^h+1\right)^2}{2}\tfrac{\sigma^2}{1-\phi_1^2}\right) - \exp\left(\tfrac{\sigma^2}{1-\phi_1^2}\right)\right). \tag{23}$$

Proof. *Appendix.* □

We next obtain the sth moment $E(Y_t^s)$, $(s \geq 1)$ for the Poisson case corresponding to $\rho^2 = 0$ and the first four moments for the negative binomial case. Assume that

$$E\left(\prod_{j=0}^{\infty} \Delta_{tj}^s\right) = \prod_{j=0}^{\infty} E\left(\Delta_{tj}^s\right), \tag{24}$$

$$\prod_{j=0}^{\infty} E\left(\Delta_{tj}^{s}\right) < \infty, \tag{25}$$

and let $\left\{ {s \atop i} \right\}$ denote the Stirling number of the second kind (see, for example, Ferland et al., 2006; Graham et al., 1989).

Proposition 4. *Assume that (7) and (24) -(25) hold.*

A) Poisson case: The sth moment of the Poisson $INSI$ process (1) -(2), corresponding to $\rho^2 = 0$, is given by

$$E\left(Y_t^s\right) = \sum_{i=0}^{s} \left\{ {s \atop i} \right\} \exp\left(\frac{i\phi_0}{1-\phi_1}\right) \prod_{j=0}^{\infty} E\left(\Delta_{tj}^{i}\right). \tag{26}$$

If, in addition, $e_t \sim N(0,1)$, then (25) reduces to (7) and $E\left(Y_t^s\right)$ is explicitly given by

$$E\left(Y_t^s\right) = \sum_{i=0}^{s} \left\{ {s \atop i} \right\} \exp\left(\frac{i\phi_0}{1-\phi_1} + \frac{i^2\sigma^2}{2\left(1-\phi_1^2\right)}\right), s \geq 1. \tag{27}$$

B) Negative binomial case: The first four moments of the negative binomial $INSI$ process (1) -(2), corresponding to $Z_t \sim \mathcal{G}(\rho^{-2},\, \rho^{-2})$ $(\rho^2 > 0)$, are given by

$$E\left(Y_t^s\right) = \sum_{i=1}^{s} A_s^i \exp\left(\tfrac{i\phi_0}{1-\phi_1}\right) \prod_{j=0}^{\infty} E\left(\Delta_{tj}^{i}\right),\ 1 \leq s \leq 4, \tag{28}$$

where $A_s^1 = 1$ $(1 \leq s \leq 4)$, $A_2^2 = 1 + \rho^2$, $A_3^2 = 3\left(1+\rho^2\right)$, $A_3^3 = 1 + 3\rho^2 + 2\rho^4$, $A_4^2 = 7\left(1+\rho^2\right)$, $A_4^3 = 6 + 18\rho^2 + 12\rho^4$, *and* $A_4^4 = 1 + 6\rho^2 + 11\rho^4 + 6\rho^6$.

If, in addition, $e_t \sim N(0,1)$, then (25) reduces to (7) and $E\left(Y_t^s\right)$ is given by

$$E\left(Y_t^s\right) = \sum_{i=1}^{s} A_s^i \exp\left(\tfrac{i\phi_0}{1-\phi_1} + \tfrac{i^2\sigma^2}{2\left(1-\phi_1^2\right)}\right),\ 1 \leq s \leq 4. \tag{29}$$

Proof. *Appendix.* □

Before we turn our attention to the posterior analysis of the P-$INSI$ and NB-$INSI$ models, we show that the mixed Poisson $INSI$ model follows from a general $INSI$ model.

Remark. *The mixed Poisson $INSI$ model may be enlarged to a more general class of conditional distributions.*

In section 2.1 we defined the mixed Poisson $INSI$ model as a process corresponding to the class of mixed Poisson conditional distributions. This model choice is motivated by the fact that with such a class of distributions, one can use the device of the mixed Poisson process to build a stochastic equation driven by $i.i.d$ innovations, so that path properties can be easily revealed. Moreover, the class of mixed Poisson distributions is quite large and contains many well known count distributions, which are useful and widely used in practice, such as the Poisson and the negative binomial.

However, we can still define the $INSI$ model for a larger class of distributions for which a corresponding stochastic equation with $i.i.d$ innovations also exists. Let F_λ be a discrete cumulative distribution function (cdf) indexed by its mean $\lambda = \int_0^{+\infty} x dF_\lambda(x) > 0$ and with support $[0,\infty)$ (i.e. $F_\lambda(x) = 0$ for all $x < 0$). A priori, no restriction on F_λ is required, so F_λ can belong, for instance, to the exponential family, to the class of mixed Poisson distributions or to any larger class (see, for example, the class of equal stochastic and mean orders proposed by Aknouche and Francq (2021)).

Let us consider the general $INSI$ process (X_t), which is defined to have F_λ as conditional distribution

$$X_t|\lambda_t \sim F_{\lambda_t}(.)\,, \tag{30}$$

where the latent intensity process (λ_t) satisfies the log-autoregression in (2).

Whatever the distribution F_{λ_t} of $X_t|\lambda_t$, model (30) can be written as a stochastic equation with $i.i.d$ inputs, as in Neumann (2011), Davis and Liu (2016) and Aknouche and Francq (2021). In particular, let F_λ^- be the quantile function associated with F_λ. It is well known that $F_\lambda^-(U)$ has the cdf F_λ, when U is uniformly distributed on $[0,1]$. Assume that (U_t) is a sequence of $i.i.d$ $U_{[0,1]}$. Then, the general $INSI$ model (30) can be expressed as the following stochastic equation

$$X_t \;=\; F_{\lambda_t}^-(U_t)\,, \tag{31}$$

where λ_t is given by (2), with $i.i.d$ inputs $\{(U_t,e_t)\}$, where $\{U_t\}$ and $\{e_t\}$ are assumed to be independent. When F_{λ_t} is the cdf of the Poisson distribution, then we obtain the Poisson $INSI$ model. When F_{λ_t} is the cdf of the negative binomial distribution, then we obtain the negative binomial $INSI$ model. We can easily study the probabilistic properties of model (31) in a way similar to that for the mixed-Poisson case. For example, the conditional mean of (31)-which is the analogue of (5)- is $E(X_t|\lambda_t) = E\left[E\left(F_{\lambda_t}^-(U_t)\,|\lambda_t\right)\right] = E(\lambda_t)\,.$

Hence, the Bayesian estimation methodology of this paper can by no means restricted to the mixed Poisson $INSI$ case. Instead, it can be easily modified to accommodate any $INSI$-type model, represented by (31).

3 MCMC inference

In this section we propose algorithmic schemes for two cases of the mixed Poisson $INSI$ model, assuming that the distribution of the innovation in the log- intensity equation is Gaussian. The first case refers to the Poisson $INSI$ model for which $\rho^2 = 0$, so the parameter vector to be estimated is $\theta = \left(\phi_0,\phi_1,\sigma^2\right)'$.

The second case refers to the negative binomial $INSI$ model, corresponding to $Z_t \sim \mathcal{G}(\rho^{-2},\,\rho^{-2})$, with $\rho^2 > 0$. The vector of parameters to be estimated is now $\theta = \left(\phi_0,\phi_1,\tau,\sigma^2\right)'$, where $\tau = \rho^{-2}$ (the dispersion parameter).

3.1 Estimating the Poisson $INSI$ model

Following the Bayesian paradigm, the parameter vector θ and the unobserved intensities $\log(\lambda) = (\log(\lambda_1),...,\log(\lambda_n))'$ are viewed as random with a prior distribution $f(\theta,\log(\lambda))$. Given a series $Y = (Y_1,\ldots,Y_n)'$ generated from (1)-(2) with Gaussian innovation $\{e_t, t\in\mathbb{Z}\}$ and $\rho^2 = 0$, our goal is to sample from the joint posterior distribution $f(\theta,\log(\lambda)|Y)$.

Under the normality of the innovation of the log-intensity equation, the conditional posteriors $f\left(\phi_0|Y,\sigma^2,\lambda,\phi_1\right)$ and $f\left(\sigma^2|Y,\phi_0,\phi_1,\lambda\right)$ are well defined. For $f\left(\phi_1|Y,\sigma^2,\lambda,\phi_0\right)$ we implement an independence Metropolis-Hastings step. The vector $\log(\lambda)$ in $f\left(\log(\lambda)|Y,\phi_0,\phi_1,\sigma^2\right)$ is updated element-by-element, using the Fast Universal Self-Tuned Sampler (FUSS) of Martino et al, (2015)[4].

[4]The FUSS algorithm is faster and has better mixing properties than alternative MCMC methods such as the MALA, slice sampling and Hamiltonian Monte Carlo sampling. The FUSS matlab function is available from Martino's webpage.

Hence, in this single-move framework, we update $f\left(\log(\lambda_t)|Y, \phi_0, \phi_1, \sigma^2, \log(\lambda_{-\{t\}})\right)$ $(1 \leq t \leq n)$, where $\log(\lambda_{-\{t\}})$ denotes the $\log(\lambda)$ vector after removing its t-th component $\log(\lambda_t)$.

Sampling ϕ_0

Assuming that the process in (2) is initialized with $\log(\lambda_1) \sim N(\phi_0|(1-\phi_1), \sigma^2|(1-\phi_1^2))$ and that the prior for ϕ_0 is the normal $N(\mu_{\phi_0}, \sigma^2_{\phi_0})$, we update ϕ_0 by sampling from

$$\phi_0|\mu_{\phi_0}, \sigma^2_{\phi_0}, Y, \phi_1, \lambda \sim N\left(D_1 d_1, D_1\right), \tag{32}$$

where

$$D_1 = \left(\frac{1}{\sigma^2_{\phi_0}} + \frac{1+\phi_1}{\sigma^2(1-\phi_1)} + \frac{n-1}{\sigma^2}\right)^{-1},\ d_1 = \frac{\mu_{\phi_0}}{\sigma^2_{\phi_0}} + \frac{\log(\lambda_1)(1+\phi_1)}{\sigma^2} + \frac{\sum_{t=1}^{n-1}[\log(\lambda_{t+1}) - \phi_1\log(\lambda_t)]}{\sigma^2}.$$

Sampling ϕ_1

Assuming a normal prior $N(\mu_{\phi_1}, \sigma^2_{\phi_1})$ for ϕ_1, its conditional posterior is given by

$$f(\phi_1|\mu_{\phi_1}, \sigma^2_{\phi_1}, Y, \phi_0, \lambda) \propto \sqrt{1-\phi^2}\times$$

$$\exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{n-1}(\log(\lambda_{t+1}) - \phi_0 - \phi_1\log(\lambda_t))^2 - (\log(\lambda_1) - \frac{\phi_0}{1-\phi_1})^2(1-\phi_1^2) - \frac{1}{2\sigma^2_{\phi_1}}(\phi_1 - \mu_{\phi_1})^2\right\}1_{(-1<\phi_1<1)}. \tag{33}$$

The above posterior is of unknown form, so we use an independence Metropolis-Hastings step with proposal density $N(B_{\phi_1}m_{\phi_1}, B_{\phi_1})1_{(-1<\phi_1<1)}$, where

$$B_{\phi_1} = \left(\frac{1}{\sigma^2_{\phi_1}} + \frac{\sum_{t=1}^{n-1}\log(\lambda_t)}{\sigma^2}\right)^{-1},\ m_{\phi_1} = \left(\frac{\mu_{\phi_1}}{\sigma^2_{\phi_1}} + \frac{\sum_{t=1}^{n-1}[\log(\lambda_t)(\log(\lambda_{t+1}) - \phi_0)]}{\sigma^2}\right).$$

Given the current value $\phi_1^{(c)}$, we move to the proposed point $\phi_1^{(p)}$ with probability

$$a_p(\phi_1^{(c)}, \phi_1^{(p)}) = min\left(\frac{\pi(\phi_1^{(p)})\sqrt{1-\phi_1^{(p)}}}{\pi(\phi_1^{(c)}\sqrt{1-\phi_1^{(c)}}}, 1\right),$$

where $\pi(\phi_1)$ is the prior of ϕ_1.

