Integer-valued stochastic volatility

Aknouche, Abdelhakim and Dimitrakopoulos, Stefanos and Touche, Nassim

USTHB and Qassim university, Leeds University, University of Bejaia

4 February 2019

Integer-valued stochastic volatility

Abdelhakim Aknouche[*], Stefanos Dimitrakopoulos[1**], and Nassim Touche[***]

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

[**]Economics Division, Leeds University, UK

[***]Department of Operational Research, University of Bejaia, Algeria

Abstract

We propose a novel class of count time series models, the mixed Poisson integer-valued stochastic volatility models. The proposed specification, which can be considered as an integer-valued analogue of the discrete-time stochastic volatility model, encompasses a wide range of conditional distributions of counts. We study its probabilistic structure and develop an easily adaptable Markov chain Monte Carlo algorithm, based on the Griddy-Gibbs approach that can accommodate any conditional distribution that belongs to that class. We demonstrate that by considering the cases of Poisson and negative binomial distributions. The methodology is applied to simulated and real data.

Keywords: Griddy-Gibbs, Markov chain Monte Carlo, mixed Poisson parameter-driven models, stochastic volatility.

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.

1 Introduction

Nowadays, time series count data models (Cameron and Trivedi, 2013) 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, undersispersion, 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. Since the model distribution is conditioned on this parameter, we suggest categorizing count data models that involve a dynamic specification for the intensity parameter into 'observed conditional intensity models' and 'unobserved conditional intensity models'[1]. This paper deals with the theory and inference of the latter models.

As is well known, observed intensity models, which mainly include integer-valued generalized autoregressive conditional heteroscedastic ($INGARCH$) 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 (2018).

In contrast, unobserved intensity models, although they do not admit a weak $ARMA$ representation, are generally of 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 principle, these models are estimated by filtering and signal extraction- based methods, such as Bayesian Markov chain Monte Carlo (MCMC) and Expectation–Maximization (EM)-type algorithms.

The literature on time series of counts has put forward parameter-driven models, which do not consider a dynamic equation for the latent intensity parameter (Zeger, 1998; Davis et al., 1999, 2000; Hay and Pettitt, 2001; Davis and Wu, 2009) and unobserved intensity models, that is, parameter-driven models with a dynamic specification for the intensity parameter (Davis and Rodriguez-Yam, 2005; Jung et al., 2006; Sorensen, 2019). In the first 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.

In the second case of unobserved intensity models, an autoregressive process (without an intercept), driven by Gaussian innovations, is assigned to a latent multiplicative or additive component of the intensity equation. Yet, all the previous research on unobserved intensity models is restricted solely to the Poisson distribution with an exponential conditional mean (which is usually a function of

[1]The terms 'observed conditional intensity models' and 'observed intensity models' are used interchangeably throughout the paper. Similarly for the 'unobserved conditional intensity models'. It is also important to highlight that the 'unobserved conditional intensity model' should not be confused with the completely different model of Bauwens and Hautsch (2006), called the 'stochastic conditional intensity model'.

covariates as well). In addition, the probabilistic properties of (Poisson-based) unobserved intensity models have not been studied so far. As such, the extant literature lacks a general framework for modeling, estimating and studying the theoretical properties of unobserved intensity count time series models. The present paper aspires to fill these gaps.

We propose a broad class of unobserved intensity models for count data, the mixed Poisson integer-valued stochastic volatility ($INSV$) models. 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 $INSV$ model, we show that the present framework can be easily generalized to account for different stochastic processes that are all based on the general $INSV$ model. Different stochastic processes lead to different $INSV$-type models that correspond to different families of conditional distributions (e.g, the exponential family). These distributions do not necessarily belong to the class of the mixed Poisson $INSV$ process.

The mixed Poisson $INSV$ model can be considered as the integer-valued analogue of the stochastic volatility model (Taylor, 1986) for real-valued time series; hence the term "integer-valued stochastic volatility". As we explain, though, in more detail in Section 2, this term is somewhat a misnomer. Furthermore, since the $INSV$ processes can be seen as flexible alternatives to the $INGARCH$ processes (see, for example, Christou and Fokianos (2014)), the present work also complements the count time series literature on observed intensity models.

We study the probabilistic path properties of the mixed Poisson $INSV$ 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, conditional of the intensity parameter, the likelihood function has a closed form and parameter estimation can be achieved by MCMC methods. The proposed posterior sampler can be easily modified to accommodate any conditional distribution that belongs to the family of the mixed Poisson $INSV$ process (or of any $INSV$-type process). To demonstrate that, we consider two specific cases of the mixed Poisson $INSV$ specification, the Poisson $INSV$ model (P-$INSV$) and the negative binomial $INSV$ model (NB-$INSV$). For both models, the parameters of the autoregression are assigned conjugate priors and are updated from well-defined conditional posterior distributions.

The only difficult updating steps concern the vector of unobserved intensities in both models and the dispersion parameter in the negative binomial case. Since the joint conditional posterior of the latent intensities is of unknown form, we adopt the Griddy-Gibbs technique (Ritter and Tanner, 1992) and sample them one at a time (element-by-element updating), in the spirit of Jacquier et al., (1994). The same technique is used for sampling the dispersion parameter of the negative binomial $INSV$ model. For the negative binomial case, a modified scale mixture representation, in the spirit of Jacquier et al, (2004), is also used to improve efficiency. Model selection is conducted using the Deviance Information Criterion (Spiegelhalter et al., 2002).

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 financial and health data. In particular, we exploit the transaction data used by Fokianos et al., (2009) and the Zeger (1988)'s poliomyelitis data. Both data sets have been widely used in the past in the context of count data models. Although no covariates are considered in our analysis, the specified drifted autoregressive latent process not only

adds serial dependence and overdispersion to the proposed model, but can also be viewed as a proxy for unknown/unavailable covariates (Davis and Wu, 2009).

The paper is structured as follows. In section 2 we set up the proposed mixed Poisson $INSV$ model, examine its probabilistic properties and show how the modelling approach taken here can be generalized to account for other $INSV$-type models. In section 3 we describe the prior-posterior analysis for the two cases of the proposed specification (P-$INSV$ and NB-$INSV$), while in section 4 we perform a simulation study. In section 5 we carry out our empirical analysis. Section 6 concludes.

2 The mixed Poisson integer-valued SV model

2.1 The set up

Consider the unknown real parameters ϕ_0 and ϕ_1 and an independent 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 volatility ($INSV$) 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 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 model (1)-(2) is novel and enlarges the existing framework of unobserved intensity models. The family of processes represented by (1) is known as mixed Poisson process with mixing variable Z_t (Mikosch, 2009). 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/λ_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 $INSV$ model (P-$INSV$). 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 $INSV$ model (NB-$INSV$). This model is characterized by conditional overdispersion, i.e., $var\left(Y_t/\lambda_t\right) = \lambda_t + p^2\lambda_t^2 > E\left(Y_t/\lambda_t\right) = \lambda_t$.

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/λ_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 $INSV$ 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 $INSV$ 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 $INSV$-type models; see a remark of section 2.2. What is more, this paper complements the research undertaken on the observed intensity models, as our modelling approach can also be viewed as a flexible alternative to the $INGARCH$ processes, for which the intensity parameters depend only on the past process.

At this point, we would like to comment on the terminology used in this paper. In the P-$INSV$ model, the intensity parameter λ_t (conditional mean) is equal to the conditional variance $var\left(Y_t/\lambda_t\right)$. Since the conditional variance is named 'volatility' in the financial literature, and the $\log(\lambda_t)$ in (2) follows a Gaussian autoregressive process, the latent log-intensity can be regarded as the equivalent of the latent log-volatility, denoted by $\log(h_t)$, of the real-valued SV model (Taylor, 1986), which also evolves according to the same process. Therefore, due to the fact that $\lambda_t = var\left(Y_t/\lambda_t\right) := h_t$, the Poisson $INSV$ model is the discrete analog of the stochastic volatility model, and as such the terminology 'Poisson $INSV$' (or simpler the 'Poisson SV') seems pertinent to describe the model (2)-(3).

On the contrary, the terminology 'negative binomial $INSV$', which is used to describe the model, given by (2) and (4), is somewhat a misnomer. The reason is that, in this case, the intensity λ_t is no longer equal to the volatility, namely, $\lambda_t \neq var\left(Y_t/\lambda_t\right) := h_t$. However, the log-volatility in the SV model generally follows a Gaussian autoregression, as the log-intensity in the NB-$INSV$ model also does. Based on this similarity, we retain the term '$INSV$' not only for the Poisson case but also for the negative binomial case and hence for the proposed mixed Poisson process. In this way, we achieve a consistent terminology throughout the paper as well.

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. As opposed to the SV specification that can be represented by a MEM form, the mixed Poisson $INSV$ 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 $INSV$ 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 $INSV$ model[2].

[2]The same strategy is used by the literature on $INGARCH$ models (Fokianos et al., 2009; Christou and Fokianos, 2014; Davis and Liu, 2016; Aknouche and Francq, 2018; Aknouche and Demouche, 2019).

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

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

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

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

It is well known that 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.* □

From Theorem 1 we see that ϕ_1 is the analog of the persistent parameter in the case of real-valued SV and $GARCH$ models. Other properties, such as geometric ergodicity and 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

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

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

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 volatility 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*

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

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

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

Proof. *Appendix.* □

To calculate the variance of the mixed Poisson INSV 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*

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

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 $INSV$ model is conditionally equidispersed but unconditionally overdispersed as

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

The negative binomial $INSV$ 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 $\gamma_h = E\left(Y_t Y_{t-h}\right) - E\left(Y_t\right)E\left(Y_{t-h}\right)$ 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 $INSV$ model and we restrict our attention to the Poisson $INSV$ 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*

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

If, in addition, $e_t\sim N\left(0,1\right)$, 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 $E\left(Y_t^s\right)$, $(s\geq 1)$ 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 $INSV$ 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 $INSV$ 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-INSV* and *NB-INSV* models, we show that the mixed Poisson $INSV$ model follows from a general $INSV$ model.

Remark. *The mixed Poisson $INSV$ model is a special case of a general $INSV$ model.*

In section 2.1 we defined the mixed Poisson $INSV$ 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 $INSV$ 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 (2018)).

Let us consider the general $INSV$ 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/λ_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 (2018). 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 $INSV$ 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 $INSV$ model. When F_{λ_t} is the cdf of the negative binomial distribution, then we obtain the negative binomial $INSV$ 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 $E(X_t/\lambda_t) = E[E(X_t/\lambda_t)] = E\left[E\left(F^-_{\lambda_t}(U_t)/\lambda_t\right)\right] = E(\lambda_t)$.