Sampling σ^2

Assuming an inverse gamma prior $\mathcal{IG}(v_a/2, v_\beta/2)$ for σ^2, we update this parameter by sampling from

$$\sigma^2|v_a, v_\beta, Y, \phi_0, \phi_1, \lambda \sim$$

$$\mathcal{IG}\left(\frac{v_a+n}{2}, \frac{v_\beta + (\log(\lambda_1) - \frac{\phi_0}{1-\phi_1})^2(1-\phi_1^2) + \sum_{t=1}^{n-1}(\log(\lambda_{t+1}) - \phi_0 - \phi_1\log(\lambda_t))^2}{2}\right). \tag{34}$$

Sampling the intensity parameters $\log(\lambda) = (\log(\lambda_1), ..., \log(\lambda_n))'$

It remains to sample from the conditional posterior distribution $f\left(\log(\lambda_t)|Y, \theta, \log(\lambda_{-\{t\}})\right)$, $t = 1, 2, ..., n$. Because of the Markovian structure of the intensity process $\{\lambda_t, t \in \mathbb{Z}\}$ and the conditional independence of Y_t and λ_{t-h} $(h \neq 0)$ given λ_t, it follows that for any $1 < t < n$

$$f\left(\log(\lambda_t)|Y, \theta, \log(\lambda_{-\{t\}})\right) \propto f\left(\log(\lambda_t)|\log(\lambda_{t-1}), \theta\right) f\left(\log(\lambda_{t+1})|\log(\lambda_t), \theta\right) f\left(Y_t|\theta, \log(\lambda_t)\right). \tag{35}$$

Using the fact that $Y_t|\theta, \lambda_t \equiv Y_t|\lambda_t \sim \mathcal{P}(\lambda_t)$, and $\log(\lambda_t) \,|\, \log(\lambda_{t-1}), \theta \sim N\left(\phi_0 + \phi_1 \log(\lambda_{t-1}), \sigma^2\right)$, the logarithmic version of (35) becomes

$$\log f\left(\log(\lambda_t)|Y, \theta, \log(\lambda_{-\{t\}})\right) \propto -\exp(\log(\lambda_t)) + Y_t \log(\lambda_t) - \tfrac{1}{2\Omega}\left(\log(\lambda_t) - \mu_t\right)^2, \quad 1 < t < n, \tag{36}$$

where

$$\mu_t = \tfrac{\log(\lambda_t)(\phi_0 - \phi_1\phi_0) + \phi_1 \log(\lambda_t)(\log(\lambda_{t-1}) + \log(\lambda_{t+1}))}{1+\phi_1^2} \tag{37}$$

$$\Omega = \tfrac{\sigma^2}{1+\phi_1^2}. \tag{38}$$

To sample from this expression, we use the FUSS sampler (Martino et al, 2015). The FUSS sampler is an MCMC method which can be used to sample efficiently from a univariate distribution. The sampler consists of four steps. In the first step, we choose an initial set of support points of the target distribution. In the second step, unnecessary support points are removed according to some pre-defined pruning criterion (optimal minimax pruning strategy). In the third step, the construction of the independent proposal density, tailored to the shape of the target takes place, with some appropriate pre-defined mechanism (interpolation). In the last step, a Metropolis–Hastings (MH) method is used.

3.2 Estimating the negative binomial $INSI$ model

For the mixed Poisson $INSI$ model (1)-(2) with $\rho^2 > 0$ and $Z_t \sim \mathcal{G}(\rho^{-2}, \rho^{-2})$, leading to $Y_t|\lambda_t, \theta \sim \mathcal{NB}\left(\tau, \frac{\tau}{\tau+\lambda_t}\right)$, we have to estimate $\theta = \left(\phi_0, \phi_1, \sigma^2, \tau\right)'$. We use again the Gibbs sampler, where the conditional posteriors for ϕ_0, ϕ_1 and σ^2 are sampled as in the Poisson case. It remains to show how to sample from $f\left(\log(\lambda)|Y, \theta\right)$ and $f\left(\tau|Y, \phi_0, \phi_1, \sigma^2, \lambda\right)$.

Sampling the augmented intensity parameters $\log(\lambda) = (\log(\lambda_1), ..., \log(\lambda_n))'$

We first derive the kernel of $f\left(\lambda_t|Y, \theta, \lambda_{-\{t\}}\right)$ for the case of the negative binomial model. It is still given by (35), where now $\theta = \left(\phi_0, \phi_1, \sigma^2, \tau\right)'$. Using the fact that $Y_t|\theta, \lambda_t \equiv Y_t|\tau, \lambda_t \sim \mathcal{NB}\left(\tau, \frac{\tau}{\lambda_t+\tau}\right)$ and $\log(\lambda_t) \,|\, \log(\lambda_{t-1}), \theta \sim N\left(\phi_0 + \phi_1 \log(\lambda_{t-1}), \sigma^2\right)$, formula (35) becomes

$$\log f\left(\log(\lambda_t)|Y, \theta, \log(\lambda_{-\{t\}})\right) \propto \tfrac{\Gamma(Y_t+\tau)}{\Gamma(\tau)}\left(\tfrac{\tau}{\tau+\lambda_t}\right)^{\tau}\left(\tfrac{\lambda_t}{\tau+\lambda_t}\right)^{Y_t}\exp\left(-\tfrac{1}{2\Omega}\left(\log(\lambda_t) - \mu_t\right)^2\right), \tag{39}$$

where μ_t and Ω are given by (37) and (38), respectively. Then, we can use the FUSS sampler (Martino et al, 2015) to draw efficiently from expression (39), as in the Poisson case.

Sampling the dispersion parameter τ

If $f(\tau)$ denotes the prior distribution of τ, then the posterior distribution $f\left(\tau|Y, \phi_0, \phi_1, \sigma^2, \lambda\right)$ is given by

$$f\left(\tau|Y, \phi_0, \phi_1, \sigma^2, \lambda\right) \propto f(\tau)\, f(Y|\theta, \lambda), \tag{40}$$

where $f(Y|\theta, \lambda)$ is the likelihood function

$$f(Y|\theta, \lambda) = \prod_{t=1}^{n} \tfrac{\Gamma(Y_t+\tau)}{\Gamma(\tau)}\left(\tfrac{\tau}{\tau+\lambda_t}\right)^{\tau}\left(\tfrac{\lambda_t}{\tau+\lambda_t}\right)^{Y_t}. \tag{41}$$

Since it is difficult to find a conjugate prior for τ, we exploit, as is usually the case, the gamma prior. In particular, we assume that $\tau > 0$ follows the gamma distribution with hyperparameters $a > 0$ and

$b > 0$, i.e.,

$$f(\tau) = \frac{b^a \tau^{a-1}}{\Gamma(a)} e^{-b\tau}.$$

Therefore, (40) becomes

$$f\left(\tau|Y, \phi_0, \phi_1, \sigma^2, \lambda\right) \propto \tau^{a-1} e^{-b\tau} \prod_{t=1}^{n} \frac{\Gamma(Y_t+\tau)}{\Gamma(\tau)} \left(\frac{\tau}{\tau+\lambda_t}\right)^{\tau} \left(\frac{\lambda_t}{\tau+\lambda_t}\right)^{Y_t}. \tag{42}$$

The posterior $f\left(\tau|Y, \phi_0, \phi_1, \sigma^2, \lambda\right)$ is not amenable to closed-form integration (Bradlow et al., 2002). To sample from this posterior we use the Metropolis adjusted Langevin algorithm (MALA); see Roberts and Tweedie (1996) and Roberts and Rosenthal (1998).

In particular, given the current value $\tau^{(c)}$, we move to the proposed point $\tau^{(p)}$ with probability

$$a_p(\tau^{(c)}, \tau^{(p)}) = min\left(\frac{f(\tau^{(p)}|Y, \phi_0, \phi_1, \sigma^2, \lambda)N(\tau^{(c)}|, \tau^{(p)} + \frac{k^2}{2}\Delta \log f(\tau^{(p)}|Y, \phi_0, \phi_1, \sigma^2, \lambda), k)I_{(0,\infty)}(\tau)}{f(\tau^{(c)}|Y, \phi_0, \phi_1, \sigma^2, \lambda)N(\tau^{(p)}|, \tau^{(c)} + \frac{k^2}{2}\Delta \log f(\tau^{(c)}|Y, \phi_0, \phi_1, \sigma^2, \lambda), k)I_{(0,\infty)}(\tau)}, 1\right),$$

where $k > 0$ is a constant and $\Delta \log f(\tau|Y, \phi_0, \phi_1, \sigma^2, \lambda)$ is the first derivative of the log-posterior of τ, evaluated at the specific value of it. The resulting value of this derivative is obtained from the MATLAB built-in function *fmincon*. Also, $N(\tau|, ;, ;, ;)I_{(0,\infty)}$ is the normal proposal density truncated in the region $(0, \infty)$. The value of k is chosen to achieve acceptance rate of 60%.

3.3 Model comparison

The conditional marginal likelihood (CML) of a model $\mathcal{M}$ with complete-data likelihood $p(Y|\mathcal{M}, \theta, \lambda)$, where θ is the parameter vector with prior $p(\theta|\mathcal{M})$ and λ is the latent variable vector, is defined as

$$p(Y|\lambda, \mathcal{M}) = \int p(Y|\mathcal{M}, \theta, \lambda)p(\theta|\mathcal{M})d\theta. \tag{43}$$

Since (43) does not have closed form, we utilize the Importance Sampling (IS) method of Chan and Eisenstat (2015), to compute it. This method is based on cross-entropy ideas. The importance sampling estimator of expression (43) is given by

$$\widehat{p(Y|\lambda, \mathcal{M})} = \frac{1}{M_1} \sum_{i=1}^{M_1} \frac{p(Y|\mathcal{M}, \theta^{(i)}, \lambda^{(i)})p(\theta^{(i)}|\mathcal{M})}{g(\theta^{(i)}, \lambda^{(i)})}, \tag{44}$$

where $g(\cdot)$ is the importance density. The optimal $g(\cdot)$ is constructed from the posterior draws. In terms of θ, the function g is defined as the product of (independent) distributions for each parameter of θ ; gamma and inverse gammas for the positive ones, truncated normal for those that are defined in $(-1, 1)$ and normal for the ones defined in $\mathbb{R}$.

- *P-INSI* model: For this model, $\theta = \left(\phi_0, \phi_1, \sigma^2\right)'$. Hence,

  $$N(\phi_0; \hat{\phi_0}, S_{\phi_0}) \times N(\phi_1; \hat{\phi_1}, S_{\phi_1})1_{(-1<\phi_1<1)} \times IG(\sigma^2; \hat{\sigma^2}, S_{\sigma^2}),$$

  where in all the above, $\hat{c}$ and S_c denote the posterior mean and variance for parameter c, respectively, obtained from the MCMC product.