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

3 Bayesian inference via Griddy-Gibbs sampling

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

The second case refers to the negative binomial $INSV$ 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 $\theta = \left(\phi', \tau, \sigma^2\right)'$, where $\tau = \rho^{-2}$ (the dispersion parameter).

3.1 Estimating the Poisson $INSV$ model

Following the Bayesian paradigm, the parameter vector θ and the unobserved intensities $\lambda = (\lambda_1, ..., \lambda_n)'$ are viewed as random with a prior distribution $f(\theta, \lambda)$. 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 estimate the joint posterior distribution $f(\theta, \lambda/Y)$. This can be done using the Gibbs sampler, provided that we can draw samples from any of the following three conditional posterior distributions $f\left(\phi/Y, \sigma^2, \lambda\right)$, $f\left(\sigma^2/Y, \phi, \lambda\right)$ and $f\left(\lambda/Y, \phi, \sigma^2\right)$.

Since the posterior distribution $f\left(\lambda/Y, \phi, \sigma^2\right)$ has a rather complicated form, the vector λ is updated (in the spirit of Jacquier et al., (1994)) element-by-element, using the Griddy-Gibbs sampler (Ritter and Tanner, 1992). In this single-move framework, the Gibbs sampler reduces to drawing samples from any of the $n+2$ conditional posterior distributions $f\left(\phi/Y, \sigma^2, \lambda\right)$, $f\left(\sigma^2/Y, \phi, \lambda\right)$ and $f\left(\lambda_t/Y, \phi, \sigma^2, \lambda_{-\{t\}}\right)$ $(1 \le t \le n)$, where $\lambda_{-\{t\}}$ denotes the λ vector after removing its t-th component λ_t. Under the normality of the innovation of the log-intensity equation and using standard linear regression theory (Box and Tiao, 1973), the conditional posteriors $f\left(\phi/Y, \sigma^2, \lambda\right)$ and $f\left(\sigma^2/Y, \phi, \lambda\right)$ can be determined directly from the conjugate priors for ϕ and σ^2.

Sampling the autoregressive parameter ϕ

Setting $\Lambda_t = (1, \log(\lambda_{t-1}))'$, equation (2) can be rewritten in the following linear regression form

$$\log(\lambda_t) = \Lambda_t'\phi + \sigma e_t, \tag{32}$$

with an *i.i.d* Gaussian innovation $\{e_t, t \in \mathbb{Z}\}$. To get a closed-form expression for the conditional posterior $f\left(\phi/Y, \sigma^2, \lambda\right)$, we use a conjugate prior for ϕ. This prior is Gaussian, $\phi \sim N\left(\phi^0, \Sigma^0\right)$, where the hyperparameters ϕ^0, Σ^0 are known and fixed to values that yield a quite reasonable diffuse prior. The conditional posterior distribution of ϕ given Y, σ^2, λ is

$$\phi/Y, \sigma^2, \lambda \sim N\left(\phi^*, \Sigma^*\right), \tag{33}$$

where

$$\Sigma^* = \left(\sum_{t=1}^{n} \tfrac{1}{\sigma^2} \Lambda_t \Lambda_t' + \left(\Sigma^0\right)^{-1}\right)^{-1} \tag{34}$$

$$\phi^* = \Sigma^* \left(\sum_{t=1}^{n} \tfrac{1}{\sigma^2} \Lambda_t \log\left(\lambda_t\right) + \left(\Sigma^0\right)^{-1} \phi^0\right). \tag{35}$$

Sampling the variance parameter σ^2

As a conjugate prior for σ^2 we use the inverted *Chi*-squared distribution, i.e.,

$$\tfrac{a\nu}{\sigma^2} \sim \chi_a^2, \tag{36}$$

where $a\nu = 1$. Given the parameters ϕ and λ, if we define

$$e_t = \log\left(\lambda_t\right) - \phi_0 - \phi_1 \log\left(\lambda_{t-1}\right), \qquad 1 \leq t \leq n, \tag{37}$$

then $e_1, e_2, ..., e_n \overset{i.i.d}{\sim} N\left(0, \sigma^2\right)$. The conditional posterior distribution of σ^2 is an inverted *Chi*-squared distribution with $a + n - 1$ degrees of freedom, that is,

$$\tfrac{a\nu + \sum_{t=1}^{n} e_t^2}{\sigma^2} / Y, \phi, \lambda \sim \chi_{a+n-1}^2. \tag{38}$$

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

It remains to sample from the conditional posterior distribution $f\left(\lambda_t/Y, \theta, \lambda_{-\{t\}}\right)$, $t = 1, 2, ..., n$. Let us first derive the kernel of this distribution and we will, then, show how to (indirectly) draw samples from it using the Griddy-Gibbs technique (Ritter and Tanner, 1992).