- *NB-INSI* model: For this model, $\theta = \left(\phi_0, \phi_1, \tau, \sigma^2\right)'$. For the distribution of τ in g, we take $G(\tau; \hat{\sigma^2_\tau}, S_{\sigma^2_\tau})$, where G denotes the gamma distribution.

Larger values of the CML indicate better model fit.

3.4 Forecast evaluation

A recursive out-of-sample forecasting exercise is conducted for the evaluation of the predictive performance of the proposed models, as is usual practice in the field (Freeland and McCabe 2004; McCabe and Martin, 2005). In this exercise we compute point and density forecasts.

The conditional predictive density of the one-step ahead y_{t+1}, given the data $Y_t = (y_1, ..., y_t)$ is given by

$$p(y_{t+1}|Y_t) = \int p(y_{t+1}|\Omega_t) dp(\Omega_t|Y_t), \tag{45}$$

where $\Omega_t = (\theta, \Lambda_t)$ and $\Lambda_t = (\lambda_1, ..., \lambda_t)$.

The above expression can be approximated by Monte Carlo integration,

$$\widehat{p}(y_{t+1}|Y_t) = \frac{1}{R} \sum_{i=1}^{R} p(y_{t+1}|\Omega_t^{(i)}), \tag{46}$$

where $\Omega_t^{(i)}$ is the posterior draw of Ω_t at iteration $i = 1, ..., R$ (after the burn-in period).

The conditional predictive likelihood of y_{t+1} is the conditional predictive density of y_{t+1} evaluated at the observed y_{t+1}^o, namely, $p(y_{t+1} = y_{t+1}^o|Y_t)$. As a metric for the evaluation of the density forecasts we use the log predictive score (LPS) which is the sum of the log predictive likelihoods

$$LPS = \sum_{t=t_0}^{T-1} \log p(y_{t+1} = y_{t+1}^o|Y_t), \tag{47}$$

where $t = t_0 + 1, ..., T$ is the evaluation period. Higher LPS values indicate better (out-of-sample) forecasting power of the model; see also Geweke and Amisano (2011b).

We also compute the one-step ahead predictive mean $E(y_{t+1}|Y_t)$, which is used as a point forecast for the observation y_{t+1}^o. As in the case of the predictive density of y_{t+1}, we also use predictive simulation for the calculation of the predictive mean. A usual metric to evaluate point forecasts is the root mean squared forecast error (RMSFE) defined as

$$RMSFE = \sqrt{\frac{\sum_{t=t_0}^{T-1} (y_{t+1}^o - E(y_{t+1}|Y_t))^2}{T - t_0}}. \tag{48}$$

Lower values of the RMSFE indicate better point forecasts.

3.5 Optimal pooling

We compare the forecasting performance of the proposed models, by implementing the approach of optimal pooling (Geweke and Amisano, 2011a). According to this approach, none of the competing models is the true data generating process, and as such it considers a linear prediction pool based on the predictive likelihood (log score function) from a set of competing models.

Given a set of models $\{\mathcal{M}_i\}_{i=1}^M$ and a set of predictive densities $\{p(y_t|y_1, ..., y_{t-1}, \mathcal{M}_i)\}_{i=1}^M$, we consider the following form of mixtured predictive densities

$$\sum_{i=1}^{M} w_i p(y_t|y_1,...,y_{t-1},\mathcal{M}_i), \text{where} \sum_{i=1}^{M} w_i = 1, \ \ w_i \geq 0, i = 1,...,M. \tag{49}$$

The optimal weight vector $\boldsymbol{w}^*$ is chosen to maximise the log pooled predictive score function; that is,

$$\underset{w_i, i=1,...,M}{\text{argmax}} \sum_{t=k_1}^{k_2} \log \left(\sum_{i=1}^{M} w_i p(y_t|y_1,...,y_{t-1},\mathcal{M}_i) \right), \tag{50}$$

where the predictive density is evaluated at the realized value y_t.

Conditional on the data up to time $t-1$, i.e., $y_1,...,y_{t-1}$, we get a large number of MCMC draws for the parameters, which are then used to evaluate the predictive likelihood $p(y_t = y_t^o|y_1,...,y_{t-1},\mathcal{M}_i)$. From the entire history of the predictive likelihood values we can then estimate the optimal weights.

4 Simulation study

In this section, we assess the performance of the proposed Bayesian methodology on simulated mixed Poisson $INSI$ series with Gaussian innovations for the log-intensity equation. In particular, we consider two cases of the mixed Poisson $INSI$ model; the P-$INSI$ model and the NB-$INSI$ model.

Throughout our simulations, we generated n=1000 data points from both models, while we run the samplers for 5000 iterations after discarding the initial 5000 cycles (burn-in period). In our empirical study the sample size consists of approximately 3000 observations. The purpose of this simulation study is to show that with much smaller sample size (n=1000), the algorithms still perform satisfactory. In an Online Appendix we repeated the same simulation study with even smaller samples ($n = 300$ and $n = 500$). Again, the simulation results were good.

We use the same priors for the common parameters in both models. In particular, we took

$$\phi_0 \sim N(0,10), \phi_1 \sim N(0,1)1_{(-1<\phi<1)}, \sigma^2 \sim IG(5,0.16).$$

For the dispersion parameter, we used

$$\tau \sim G(5,0.1).$$

To monitor the performance of our sampling algorithms, we estimated the inefficiency factor (IF) that measures how well the chain mixes. The IF is defined as $1 + \sum_{s=1}^{\infty} \varrho_s$, where ϱ_s is the sample autocorrelation at lag s; for further details see Chib (2001). The inefficiency factor quantifies the relative efficiency loss owing to the correlation in the samples drawn. A well designed algorithm will generate low correlations across draws and therefore a low IF.

To monitor any lack of convergence, we also computed the values of the Convergence Diagnostics (CD) statistic of Geweke (1992). The CD statistic compares draws from the early part of the chain to those from the last part of the chain. Lower absolute values of CD statistic indicate a better convergence.

The simulation results (posterior means, standard deviations, IF and CD values) for the P-$INSI$ and the NB-$INSI$ are summarized in Tables 1 and 2, respectively for three different sets of true values of the parameters. In any case, the estimated parameters are close to their true values. Based on the IF and CD values there are no convergence issues, while the mixing of the produced chains is satisfactory.

In an Online Appendix, we display the figures for the posterior paths, the posterior autocorrelation functions, and the posterior histograms for the parameters of this simulation study. The posterior paths were stable and the posterior autocorrelations decayed quickly, indicating that the proposed algorithmic schemes were efficient.

The estimation algorithms were implemented in Matlab on a desktop with Intel Core i5-3470 @ 3.20 GHz 3.20 GHz processes with 8 GB RAM. In terms of computation time, it takes 20.743 seconds to obtain 100 posterior draws from the $P\text{-}INSI$ model and 57.851 seconds from the $NB\text{-}INSI$ model.

5 Empirical applications

5.1 Application I: Financial data

To illustrate the Bayesian methodology of this paper, we use three time series that record the number of trades in five-minute intervals between 9:45 AM and 4:00 PM for three stocks (Glatfelter Company (GLT), Wausau Paper Corporation (WPP), Empire District Electric Company (EDE)). These stocks are traded on the New York Stock Exchange (NYSE) and the time period spans from January 3, 2005 to February 18, 2005. The time series in question, which are plotted in Figure 1, consist of $n = 2925$ observations and have also been used by Jung et al., (2011). As can be seen from Table 3, each time series is strongly overdispersed; the sample variance is greater than the sample mean. This is also verified by the histogram of these series; see Figure 2.

The number of Gibbs iterations was set equal to 5000 with a burn-in period of 5000 updates. Furthermore, the hyperparameters of the prior distributions for the $P\text{-}INSI$ and $NB\text{-}INSI$ models are similar to those used in the simulation study. The significance of the parameters was examined using 95% highest posterior density intervals; see Koop et al. (2007).

5.1.1 Results

The empirical results are presented in Table 4. In particular, it reports the posterior means, standard deviations, inefficiency factors (IF) and CD statistics for both models and for all data sets. According to the CD values of this table, the generated sequences of MCMC draws converge for all parameters of the models. Also, the reported IF values (Table 4) signal a good mixing of the corresponding MCMC chains, produced by the proposed algorithms.

The figures for the posterior paths, the posterior histograms, and the posterior autocorrelation functions, for the parameters of both models and for each of the data set in question are given in an Online Appendix. These figures verify that the posterior paths are stable, while the posterior autocorrelations decay satisfactory, suggesting that the proposed algorithmic schemes are efficient.

All the parameters in Table 4 are statistically significant. We also observe that the estimated persistence parameter ϕ_1 is systematically smaller in the $P\text{-}INSI$ model (around 0.7) than in the $NB\text{-}INSI$ model (around 0.9), whereas the opposite holds for the estimated posterior mean of the variance σ^2 of the log-volatility equation. From the corresponding posterior histograms (see Online Appendix), $\widehat{\phi}_1$ lies in the stability domain, defined by expression (7). The dispersion parameter in the $NB\text{-}INSI$ model is the highest in GLT (0.911) and the smallest in EDE (4.89).

According to the reported values of the log of the conditional marginal likelihood (log-CML), the most preferred model is the $NB\text{-}INSI$. It has the largest log-CML value for each data set. This is also supported by the significance of the τ parameter, which, in any case, is far away from zero (Table 4).

We also considered weighted linear combinations of linear pools (prediction models), which are evaluated using the log predictive scoring rule (Geweke and Amisano, 2011a). In particular, optimal weights can be computed in each time t to form a prediction pool for the one-step-ahead forecasting. Figure 3 shows the evolution of these weights for the last 20 observations of the GLT data set. The largest weight over the entire out-of sample forecasting period was received by the NB-$INSI$ model. The corresponding figures for the EDE and WPP data have been moved to the Online Appendix. In addition, Table 5 displays the weights for the pool optimized over all of the last 20 observations for the three financial data sets. The sum of these numbers for each data set is equal to 1. All the weight is allocated to the NB-$INSI$ model, with the P-$INSI$ being excluded from the optimal prediction pool.

We compared the forecasting performance of the two models, using log predictive scores (LPS) and RMSFEs. For this out-of-sample forecasting exercise, we calculated density and point forecasts for the last 20 data points. Table 4 presents the results. Based on the density forecast results, the NB-$INSI$ model is the best forecasting model. As far as the point forecasts are concerned, the NB-$INSI$ model delivered the lowest RMSFE. In any case, the NB-$INSI$ model dominates the P-$INSI$ model, both in terms of goodness of fit and out-of-sample forecasting ability.

Since NB-$INSI$ outperforms P-$INSI$ we focus on the former. Figure 4 shows the real time path of each time series along with the estimated conditional mean $\widehat{\lambda}_t$ and the estimated conditional variance, given by $\widehat{\lambda}_t\left(1+\frac{1}{\widehat{\tau}}\widehat{\lambda}_t\right)$, obtained from the NB-$INSI$ model. There is an observed consistent evolution of these two quantities with the evolution of the original series, where the (conditional) overdispersion phenomenon is visually highlighted as well.

Given that the proposed specification of this paper attempts to capture the autocorrelation in the data, it would be appropriate to compare the sample autocorrelations from the data and the autocorrelations implied by the NB-$INSI$ model. Hence, we focus on two types of its residuals. First, the Pearson conditional mean residual $(\widehat{\epsilon}_t)_{1\le t\le n}$ (Y-residuals in short) is defined by

$$\widehat{\epsilon}_t = \frac{Y_t-\widehat{\lambda}_t}{\sqrt{\widehat{\lambda}_t\left(1+\frac{1}{\widehat{\tau}}\widehat{\lambda}_t\right)}}, \quad 1\le t\le n,$$

where $\widehat{\lambda}_t$ is the posterior estimate of λ_t, and $\widehat{\lambda}_t\left(1+\frac{1}{\widehat{\tau}}\widehat{\lambda}_t\right)$ is the estimated conditional variance of the model.

Second, we analysed the randomized quantile Y-residuals, defined by Dunn and Smyth (1996) and used e.g. by Benjamin et al., (2003), Zhu (2011) and Aknouche et al., (2018). The randomized quantile Y-residuals are especially useful to achieve continuous residuals, when the series is discrete, as in our case. In the NB-$INSI$ context, they are given by $\widehat{\varepsilon}_t = \Phi^{-1}(p_t)$, where Φ^{-1} is the inverse of the standard normal cumulative distribution and p_t is a random number uniformly chosen in the interval

$$\left[F\left(Y_t-1;\widehat{\theta}\right), F\left(Y_t;\widehat{\theta}\right)\right],$$

$F\left(x;\widehat{\theta}\right)$ being the cumulative function of the negative binomial distribution $\mathcal{NB}\left(\widehat{\tau}, \frac{\widehat{\tau}}{\widehat{\lambda}_t+\widehat{\tau}}\right)$ evaluated at x with parameter $\widehat{\theta}=\left(\widehat{\phi}_0,\widehat{\phi}_1,\widehat{\sigma}^2,\widehat{\tau}\right)'$.

The results from the residual analysis are presented in Figure 5 for the GLT data. For example, in Figure 5 we compare the sample autocorrelations of the observed GLT data (Figure 5a) with those of the residuals (Pearson and randomized). We can see that there is significant autocorrelation in the GLT data series but there is no significant evidence of correlation within the Pearson (Figure 5b) or randomized Y-residuals (Figure 5c). We repeated the same analysis for the rest of the data sets (EDE and WPP) and reported the graphical results in the Online Appendix, as the conclusions remained unchanged.

5.2 Application II: Health data

We also considered, as a second empirical example, the weekly number of a disease caused by Escherichia coli (E.coli) in the state of North Rhine-Westphalia (Germany). The time period of the data set is from January 2001 to May 2013 with $n = 646$. The series has been analyzed by various researchers (Doukhan et al., 2006; Silva and Barreto-Souza, 2019) and a graphical view of it is provided in Figure 6. It is also highly overdispersed, as its sample mean is 20.3344 and its sample variance is 88.7531.

5.2.1 Results

The estimation results are presented in Table 6. No convergence or mixing issues were detected by inspecting the IF and CD values. The trace plots for the parameters of both models along with their posterior histograms and autocorrelations are given in the Online Appendix. The estimated ϕ_1 parameters obtained from the two models are high in magnitude; 0.8075 in the P-$INSI$ model and 0.9012 in the NB-$INSI$ model. Therefore, there is strong persistence in the evolution of the latent log-intensities. We also note that the posterior mean of the dispersion parameter is approximately 32 and statistically significant.

Moreover, the NB-$INSI$ model outperformed the P-$INSI$ model, both in terms of one-step ahead (out-of-sample) point and density forecasts (Table 6) for the last 20 observations of the sample. In Figure 7, the P-$INSI$ model received the least of the weight over the forecasting period and it is also excluded from the optimal pool of models; its optimal weight was zero. The conditional marginal likelihood suggests the NB-$INSI$ model, as having the best in-sample fit (Table 6).

In Figure 8 we plotted the E.coli series along with the conditional means and variances of the log-intensities for the NB-$INSI$ model, which is the dominant one. This figure shows visually that the data are conditionally overdispersed. This model is also able to pick up most of the serial correlation present in the E.coli series. This can be seen from the sample autocorrelations of the Y-residuals and the randomized Y-residuals, given in Figure 9.

6 Conclusions

We proposed an integer-valued stochastic intensity ($INSI$) model that conceptually parallels the stochastic volatility model for real-valued time series. However, like the difference between the $INGARCH$ and the $GARCH$ models, the underlying source of dynamics in our model (time-varying parameter nature of our model) stems from the conditional mean and not the conditional variance, as in the stochastic volatility framework. The proposed specification is a discrete-valued parameter-driven model, depending on a latent time-varying intensity parameter, the logarithm of which follows a drifted first-order autoregression. We focused on a rich class of Poisson mixture distributions that form a particular $INSI$-type model, the mixed Poisson $INSI$ model.

Unlike the standard stochastic volatility model, the mixed Poisson $INSI$ model does not admit a weak ARMA representation nor a multiplicative error representation but, in spite of that, we easily studied its probabilistic properties, such as ergodicity, mixing, covariance structure and existence of higher order moments. The Poisson mixture paradigm has the advantage that its probabilistic structure allows us to write it as a non-linear stochastic difference equation with $i.i.d$ innovations, simplifying the study of its probabilistic properties.

Under the Gaussianity assumption for the innovation of the log-intensity equation, we constructed Markov Chain Monte Carlo algorithms in order to estimate the parameters of the mixed Poisson $INSI$ model for two particular conditional distributions; the Poisson and the negative binomial. The proposed specifications were applied to financial and health data.

It would be interesting to compare the Poisson and negative binomial $INSI$ models with the corresponding $INGARCH$ models, both in terms of model fit and forecasting performance. We leave that for future research.

Appendix: Proofs

Proof of Theorem 1. In view of (9), we know that Y_t is a causal measurable function of the *i.i.d* sequence $\{(N_t, Z_t, e_t), t \in \mathbb{Z}\}$. Hence, $\{Y_t, t \in \mathbb{Z}\}$ is strictly stationary and ergodic. If $|\phi_1| \geq 1$, then clearly there does not exist a nonanticipative strictly stationary solution of (1)-(2), like expression (9). $\square$

Proof of Theorem 2. Under the assumptions of Theorem 2 it is clear that the autoregressive process $\{\log(\lambda_t), t \in \mathbb{Z}\}$ is geometrically ergodic and hence is β-mixing (Meyn and Tweedie, 2009). By the properties of the β-mixing coefficient (e.g. Francq and Zakoian, 2019) it follows that $\{\lambda_t, t \in \mathbb{Z}\}$ is also β-mixing. Now, since Y_t is a measurable function of λ_t, then

$$\beta_Y(h) \leq \beta_\lambda(h),$$

so the result follows, where $\beta_Y(h) = E\left(\sup_{A \in \sigma\{Y_t, t \geq h\}} |P(A/\sigma\{Y_t, t \leq 0\}) - P(A)|\right)$ and so on, with $\sigma\{Y_t, a \leq t \leq b\}$ being the σ-algebra, generated by $\{Y_t, a \leq t \leq b\}$. [5]

$\square$

Proof of Proposition 1. Under (7) and (10)-(12) we have

$$\begin{aligned} E(Y_t) &= E\left(E\left(N_t(Z_t\lambda_t)|Z_t\right)\right) \\ &= E\left[\exp\left(\tfrac{\phi_0}{1-\phi_1} + \sum_{j=0}^{\infty} \phi_1^j \sigma e_{t-j}\right)\right] \\ &= \left[\prod_{j=0}^{\infty} E(\Delta_{tj})\right] \exp\left(\tfrac{\phi_0}{1-\phi_1}\right), \end{aligned}$$

which is (13). If, in addition, $e_t \sim N(0,1)$, then using the fact that if $X \sim N(0,1)$ it holds $E(\exp(\varphi X)) = \exp(\frac{\varphi^2}{2})$ for any nonnull real constant φ, we get

$$\begin{aligned} E(Y_t) &= \exp\left(\tfrac{\phi_0}{1-\phi_1}\right) \prod_{j=0}^{\infty} \exp \tfrac{1}{2}\left(\phi_1^j \sigma\right)^2 \\ &= \exp\left(\tfrac{\phi_0}{1-\phi_1}\right) \exp \tfrac{\sigma^2}{2} \sum_{j=0}^{\infty} \phi_1^{2j} \\ &= \exp\left(\tfrac{\phi_0}{1-\phi_1} + \tfrac{\sigma^2}{2\left(1-\phi_1^2\right)}\right), \end{aligned}$$

which is (14).

$\square$

[5]Note that $\{Y_t, t \in \mathbb{Z}\}$ is not a Markov chain but a Hidden Markov chain in the sense of Leroux (1992). So the result also follows from Proposition 4 of Carrasco and Chen (2002).

Proof of Proposition 2. Under (7) and (15)-(17) and using (5)-(6) we have

$$\begin{aligned}
var\left(Y_t\right) &= E\left(var\left(Y_t|\lambda_t\right)\right)+var\left(E\left(Y_t|\lambda_t\right)\right)\\
&= \left[\prod_{j=0}^{\infty}E\left(\Delta_{tj}\right)\right]\exp\left(\tfrac{\phi_0}{1-\phi_1}\right)+\left(\rho^2+1\right)E\left(\left(\exp\left(\tfrac{\phi_0}{1-\phi_1}+\sum_{j=0}^{\infty}\phi_1^j\sigma e_{t-j}\right)\right)^2\right)\\
&\quad -\left(\prod_{j=0}^{\infty}E\left(\Delta_{tj}\right)\exp\left(\tfrac{\phi_0}{1-\phi_1}\right)\right)^2\\
&= \left(\prod_{j=0}^{\infty}E\left(\Delta_{tj}\right)\right)\exp\left(\tfrac{\phi_0}{1-\phi_1}\right)+\left(\left(\rho^2+1\right)\prod_{j=0}^{\infty}E\left(\Delta_{tj}^2\right)-\prod_{j=0}^{\infty}\left(E\left(\Delta_{tj}\right)\right)^2\right)\exp\left(\tfrac{2\phi_0}{1-\phi_1}\right),
\end{aligned}$$

which is (18). If, in addition, $e_t \sim N(0,1)$, then using the above normality result we get

$$\prod_{j=0}^{\infty}E\left(\Delta_{tj}\right) = \prod_{j=0}^{\infty}E\exp\left(\phi_1^j\sigma e_{t-j}\right)=\exp\left(\tfrac{\sigma^2}{2\left(1-\phi_1^2\right)}\right)$$

and

$$\prod_{j=0}^{\infty}E\left(\Delta_{tj}^2\right) = \prod_{j=0}^{\infty}E\left(\left(\exp\left(\phi_1^j\sigma e_{t-j}\right)\right)^2\right)=\exp\left(\tfrac{2\sigma^2}{1-\phi_1^2}\right).$$

Using these results, we get (19).

□

Proof of Proposition 3. Under (3) and (20)-(21) we have for $h>0$

$$\begin{aligned}
E\left(Y_tY_{t-h}\right) &= E\left[E\left(Y_tY_{t-h}\right)|\lambda_t,\lambda_{t-h}\right]\\
&= E\left[\exp\left(\tfrac{\phi_0}{1-\phi_1}+\sum_{j=0}^{\infty}\phi_1^j\sigma e_{t-j}+\tfrac{\phi_0}{1-\phi_1}+\sum_{j=0}^{\infty}\phi_1^j\sigma e_{t-h-j}\right)\right]\\
&= \exp\left(\tfrac{2\phi_0}{1-\phi_1}\right)\prod_{j=0}^{h-1}E\left(\Delta_{tj}\right)\prod_{j=0}^{\infty}E\left[\exp\left(\left(\phi_1^h+1\right)\phi_1^j\sigma e_{t-h-j}\right)\right].
\end{aligned}$$

Hence, (22) follows by calculating $E\left(Y_t\right)E\left(Y_{t-h}\right)$, using (13).