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 $1 < t < n$

$$\begin{aligned} f\left(\lambda_t/Y, \theta, \lambda_{-\{t\}}\right) &= \tfrac{f(\lambda_t/\lambda_{t-1}, \theta) f(\lambda_{t+1}/\lambda_t, \theta) f(Y_t/\theta, \lambda_t)}{f(\lambda_{t+1}/\lambda_{t-1}, \theta) f(Y_t/\theta, \lambda_{t-1}, \lambda_{t+1})} \\ &\propto f\left(\lambda_t/\lambda_{t-1}, \theta\right) f\left(\lambda_{t+1}/\lambda_t, \theta\right) f\left(Y_t/\theta, \lambda_t\right). \end{aligned} \tag{39}$$

Using the fact that $Y_t/\theta, \lambda_t \equiv Y_t/\lambda_t \sim \mathcal{P}\left(\lambda_t\right), \log\left(\lambda_t\right)/\log\left(\lambda_{t-1}\right), \theta \sim N\left(\phi_0 + \phi_1 \log\left(\lambda_{t-1}\right), \sigma^2\right)$, and $d\log(\lambda_t) = \frac{1}{\lambda_t} d\lambda_t$, formula (39) becomes

$$f\left(\lambda_t/Y, \theta, \lambda_{-\{t\}}\right) \propto \exp\left(-\lambda_t + (Y_t - 1)\log\left(\lambda_t\right) - \tfrac{1}{2\Omega}\left(\log\left(\lambda_t\right) - \mu_t\right)^2\right), \quad 1 < t < n, \tag{40}$$

where

$$\mu_t = \frac{\phi_0(1-\phi_1)+\phi_1(\log(\lambda_{t-1})+\log(\lambda_{t+1}))}{1+\phi_1^2} \quad (41)$$

$$\Omega = \frac{\sigma^2}{1+\phi_1^2}. \quad (42)$$

Once the kernel of $f\left(\lambda_t/Y,\theta,\lambda_{-\{t\}}\right)$ is determined, one can use some indirect sampling algorithms to draw the intensity λ_t. In this paper, we adopt the Griddy-Gibbs technique (Ritter and Tanner, 1992), which works as follows:

(i) Choose a grid of m points from a given interval $[\lambda_{t1},\lambda_{tm}]$ of λ_t: $\lambda_{t1} \leq \lambda_{t2} \leq ... \leq \lambda_{tm}$. Then, evaluate the conditional posterior $f\left(\lambda_t/Y,\theta,\lambda_{-\{t\}}\right)$ via (40)-(42) at each one of these points, to obtain $f_{ti} = f\left(\lambda_{ti}/Y,\theta,\lambda_{-\{t\}}\right)$, $i = 1,...,m$.

(ii) From the values $f_{t1}, f_{t2}, ..., f_{tm}$ construct the discrete distribution $p(.)$, defined at λ_{ti} $(1 \leq i \leq m)$, by setting $p(\lambda_{ti}) = \frac{f_{ti}}{\sum_{j=1}^{m} f_{tj}}$. This may be seen as an approximation to the inverse cumulative distribution of $f\left(\lambda_t/Y,\theta,\lambda_{-\{t\}}\right)$.

(iii) Generate a number from the uniform distribution on the interval $(0,1)$ and transform it using the discrete distribution $p(.)$, obtained in (ii), to get a random draw for λ_t.

It is worth noting that the choice of the grid $[\lambda_{t1},\lambda_{tm}]$ is crucial for the efficiency of the Griddy-Gibbs algorithm. We follow here an approach similar to Tsay (2010), which consists of taking the range of λ_t at the l-th Gibbs iteration to be $[\lambda_{1t}^{(l)},\lambda_{2t}^{(l)}]$, with

$$\lambda_{1t}^{(l)} = a_1 * \max\left(\lambda_t^{(l-1)},\lambda_t^{(0)}\right), \quad \lambda_{2t}^{(l)} = b_1 * \min\left(\lambda_t^{(l-1)},\lambda_t^{(0)}\right), \quad (43)$$

where a_1 and $b_1 > a_1$ are known and fixed, and $\lambda_t^{(l-1)}$ and $\lambda_t^{(0)}$ are, respectively, the estimate of λ_t at the $(l-1)$-th iteration and the initial value.

The following algorithm summarizes the proposed Gibbs sampler for drawing from the joint posterior distribution $f(\theta,\lambda/Y)$ of the Poisson $INSV$ model. For iteration $l = 0,1,...,M$, consider the notation $\lambda^{(l)} = \left(\lambda_1^{(l)},...,\lambda_n^{(l)}\right)'$, $\phi^{(l)} = \left(\phi_0^{(l)},\phi_1^{(l)}\right)'$ and $\sigma^{2(l)}$.

Algorithm 1 (BGG sampler for the P-$INSV$ model)

Step 0 Specify starting values $\lambda^{(0)}$, $\phi^{(0)}$ and $\sigma^{2(0)}$.

Step 1 Repeat for $l = 0,1,...,M-1$,

1.1. Draw $\phi^{(l+1)}$ from $f\left(\phi/Y,\sigma^{2(l)},\lambda^{(l)}\right)$ using (33).

1.2. Draw $\sigma^{2(l+1)}$ from $f\left(\sigma^2/Y,\phi^{(l+1)},\lambda^{(l)}\right)$ using (38).

1.3. Repeat for $t = 1,2,...,n$

Griddy-Gibbs sampler:

Select a grid of m points $\left(\lambda_{ti}^{(l+1)}\right)$: $\lambda_{t1}^{(l+1)} \leq \lambda_{t2}^{(l+1)} \leq ... \leq \lambda_{tm}^{(l+1)}$.

For $1 \leq i \leq m$ calculate $f_{ti}^{(l+1)} = f\left(\lambda_{ti}^{(l+1)}/Y,\theta^{(l)},\lambda_{-\{t\}}^{(l)}\right)$ from (40).