If $e_t \sim N(0,1)$, then $E\left(\Delta_{tj}\right)=\exp\left(\frac{\sigma^2}{2}\phi_1^{2j}\right)$ and

$$\begin{aligned}
\prod_{j=0}^{h-1}E\left(\Delta_{tj}\right)\prod_{j=0}^{\infty}E\exp\left(\left(\phi_1^h+1\right)\phi_1^j\sigma e_{t-h-j}\right) &= \prod_{j=0}^{h-1}\exp\left(\tfrac{\sigma^2}{2}\phi_1^{2j}\right)\prod_{j=0}^{\infty}\exp\left(\tfrac{\left(\phi_1^h+1\right)^2\sigma^2}{2}\phi_1^{2j}\right)\\
&= \exp\left(\tfrac{\sigma^2}{2}\tfrac{1-\phi_1^{2h}}{1-\phi_1^2}\right)\exp\left(\tfrac{\left(\phi_1^h+1\right)^2\sigma^2}{2}\tfrac{1}{1-\phi_1^2}\right).
\end{aligned}$$

Hence, (23) follows by combining the last two expression. Observe that as $h\to\infty$ (under $|\phi_1|<1$) $\gamma_h \to 0$, which is consistent with asymptotic theory.

□

Proof of Proposition 4. A) Poisson case: When $Y_t|\lambda_t \sim \mathcal{P}\left(\lambda_t\right)$, it is well known that (e.g. Ferland et al., 2006)

$$E\left(Y_t^s|\lambda_t\right) = \sum_{i=0}^{r} \left\{ \begin{matrix} s \\ i \end{matrix} \right\} \lambda_t^i.$$

Hence, under (7) and (24)-(25) we have

$$\begin{aligned} E\left(Y_t^s\right) &= E\left(E\left(Y_t^s|\lambda_t\right)\right) \\ &= \sum_{i=1}^{s} \left\{ \begin{matrix} s \\ i \end{matrix} \right\} E\left(\exp\left(\tfrac{i\phi_0}{1-\phi_1} + i\sum_{j=0}^{\infty} \phi_1^j \sigma e_{t-j}\right)\right) \\ &= \sum_{i=1}^{s} \left\{ \begin{matrix} s \\ i \end{matrix} \right\} \exp\left(\tfrac{i\phi_0}{1-\phi_1}\right) \prod_{j=0}^{\infty} E\left(\Delta_{tj}^i\right), \end{aligned}$$

which is (26).

If, in addition, $e_t \sim N(0,1)$, then using the above normality result we get (27).

B) Negative binomial case: The first four moments of $Y_t|\lambda_t \sim \mathcal{NB}\left(\rho^{-2}, \frac{\rho^{-2}}{\rho^{-2}+\lambda_t}\right)$ are explicitly given by

$$\begin{aligned} E\left(Y_t|\lambda_t\right) &= \lambda_t \\ E\left(Y_t^2|\lambda_t\right) &= \lambda_t + \left(1+\rho^2\right)\lambda_t^2 \\ E\left(Y_t^3|\lambda_t\right) &= \lambda_t + 3\left(1+\rho^2\right)\lambda_t^2 + \left(1+3\rho^2+2\rho^4\right)\lambda_t^3 \\ E\left(Y_t^4|\lambda_t\right) &= \lambda_t + 7\left(1+\rho^2\right)\lambda_t^2 + \left(6+18\rho^2+12\rho^4\right)\lambda_t^3 + \left(1+6\rho^2+11\rho^4+6\rho^6\right)\lambda_t^4, \end{aligned}$$

so (28) follows by combining (8), the formula $E\left(\lambda_t^i\right) = \exp\left(\frac{i\phi_0}{1-\phi_1}\right)\prod_{j=0}^{\infty} E\left(\Delta_{tj}^i\right)$ and the fact that $E\left(Y_t^s\right) = E\left(E\left(Y_t^s|\lambda_t\right)\right)$ $(1 \le s \le 4)$. If, in addition, $e_t \sim N(0,1)$, then since $\prod_{j=0}^{\infty} E\left(\Delta_{tj}^i\right) = \frac{i^2\sigma^2}{2\left(1-\phi_1^2\right)}$, we finally get (29).

□

References

Ahmad, A., Francq, C. (2016). Poisson QMLE of count time series models. *Journal of Time Series Analysis* 37: 291-314.

Aknouche, A. (2017). Periodic autoregressive stochastic volatility. *Statistical Inference for Stochastic Processes* 20: 139–177.

Aknouche, A., Bendjeddou, S., Touche, N. (2018). Negative binomial quasi-likelihood inference for general integer-valued time series models. *Journal of Time Series Analysis* 39: 192-211.

Aknouche, A., Demmouche, N. (2019). Ergodicity conditions for a double mixed Poisson autoregression. *Statistics and Probability Letters*: 147: 6–11.

Aknouche, A., Francq, C. (2021). Count and duration time series with equal conditional stochastic and mean orders. *Econometric Theory, forthcoming.*

Atchadé, Y.F. (2006). An adaptive version for the Metropolis adjusted Langevin algorithm with a truncated drift. *Methodology and Computing in Applied Probability* 8: 235–254.

Barra, I., Borowska, A., Koopman, S.J. (2018). Bayesian dynamic modeling of high-frequency integer price changes. *Journal of Financial Econometrics* 16: 384–424.

Benjamin, M.A., Rigby, R.A., Stasinopoulos, D.M. (2003). Generalized autoregressive moving average models. *Journal of the American Statistical Association* 98: 214–23.

Berg, A., Meyer, R., Yu, J. (2004). Deviance information criterion for comparing stochastic volatility models. *Journal of Business & Economic Statistics* 22: 107–120.

Bradlow, E.T., Hardie, B.G.S, Fader, P.S. (2002). Bayesian inference for the negative binomial distribution via polynomial expansions. *Journal of Computational & Graphical Statistics* 11: 189–201.

Carrasco, M., Chen, X. (2002). Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory* 18: 17-39.

Chan, J.C.C., Eisenstat, E. (2015). Marginal likelihood estimation with the Cross-Entropy method. *Econometric Reviews*, 34: 256–285.

Chan, K.S., Ledholter, J. (1995). Monte Carlo EM Estimation for Time Series Models Involving Counts. *Journal of the American Statistical Association*, 90, 242–252.

Chen, C.W.S., So, M., Li, J.C., Sriboonchitta, S. (2016). Autoregressive conditional negative binomial model applied to over-dispersed time series of counts. *Statistical Methodology* 31: 73–90.

Chib, S. (2001). Markov chain Monte Carlo methods: computation and inference. In: *Handbook of Econometrics*, Volume 5. North Holland, Amsterdam, pp. 3569-3649.

Christou, V., Fokianos, K. (2014). Quasi-likelihood inference for negative binomial time series models. *Journal of Time Series Analysis* 35: 55–78.

Christou, V., Fokianos, K. (2015). Estimation and testing linearity for non-linear mixed poisson autoregressions. *Electronic Journal of Statistics* 9: 1357—1377.

Cox, D.R. (1981). Statistical analysis of time series: Some recent developments. *Scandinavian*

Journal of Statistics 8: 93-115.

Davis, R.A., Dunsmuir, W.T.M., Wang, Y. (1999). Modelling time series of count data. In *Asymptotics, Nonparametrics, and Time Series*. CRC Press, New York, pp. 63–114.

Davis, R.A., Dunsmuir, W.T.M., Wang, Y. (2000). On autocorrelation in a Poisson regression model. *Biometrika* 87: 491-505.

Davis, R.A., Dunsmuir, W.T.M. (2016). State space models for count time series. In: *Handbook of Discrete-Valued Time Series*. Chapman and Hall: CRC Press, pp. 121–144.

Davis, R.A., Holan, S.H., Lund, R., Ravishanker, N. (2016). *Handbook of Discrete-Valued Time Series*. Chapman and Hall: CRC Press.

Davis, R.A., Liu, H. (2016). Theory and inference for a class of observation-driven models with application to time series of counts. *Statistica Sinica* 26: 1673-1707.

Davis, R.A., Rodriguez-Yam, G. (2005). Estimation for state-space models based on likelihood approximation. *Statistica Sinica* 15: 381–406.

Davis, R.A., Wu, R. (2009). A negative binomial model for time series of counts. *Biometrika* 96: 735–749.

Dean, C., Lawless, J., Willmot, G. (1989). A mixed Poisson-Inverse-Gaussian regression model. *The Canadian Journal of Statistics* 17: 171–181.

Doukhan, P., Fokianos, K., Tjøstheim, D. (2012). On weak dependence conditions for Poisson autoregressions. *Statistics and Probability Letters* 82: 942–948.

Doukhan, P., Latour, A., Oraichi, D. (2006). A simple integer-valued bilinear time series model. *Advances in Applied Probability* 32: 559–578.

Dunn, P.K., Smyth, G.K. (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics* 5: 236–244.

Efron, B. (1986). Double exponential families and their use in generalized linear Regression. *Journal of the American Statistical Association* 81: 709–721.

Engle, R. (2002). New frontiers for ARCH models. *Journal of Applied Econometrics* 17: 425–446.

Feller, W. (1943). On a general class of “contagious” distributions. *Annals of Mathematical Statistics* 14: 389–400.

Ferland, R., Latour, A., Oraichi, D. (2006). Integer-valued GARCH process. *Journal of Time Series Analysis* 27: 923-942.

Fokianos, K., Rahbek, A., Tjøstheim, D. (2009). Poisson autoregression. *Journal of the American Statistical Association* 140: 1430–1439.

Fokianos, K. (2016). Statistical analysis of count time Series models: A GLM perspective. In: *Handbook of Discrete-Valued Time Series*. Chapman and Hall: CRC Press.

Francq, C., Zakoian, J.M. (2019). *GARCH Models: Structure, Statistical Inference and Applications*. Wiley.

Freeland, R.K., McCabe, B.P.M. (2004). Forecasting Discrete Valued Low Count Time Series. *International Journal of Forecasting*, 20: 427–434.

Frühwirth-Schnatter, S., Frühwirth, R., Held, L., Rue, H. (2009). Improved auxiliary mixture sampling for hierarchical models of non-Gaussian data. *Statistics and Computing*, 19: 479.

Frühwirth-Schnatter, S., Wagner, H. (2006). Auxiliary mixture sampling for parameter-driven models of time series of counts with applications to state space modelling. *Biometrika* 93: 827–841.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In: Bayesian Statistics 4, Oxford: Clarendon Press, pp. 641-649.

Geweke, J., Amisano, G. (2011a). Optimal prediction pools. *Journal of Econometrics* 164: 130–141.

Geweke, J., Amisano, G. (2011b). Hierarchical Markov normal mixture models with applications to financial asset returns. *Journal of Applied Econometrics* 26:1–29.

Graham, R.L., Knuth, D.E., Patashnik, O. (1989). *Concrete Mathematics. A Foundation for Computer Science*. Reading, MA: Addison Wesley.

Grunwald, G., Hyndman, R.J.S., Tedesco, L., Tweedie, R.L. (2000). Theory and methods: Non-Gaussian conditional linear AR(1) models. *Australian & New Zealand Journal of Statistics* 42: 479–495.

Hay, J., Pettitt, A. (2001). Bayesian analysis of a time series of counts with covariates: an application to the control of an infectious disease. *Biostatistics* 2: 433–444.

Hinde, J.P. (1982). Compound Poisson regression models. In: Proceedings of the International Conference on Generalised Linear Models. Lecture Notes in Statistics, Vol 14. Springer, pp. 109–121.

Jørgensen, B. (1997). *The Theory of Dispersion Models*. Chapman & Hall, London.

Jung, R.C., Kukuk, M., Liesenfeld, R. (2006). Time series of count data: modeling, estimation and diagnostics. *Computational Statistics & Data Analysis* 51: 2350–2364.

Jung, R.C., Liesenfeld, R., Richard, J.-F. (2011). Dynamic Factor Models for Multivariate Count Data: An Application to Stock-Market Trading Activity. *Journal of Business and Economic Statistics*, 29: 73–86.

Kim, S., Shephard, N., Chib, S. (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models. *Review of Economic Studies* 65: 361-393.

Kokonendji, C.C., Dossou-Gbété, S., Demétrio, C.G.B. (2004). Some discrete exponential dispersion models: Poisson-Tweedie and Hinde-Demétrio classes. *Statistics and Operations Research Transactions* 28: 201—213.

Koop, G., Poirier, D.J., Tobias, J.L. (2012). *Bayesian Econometric methods (1st edn.* Cambridge University Press, New York.

Liesenfeld, R., Nolte, I., Pohlmeier, W. (2006). Modelling financial transaction price movements: a dynamic integer count data model. *Empirical Economics* 30: 795–825.

Leroux, B.G. (1992). Maximum likelihood estimation for hidden Markov models. *Stochastic Processes and their Applications* 40: 127–143.

McCabe, B.P.M., Martin, G.M. (2005). Bayesian predictions of low count time series. *International Journal of Forecasting* 21: 315–330.

Martino, L., Yang, H., Luengo, D., Kanniainen, J., Corander, J. (2015). A Fast Universal Self-Tuned Sampler Within Gibbs Sampling. *Digital Signal Processing* 47: 68–83.

Meyn, S.P., Tweedie, R.L. (2009). *Markov Chains and Stochastic Stability.* Springer Verlag, New York.

Mikosch, T. (2009). *Non-life Insurance Mathematics, An Introduction with the Poisson Process.* Springer-Verlag, Berlin.

Neumann, M.H. (2011). Absolute regularity and ergodicity of Poisson count processes. *Bernoulli* 17: 1268–1284.

Priestley, M.B. (1981). *Spectral analysis and time series.* New York: Academic Press.

Roberts, G.O., Rosenthal, J.S. (1998). Optimal scaling of discrete approximations to Langevin diffusions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 60: 255-–268.

Roberts, G.O., Tweedie, R.L. (1996). Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli* 2: 341-363.

Rydberg, T.H., Shephard, N. (2000). BIN models for trade-by-trade data. Modelling the number of trades in a fixed interval of time. *Technical Report 0740, Econometric Society.*

Silva, R.B., Barreto-Souza, W. (2019). Flexible and robust mixed Poisson INGARCH models. *Journal of Time Series Analysis* 40: 788–814.

Sørensen, H. (2019). Independence, successive and conditional likelihood for time series of counts. *Journal of Statistical Planning and Inference* 200: 20–31.

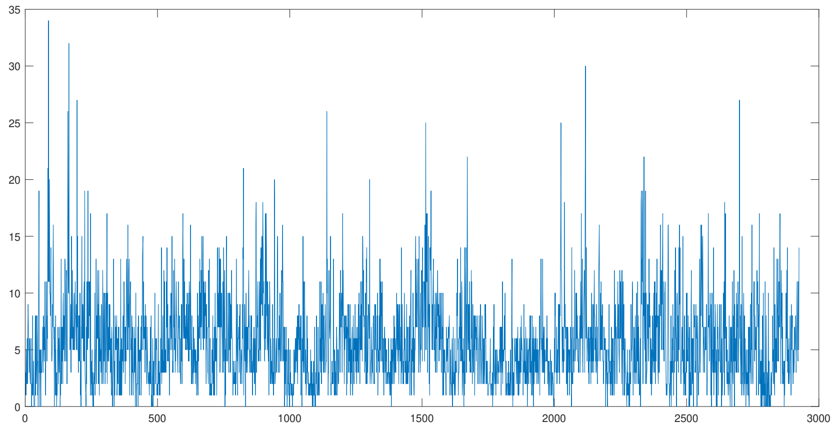
Weiß, C.H (2009). Modelling time series of counts with overdispersion. *Statistical Methods and Applications*, 18:507–519.

Weiß, C.H. (2017). *An Introduction to Discrete-Valued Time Series.* Wiley.

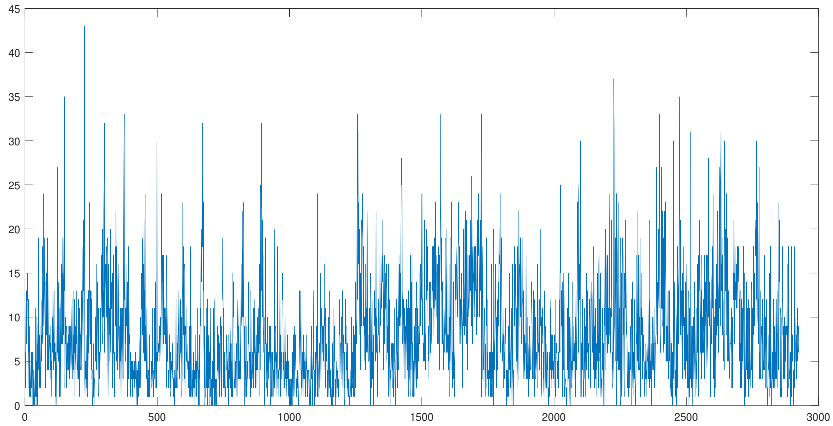
Zeger, S.L. (1988). A regression model for time series of counts. *Biometrika* 75: 621–629.

Zeger, S., Qaqish, B. (1988). Markov regression models for time series: A quasi-likelihood approach. *Biometrics* 44: 1019–1031.

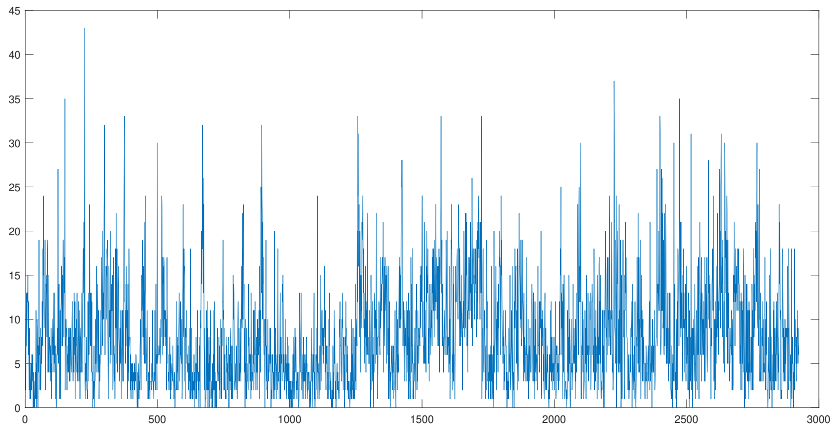
Zhu, F. (2011). A negative binomial integer-valued GARCH model. *Journal of Time Series Analysis* 32: 54–67.



(a) GLT.

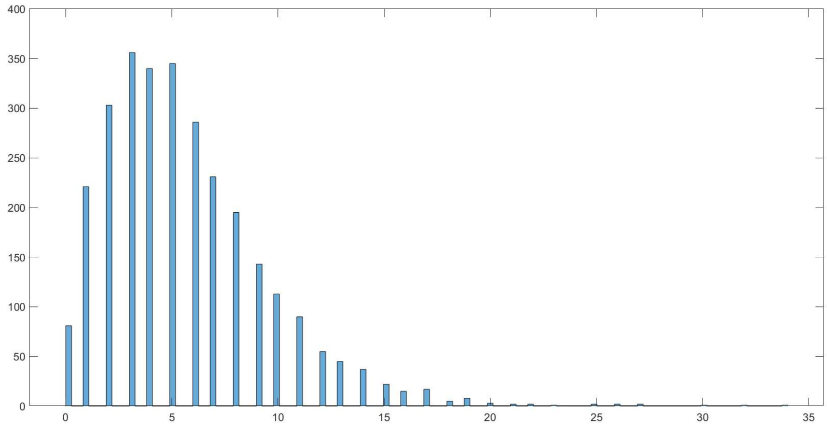


(b) EDE.

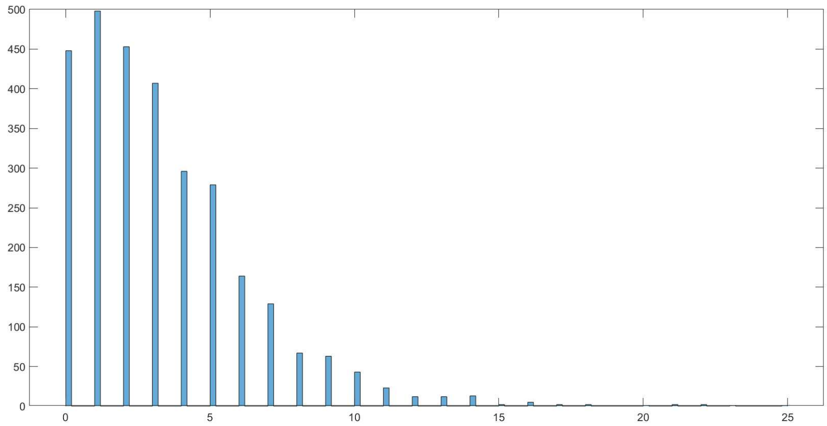


(c) WPP.

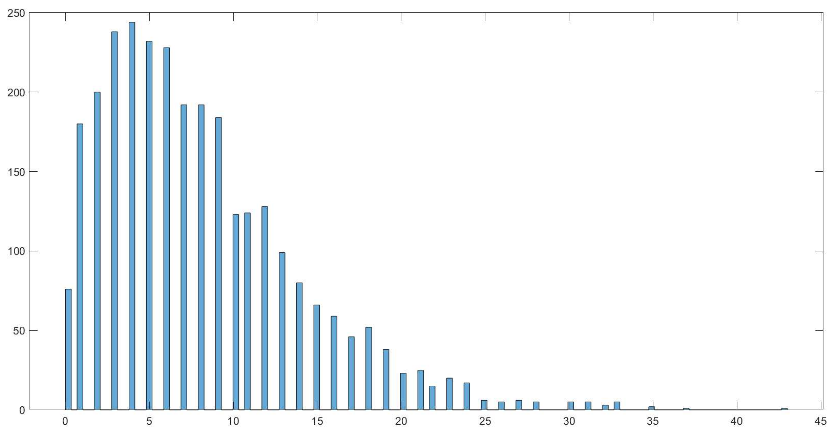
Figure 1: Empirical results: Plot of the three financial time series.



(a) GLT.



(b) EDE.



(c) WPP.

Figure 2: Empirical results: Histograms of the three financial time series.