Define the inverse distribution $p\left(\lambda_{ti}^{(l+1)}\right) = \frac{f_{ti}^{(l+1)}}{\sum_{j=1}^{m} f_{tj}^{(l+1)}}$, $1 \leq i \leq m$.

Generate a number u from the uniform $(0,1)$ distribution and transform u using the inverse distribution $p(.)$ to get $\lambda_t^{(l+1)}$, which is considered to be a draw from $f\left(\lambda_t/Y,\theta^{(l+1)},\lambda_{-\{t\}}^{(l)}\right)$.

Step 2 Return the values $\lambda^{(l)}$, $\phi^{(l)}$ and $\sigma^{2(l)}$, $l = 1,...,M$.

3.2 Estimating the negative binomial $INSV$ model

It is important to mention that the proposed Markov chain Monte Carlo (MCMC) methodology of this paper is not model-dependent but exhibits the advantage of being easily adaptable to other conditional distributions that belong to the class of the mixed Poisson $INSV$ models. Since the negative binomial conditional model is often more flexible in representing overdispersion, we shall estimate the negative binomial $INSV$ model.

We consider two estimation approaches. The first one refers to the direct representation of the negative binomial conditional distribution, i.e. $Y_t/\lambda_t, \theta \sim \mathcal{NB}\left(\tau, \frac{\tau}{\tau+\lambda_t}\right)$ with $\tau = \rho^{-2} > 0$ and $\theta = \left(\phi', \sigma^2, \tau\right)'$. The second one, analogously to the Jacquier et al., (2004) approach for the Student-t distribution [see also, Wang et al., (2011); Abanto-Valle et al., (2011)], uses the scale mixture form of the negative binomial distribution through the latent variable $W_t = \lambda_t Z_t$ with a slightly different parametrization.

For highly volatile series, the former approach may become unstable and thus we prefer the latter approach, which gives better results. Another advantage of the scale mixture representation is that it allows us to use a conjugate prior for the latent variable W_t and the kernel of λ_t has a more simplified expression than (44); see below.

3.2.1 The direct representation of the NB-$INSV$ model

For the mixed Poisson $INSV$ 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 $\theta = \left(\phi', \sigma^2, \tau\right)'$. We use again the Gibbs sampler, where the conditional posteriors $f\left(\phi/Y, \sigma^2, \lambda, \tau\right) = f\left(\phi/Y, \sigma^2, \lambda\right)$, $f\left(\sigma^2/Y, \phi, \lambda, \tau\right) = f\left(\sigma^2/Y, \phi, \lambda\right)$ are sampled as in the Poisson case. So it remains to show how to sample from $f\left(\lambda/Y, \theta\right)$ and $f\left(\tau/Y, \phi, \sigma^2, \lambda\right)$. Since $f\left(\tau/Y, \phi, \sigma^2, \lambda\right)$ is not amenable to closed-form integration (Bradlow et al., 2002), we also sample from it using the Griddy-Gibbs sampler, whenever its kernel is defined.

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

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 (39), where now $\theta = \left(\phi', \sigma^2, \tau\right)'$. Using the fact that $Y_t/\theta, \lambda_t \equiv Y_t/\tau, \lambda_t \sim \mathcal{NB}\left(\tau, \frac{\tau}{\lambda_t+\tau}\right)$, $\log\left(\lambda_t\right)/\log\left(\lambda_{t-1}\right), \theta \sim N\left(\phi_0 + \phi_1 \log\left(\lambda_{t-1}\right), \sigma^2\right)$, and $d\log(\lambda_t) = \frac{1}{\lambda_t} d\lambda_t$, formula (39) becomes

$$f\left(\lambda_t/Y, \theta, \lambda_{-\{t\}}\right) \propto \frac{1}{\lambda_t} \frac{\Gamma(Y_t+\tau)}{\Gamma(\tau)} \left(\frac{\tau}{\tau+\lambda_t}\right)^{\tau} \left(\frac{\lambda_t}{\tau+\lambda_t}\right)^{Y_t} \exp\left(-\frac{1}{2\Omega}\left(\log\left(\lambda_t\right) - \mu_t\right)^2\right), \tag{44}$$

where μ_t and Ω are given by (41) and (42), respectively. Then, we can use the Griddy-Gibbs sampler, as in Algorithm 1, to draw from the conditional posterior $f\left(\lambda_t/Y, \theta, \lambda_{-\{t\}}\right)$.

Sampling the dispersion parameter τ

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

$$f\left(\tau/Y, \phi, \sigma^2, \lambda\right) \propto f\left(\tau\right) f\left(Y/\theta, \lambda\right), \tag{45}$$

where $f\left(Y/\theta, \lambda\right)$ is the likelihood function

$$f\left(Y/\theta, \lambda\right) = \prod_{t=1}^{n} \frac{\Gamma(Y_t+\tau)}{\Gamma(\tau)} \left(\frac{\tau}{\tau+\lambda_t}\right)^{\tau} \left(\frac{\lambda_t}{\tau+\lambda_t}\right)^{Y_t}. \tag{46}$$

Since it is difficult to find a conjugate prior for τ, we exploit, as is usually the case, the gamma prior (although in some cases we use the uniform prior; see empirical applications). 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, (45) becomes

$$f\left(\tau/Y, \phi, \sigma^2, \lambda\right) \propto \tau^{a-1} e^{-b\tau} \prod_{t=1}^{n} \frac{\Gamma(Y_t+\tau)}{\Gamma(\tau)} \left(\frac{\tau}{\tau+\lambda_t}\right)^{\tau} \left(\frac{\lambda_t}{\tau+\lambda_t}\right)^{Y_t}. \tag{47}$$

After determining the kernel of $f\left(\tau/Y, \phi, \sigma^2, \lambda\right)$, we use the Griddy-Gibbs sampler, as in the case of λ.

We summarize the proposed Gibbs sampler for drawing from the joint posterior distribution $f(\theta, \lambda/Y)$ of the negative binomial $INSV$ model. For iteration $l = 0, 1, ..., M$, consider the notation $\lambda^{(l)} = \left(\lambda_1^{(l)}, ..., \lambda_n^{(l)}\right)'$, $\phi^{(l)} = \left(\phi_0^{(l)}, \phi_1^{(l)}\right)'$, $\sigma^{2(l)}$ and $\tau^{(l)}$.

Algorithm 2 (BGG sampler for the NB-$INSV$ model-direct representation)
Step 0 Specify starting values $\lambda^{(0)}$, $\phi^{(0)}$, $\sigma^{2(0)}$ and $\tau^{(0)}$.
Step 1 Repeat for $l = 0, 1, ..., M-1$,
1.1. Draw $\phi^{(l+1)}$ from $f\left(\phi/Y, \sigma^{2(l)}, \lambda^{(l)}\right)$ using (33).
1.2. Draw $\sigma^{2(l+1)}$ from $f\left(\sigma^2/Y, \phi^{(l+1)}, \lambda^{(l)}\right)$ (38).
1.3. Repeat for $t = 1, 2, ..., n$
Draw $\lambda_t^{(l)}$ from $f\left(\lambda_t/Y, \phi^{(l+1)}, \sigma^{2(l+1)}, \tau^{(l)}, \lambda_{-\{t\}}^{(l)}\right)$ in (44) using the Griddy-Gibbs sampler as in Step 1.3 of Algorithm 1.
1.4. Draw $\tau^{(l+1)}$ from $f\left(\tau/Y, \phi^{(l+1)}, \sigma^{2(l+1)}, \lambda^{(l+1)}\right)$ in (47) using the Griddy-Gibbs method as in Step 1.3 of Algorithm 1.
Step 2 Return the values $\lambda^{(l)}$, $\phi^{(l)}$, $\sigma^{2(l)}$ and $\tau^{(l)}$, $l = 1, ..., M$.
As for λ_t, the range of τ at the l-th Gibbs iteration is taken to be $[\tau_1^{(l)}, \tau_2^{(l)}]$, where

$$\tau_1^{(l)} = a_2 * \max\left(\tau^{(0)}, \tau^{(l-1)}\right), \quad \tau_2^{(l)} = b_2 * \min\left(\tau^{(0)}, \tau^{(l-1)}\right). \tag{48}$$

3.2.2 The scale mixture representation of the NB-$INSV$ model

It is well-known that the negative binomial distribution of a random variable $Y \sim \mathcal{NB}\left(\tau, \frac{\tau}{\lambda+\tau}\right)$ may be written in the following scale mixture form

$$f(y/\lambda, \tau) = \int_0^{\infty} f(w/\lambda, \tau) f(y/w)\, dw,$$

where $f(w/\lambda, \tau) = \frac{1}{\Gamma(\tau)} \left(\frac{\tau}{\lambda}\right)^{\tau} w^{\tau-1} e^{-\frac{\tau}{\lambda} w}$ $(w > 0)$ and $f(y/w) = e^{-w} \frac{w^y}{y!}$. For a sequence of latent variables $\{W_t, t \in \mathbb{Z}\}$, the conditional distribution $Y_t/\lambda_t \sim \mathcal{NB}\left(\tau, \frac{\tau}{\lambda_t+\tau}\right)$, may then be written hierarchically as follows:

$$Y_t/\lambda_t, W_t, \tau \sim \mathcal{P}(W_t), \tag{49}$$

$$W_t/\lambda_t, \tau \sim \mathcal{G}\left(\tau, \frac{\tau}{\lambda_t}\right). \tag{50}$$

So we have to estimate $\theta = \left(\phi, \sigma^2, \tau\right)'$ and the latent variables $(W_t)_{1\leq t\leq n}$ and $(\lambda_t)_{1\leq t\leq n}$. Comparing the mixture representation (49)-(50) with (4), it is clear that $W_t = \lambda_t Z_t$ and unlike $\{Z_t, t \in \mathbb{Z}\}$, the sequence $\{W_t, t \in \mathbb{Z}\}$ is not *i.i.d.*

The conditional posteriors for the components of θ are given as in Algorithm 2. From the conjugate prior (50), the conditional posterior of W_t $(1 \leq t \leq n)$ is

$$W_t/Y, \lambda_t, \tau \sim \mathcal{G}\left(\tau + \sum_{t=1}^{n} Y_t, \tfrac{\tau}{\lambda_t} + n\right), \tag{51}$$

which is another advantage of the scale mixture representation.