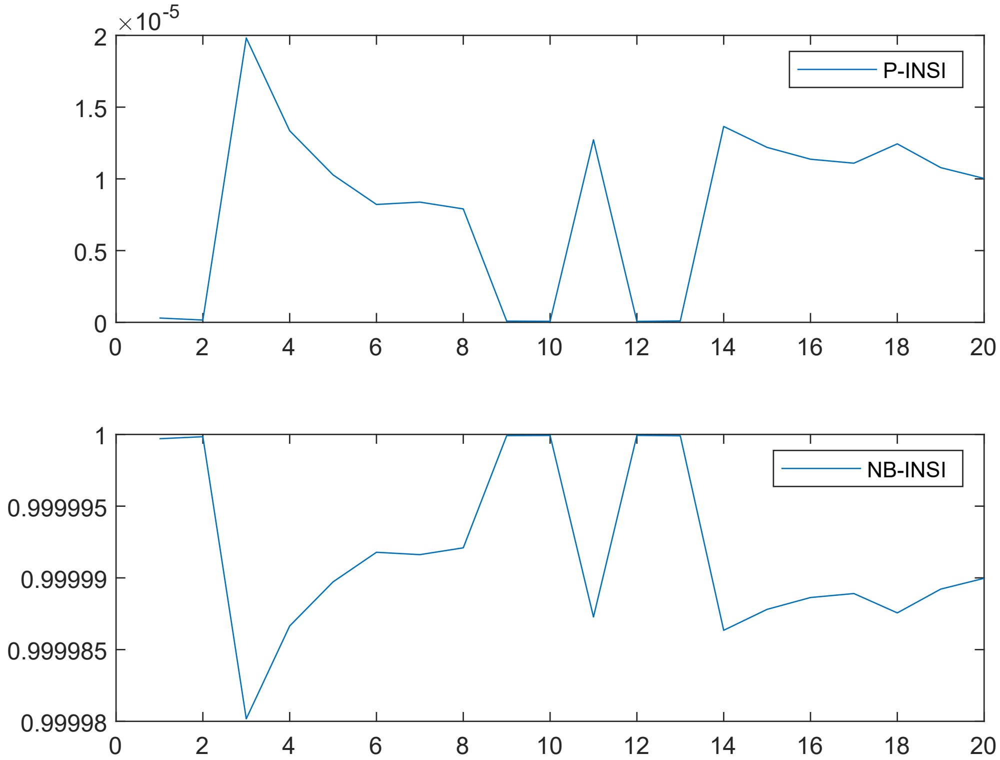
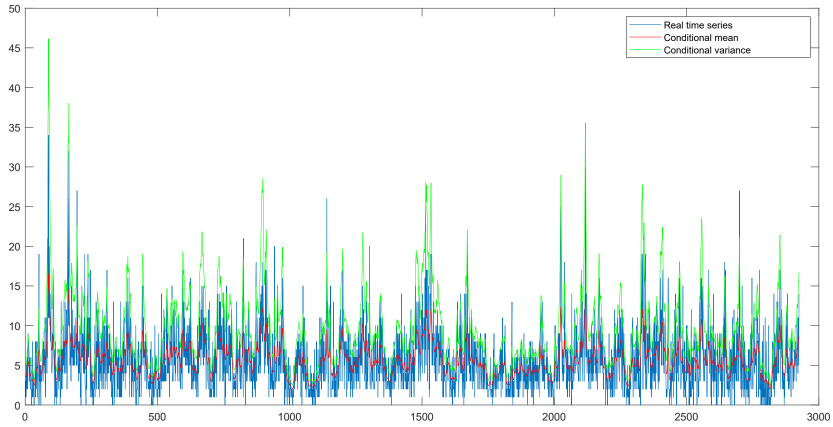
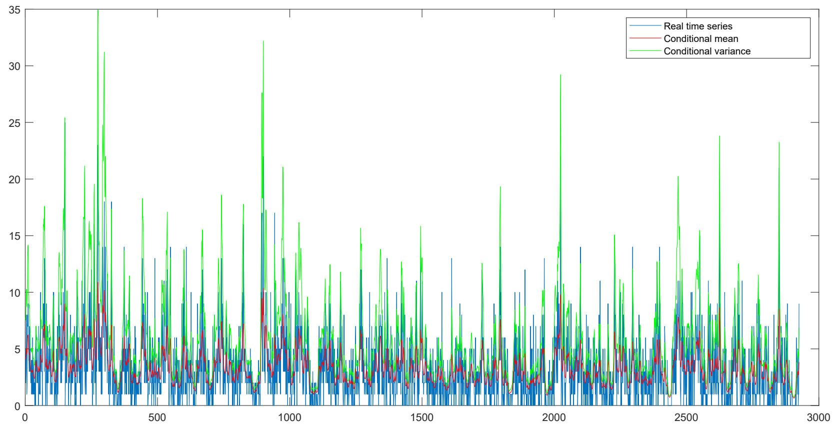


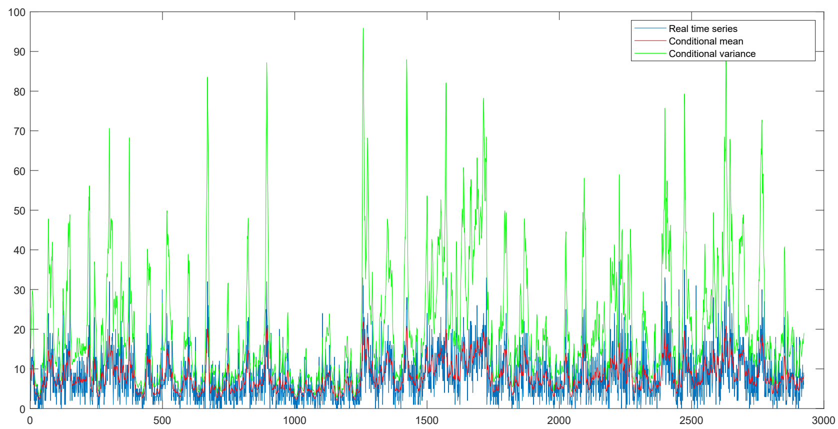
Figure 3: Empirical results for the financial data set GLT. Evolution of model weights in the two-model pool of one-step ahead predictive densities.



(a) GLT.

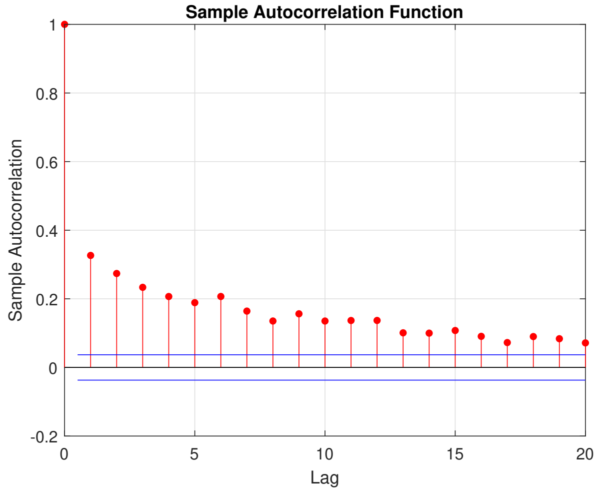


(b) EDE.

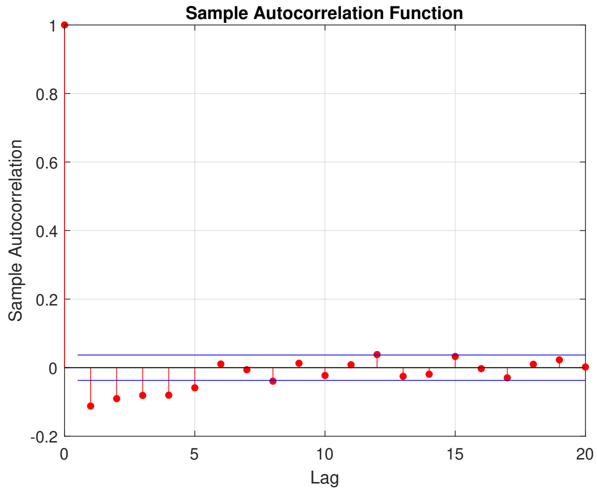


(c) WPP.

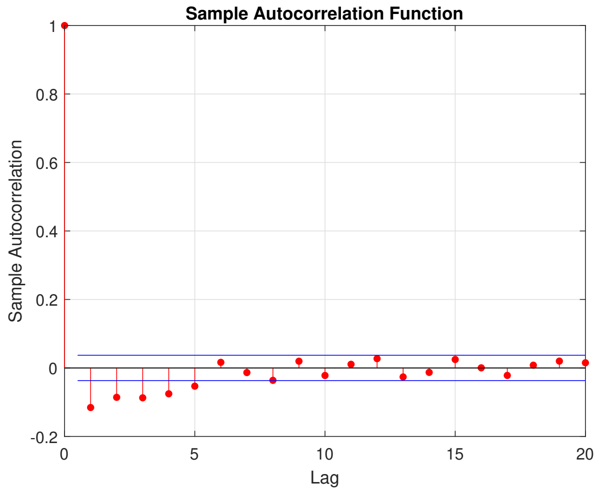
Figure 4: Empirical results for the financial data: Real series and estimated conditional mean and variance obtained from the NB- $INSI$ model.



(a) Sample autocorrelations of the GLT series.



(b) Sample autocorrelations of the Y-residuals.



(c) Sample autocorrelations of the randomized Y-residuals.

Figure 5: Empirical results for the financial data GLT. Residual autocorrelation analysis for the NB-$INSI$ model.

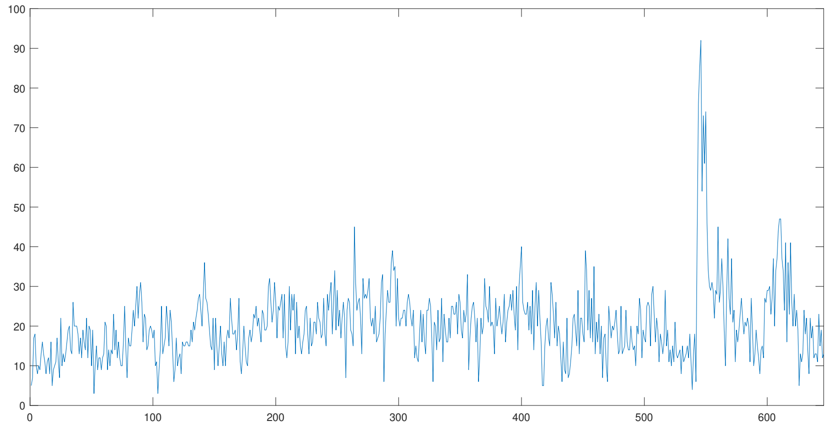


Figure 6: Empirical results: Plot of the E.coli time series.

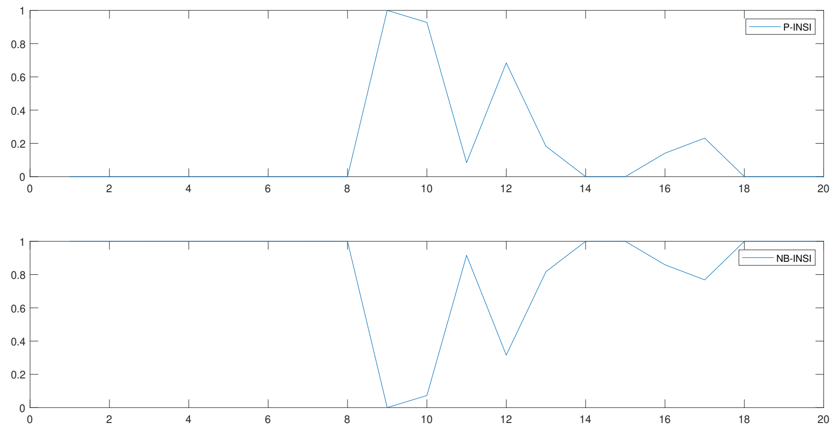


Figure 7: Empirical results for the E.coli data. Evolution of model weights in the two-model pool of one-step ahead predictive densities.