Now sampling from $f\left(\lambda_t/Y, \theta, \underline{\lambda}_{-\{t\}}, W\right)$, where $W = (W_1, ..., W_n)'$, is done as in (40) but with a slight modification. In particular, similarly to (39), we have

$$\begin{aligned} f\left(\lambda_t/Y, \theta, \underline{\lambda}_{-\{t\}}, W\right) &= f\left(\lambda_t/Y_t, \theta, \lambda_{t-1}, \lambda_{t+1}, W_t\right) \\ &\propto f\left(\lambda_t/\lambda_{t-1}, \theta\right) f\left(\lambda_{t+1}/\lambda_t, \theta\right) f\left(W_t/\lambda_t, \theta\right). \end{aligned}$$

Using similar devices as for (40) we therefore get

$$f\left(\lambda_t/Y, \theta, \underline{\lambda}_{-\{t\}}, W_t\right) \propto \exp\left(-\tfrac{\tau W_t}{\lambda_t} - (\tau+1)\log\left(\lambda_t\right) - \tfrac{1}{2\Omega}\left(\log\left(\lambda_t\right) - \mu_t\right)^2\right), \quad 1 < t < n, \tag{52}$$

where μ_t and Ω are given by (41) and (42), respectively. The following scheme recapitulates the Griddy-Gibbs sampler based on the scale mixture representation of the negative binomial $INSV$ model.

Algorithm 3 (BGG sampler for the NB-$INSV$ model-scale mixture representation)

Step 0 Specify starting values $\lambda^{(0)}$, $W^{(0)}$, $\phi^{(0)}$, $\sigma^{2(0)}$ and $\tau^{(0)}$.

Step 1 Repeat for $l = 0, 1, ..., M-1$,

1.1. Draw $\phi^{(l+1)}$ from $f\left(\phi/Y, \sigma^{2(l)}, \lambda^{(l)}\right)$ using (33).

1.2. Draw $\sigma^{2(l+1)}$ from $f\left(\sigma^2/Y, \phi^{(l+1)}, \lambda^{(l)}\right)$ using (38).

1.3. Repeat for $t = 1, 2, ..., n$
Draw $W_t^{(l+1)}$ from $f\left(W_t/Y, \lambda_t^{(l)}, \tau^{(l)}\right)$ using (51).

1.4. Repeat for $t = 1, 2, ..., n$
Draw $\lambda_t^{(l+1)}$ from $f\left(\lambda_t/Y, \phi^{(l+1)}, \sigma^{2(l+1)}, W_t^{(l+1)}, \tau^{(l)}, \lambda_{-\{t\}}^{(l+1)}\right)$ in (52)
using the Griddy-Gibbs sampler.

1.5. Draw $\tau^{(l+1)}$ from $f\left(\tau/Y, \phi^{(l+1)}, \sigma^{2(l+1)}, \lambda^{(l+1)}\right)$ in (47)
using the Griddy-Gibbs sampler.

Step 2 Return the values $\lambda^{(l)}$, $W^{(l)}$, $\phi^{(l)}$, $\sigma^{2(l)}$ and $\tau^{(l)}$, $l = 1, ..., M$.

Another advantage, rather statistical, of the mixture representation of the mixed Poisson $INSV$ model is that we can use the scale mixture form (Algorithm 3) for any conditional distribution of Y_t/λ_t that belongs to the class of Poisson mixtures, not only for the negative binomial law. It suffices to sample from the given distribution of the mixing variable W_t (Step 1.3 of Algorithm 3), which, in the general case, is not necessarily gamma distributed.

3.3 $MCMC$ diagnostics

Given our single-move framework, it is of principal interest to discuss the numerical properties of the proposed BGG methods. Despite its ease of implementation, the main drawback of the single-move approach (see, for example, Kim et al., (1998)) is that the posterior draws are often highly correlated, causing slow mixing and slow convergence properties. Note that, due to the non-MEM form of the mixed Poisson $INSV$ model, a multi-move approach does not seem possible to follow.

Among several $MCMC$ diagnostic measures, we first consider the Relative Numerical Inefficiency (RNI) (see, for example, Geweke, 1989; Geyer, 1992; Tsiakas, 2006; Aknouche, 2017), which is given by

$$RNI = 1 + 2\sum_{h=1}^{B} K\left(\tfrac{h}{B}\right)\widehat{\rho}_h^G,$$

where $B = 500$ is the bandwidth, $K(.)$ is the Parzen kernel (see, for example, Priestley (1981), Ch. 6) and $\widehat{\rho}_h^G$ is the sample autocorrelation at lag h of the BGG parameter draws. The RNI measures the inefficiency stemming from the serial correlation of the BGG draws.

In addition, we use the Numerical Standard Error (NSE), (see, for example, Geweke, 1989; Tsiakas, 2006; Aknouche, 2017), which is the square root of the estimated asymptotic variance of the $MCMC$ estimator. In fact, the NSE is given by

$$NSE = \sqrt{\tfrac{1}{M}\left(\widehat{\gamma}_0^G + 2\sum_{h=1}^{B} K\left(\tfrac{h}{B}\right)\widehat{\gamma}_h^G\right)},$$

where $\widehat{\gamma}_h^G$ is the sample autocovariance at lag h of the BGG parameter draws and M is the number of draws.

We also computed the Convergence Diagnostics (CD) statistics of Geweke (1992) in order to monitor convergence. The CD statistics compares the mean $\bar{x}_0$ of the first part $l = 1, ..., M_1$ of the chain to the mean $\bar{x}_1$ of the last part $l = M_2 + 1, ..., M$ of the chain, after discarding the middle part and is calculated as

$$CD = \frac{\bar{x}_0 - \bar{x}_1}{\sqrt{NSE_0^2/ + NSE_1^2}}.$$

The CD statistics converges in distribution to the standard normal and if a sufficiently large number of draws has been obtained, it attains low values.