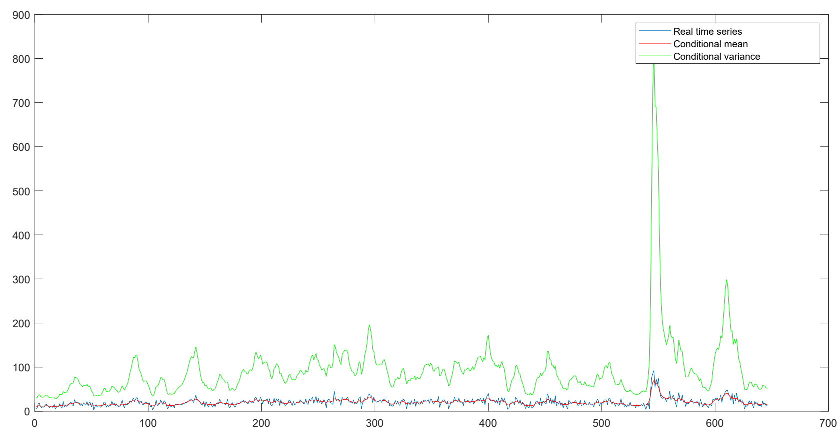
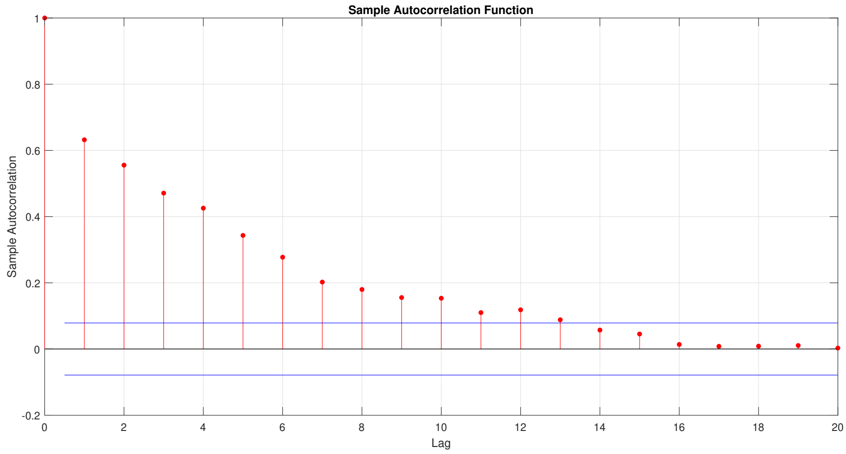
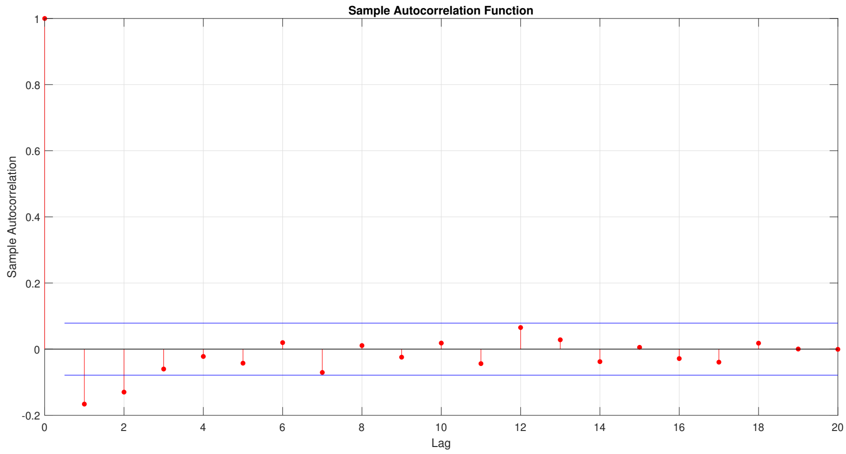


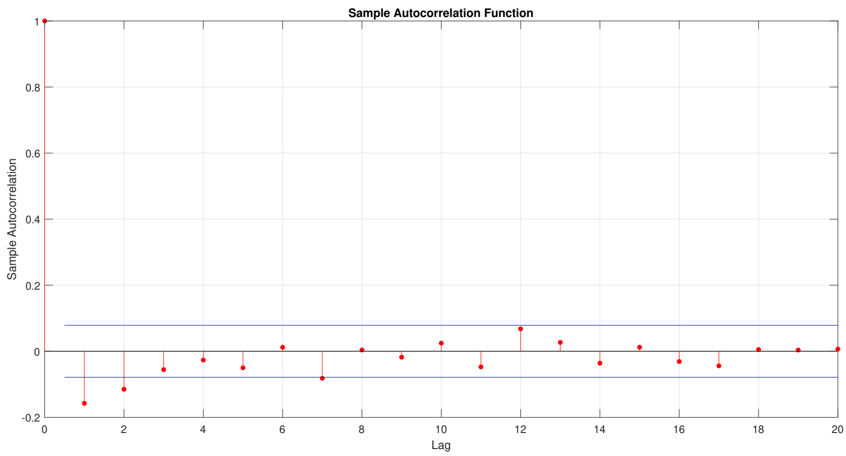
Figure 8: Empirical results for the E.coli data: Real series and estimated conditional mean and variance obtained from the NB- $INSI$ model.



(a) Sample autocorrelations of the E.coli series.



(b) Sample autocorrelations of the Y-residuals.



(c) Sample autocorrelations of the randomized Y-residuals.

Figure 9: Empirical results for the E.coli data. Residual autocorrelation analysis for the $NB\text{-}INSI$ model.

Table 1: Simulation results for the *P-INSI* model (n=1000).

	ϕ_0	ϕ_1	σ^2
True	0.5	0.5	0.5
Mean	0.4521	0.4753	0.4553
Std	0.0468	0.0480	0.0443
IF	1.5421	6.6377	10.237
CD	0.8217	0.1832	0.6925
True	1.2	0.9	0.2
Mean	1.141	0.8985	0.1711
Std	0.1317	0.0157	0.0135
IF	1.0298	2.1512	5.3778
CD	1.6501	−0.5089	−0.2514
True	0.1	0.3	0.1
Mean	0.0694	0.2336	0.1135
Std	0.0266	0.1312	0.0217
IF	9.7371	45.218	39.082
CD	0.8656	0.3712	−1.3626

Table 2: Simulation results for the *NB-INSI* model (n=1000).

True	$\phi_0 = 1$	$\phi_1 = 0.6$	$\sigma^2 = 0.2$	$\tau = 7$
Mean	0.94128	0.5938	0.2009	6.9872
Std	0.0399	0.0487	0.0245	1.6761
IF	1.5057	16.158	22.842	26.054
CD	0.1072	−0.1116	−0.1544	0.3449
True	$\phi_0 = 1.5$	$\phi_1 = 0.8$	$\sigma^2 = 0.4$	$\tau = 6$
Mean	1.5695	0.79825	0.40576	6.2843
Std	0.1037	0.0252	0.0404	1.3102
IF	1.0468	7.9823	15.687	29.094
CD	1.1652	0.0576	−0.0691	0.4560
True	$\phi_0 = 2$	$\phi_1 = 0.5$	$\sigma^2 = 0.1$	$\tau = 5$
Mean	2.005	0.4897	0.1061	4.9326
Std	0.0247	0.0625	0.0147	0.5529
IF	1.7169	17.183	24.573	20.671
CD	0.5869	−0.9878	0.9223	0.7136

Table 3: Empirical results for financial data. Descriptive statistics.

	Sample Mean	Sample Variance
GLT	5.6632	15.2070
EDE	3.3138	9.2086
WPP	8.1115	35.5382

Table 4: Empirical results for financial data. Competing mixed Poisson models

Model	GLT (P-$INSI$)			EDE (P-$INSI$)			WPP (P-$INSI$)			GLT (NB-$INSI$)			EDE (NB-$INSI$)			WPP (NB-$INSI$)		
	Mean	IF	CD	Mean	IF	CD	Mean	IF	CD	Mean	IF	CD	Mean	IF	CD	Mean	IF	CD
ϕ_0	1.6059*	1.4899	-1.1431	0.9643*	1.8106	-0.3830	1.8933*	1.2096	-1.4585	1.6556*	1.3474	-0.4803	1.0557*	3.7415	-1.1944	1.9677*	1.1736	0.2220
	(0.0248)			(0.0316)			(0.0293)			(0.0377)			(0.0432)			(0.0458)		
ϕ_1	0.7052*	13.101	-2.2474	0.6638*	14.99	-0.1564	0.6899*	5.2679	0.1095	0.9137*	46.139	-1.1018	0.8750*	59.074	-0.8670	0.9093*	37.095	0.2851
	(0.0245)			(0.027)			(0.0192)			(0.0134)			(0.0180)			(0.0128)		
σ^2	0.1323*	18.426	2.2985	0.2748*	19.024	-0.0600	0.2208*	10.284	1.3006	0.0277*	60.342	1.4457	0.0695*	45.132	1.0877	0.0462*	66.842	-0.0862
	(0.0105)			(0.0197)			(0.0120)			(0.0041)			(0.0115)			(0.0066)		
τ										9.1157*	21.459	1.0989	4.8963*	69.381	0.777	6.4353*	30.039	0.6873
										(0.8125)			(0.5314)			(0.4921)		
Log CML	-6888.2			-5848.4			-7721.6			-5938.7			-4878.1			-6179.3		
	(0.21)			(0.11)			(0.16)			(0.70)			(0.82)			(0.99)		
LPS	-45.136			-34.893			-52.23			-44.006			-29.7885			-48.5892		
RMSFE	294.63			20.912			640.21			285.98			19.25			632.41		

*Significant based on the 95% highest posterior density interval. Standard deviation in parentheses (for the estimated parameters). For the Log ML estimates, we also report the numerical standard errors in parentheses. CD stands for Converge Diagnostics and IF stands for Inefficiency Factor.

Table 5: Empirical results for financial data. Optimal pool.

	P-$INSI$	NB-$INSI$
GLT	1.003e-05	0.999
EDE	3.0901e-07	1
WPP	2.7304e-07	1

Table 6: Empirical results for E.coli data. Competing mixed Poisson models

Model	*P-INSI*			*NB-INSI*		
	Mean	IF	CD	Mean	IF	CD
ϕ_0	2.9397*	1.0563	-0.6737	2.9466*	1.08	2.0597
	(0.0463)			(0.0599)		
ϕ_1	0.8075*	6.0239	0.2988	0.9012*	16.698	0.5741
	(0.032)			(0.0238)		
σ^2	0.0462*	8.4716	0.5575	0.0194*	31.863	-1.2733
	(0.0053)			(0.0034)		
τ				34.333*	9.3764	-0.9397
				(5.3834)		
Log CML	-1929.3			-1744.1		
LPS	-57.894			-55.398		
RMSFE	19.321			18.555		

*Significant based on the 95% highest posterior density interval. Standard deviation in parentheses (for the estimated parameters). For the Log ML estimates, we also report the numerical standard errors in parentheses. CD stands for Converge Diagnostics and IF stands for Inefficiency Factor.