3.4 Model selection via the Deviance Information Criterion

We also carry out model selection using the *Deviance Information Criterion* (DIC, Spiegelhalter et al., 2002); see also Berg et al., (2004). The DIC value can be obtained from the $MCMC$ output, without extra calculations. In the context of the mixed Poisson $INSV$ model, the DIC formula is defined as

$$DIC = -4E_{\theta,\lambda/Y}\left(\log\left(f\left(Y/\theta,\lambda\right)\right)\right) + 2\log\left(f\left(Y/\overline{\theta},\overline{\lambda}\right)\right),$$

where $f(Y/\theta,\lambda)$ is the (conditional) likelihood of the mixed Poisson INSV model and $\left(\overline{\theta},\overline{\lambda}\right) = E\left((\theta,\lambda)/Y\right)$ is the posterior mean of (θ,λ). The expectation $E_{\theta,\lambda/Y}\left(\log\left(f\left(Y/\theta,\lambda\right)\right)\right)$ can be estimated from the Griddy-Gibbs draws by averaging the conditional log-likelihood, $\log f(Y/\theta,\lambda)$, over the posterior draws of (θ,λ). Furthermore, the joint posterior mean estimate of $(\overline{\theta},\overline{\lambda})$ can be approx-

imated by the mean of the posterior draws of $(\theta^{(l)}, \lambda^{(l)})$. Since

$$\log f\left(Y/\theta,\lambda\right)=\begin{cases}\sum_{t=1}^{n}\left(-\lambda_t+Y_t\log\left(\lambda_t\right)\right)\\ \sum_{t=1}^{n}\log\left(\frac{\Gamma(Y_t+\tau)}{\Gamma(\tau)}\right)+\tau\log\left(\frac{\tau}{\tau+\lambda_t}\right)+Y_t\log\left(\frac{\lambda_t}{\tau+\lambda_t}\right),\end{cases}$$

the DIC is estimated by

$$\begin{cases}\frac{2}{M}\sum\limits_{l=l_0}^{l_0+M}\sum\limits_{t=1}^{n}\left(-\lambda_t^{(l)}+Y_t\log\left(\lambda_t^{(l)}\right)\right)-\sum_{t=1}^{n}\left(-\overline{\lambda}_t+Y_t\log\left(\overline{\lambda}_t\right)\right), & \text{if } \rho^2=0\\ \frac{2}{M}\sum\limits_{l=l_0}^{l_0+M}\sum\limits_{t=1}^{n}\left(\log\left(\frac{\Gamma\left(Y_t+\tau^{(l)}\right)}{\Gamma\left(\tau^{(l)}\right)}\right)+\tau^{(l)}\log\left(\frac{\tau^{(l)}}{\tau^{(l)}+\lambda_t^{(l)}}\right)+Y_t\log\left(\frac{\lambda_t^{(l)}}{\tau^{(l)}+\lambda_t^{(l)}}\right)\right)-\\ \sum\limits_{t=1}^{n}\left(\log\left(\frac{\Gamma(Y_t+\overline{\tau})}{\Gamma(\overline{\tau})}\right)+\overline{\tau}\log\left(\frac{\overline{\tau}}{\overline{\tau}+\overline{\lambda}_t}\right)+Y_t\log\left(\frac{\overline{\lambda}_t}{\overline{\tau}+\overline{\lambda}_t}\right)\right), & \text{if } \rho^2>0\end{cases}$$

where $\lambda_t^{(l)}$and $\tau^{(l)}$ denote the l-th BGG draw of λ_t and τ from $f\left(\lambda_t/Y_t,\theta\right)$ and $f\left(\tau/Y,\phi,\sigma^2,\lambda\right)$, respectively, M is the number of draws, l_0 is the burn-in size and $\overline{\lambda}_t:=E\left(\lambda_t/Y\right)$ and $\overline{\tau}=:E\left(\tau/Y\right)$ are estimated by $\frac{1}{M}\sum_{l=l_0}^{l_0+M}\lambda_t^{(l)}$ $(1\leq t\leq n)$ and $\frac{1}{M}\sum_{l=l_0}^{l_0+M}\tau^{(l)}$, respectively.

A model is preferred if it has the smallest DIC value.

4 Simulation study

In this section, we assess the performance of the proposed Bayesian methodology on simulated mixed Poisson $INSV$ series with Gaussian innovations for the log-intensity equation. In particular, we consider two cases of the mixed Poisson $INSV$ model; the P-$INSV$ model and the NB-$INSV$ model. For the NB-$INSV$ model we implement the direct approach of Algorithm 2 and the scale mixture approach of Algorithm 3.

For each model, three Monte Carlo experiments (MCE) are conducted, each corresponding to a different set of real values. In each case, we perform 3000 Monte Carlo replications with sample size $n=1000$ for which we compute the BGG estimates, their means ($Mean$), their standard deviations (Std) and their Root Mean Square Errors ($RMSE$) over the 3000 replications. The $RMSE$ of an estimate $\widehat{\theta}$ of θ is calculated from the formula $RMSE=\sqrt{bias^2+Std^2}$, where $bias$ is the sample mean of $\widehat{\theta}-\theta$ over the 3000 replications.

In implementing the BGG samplers, the initial intensity $\lambda^{(0)}$ is taken to be the intensity generated by the mixed Poisson $INGARCH\left(1,1\right)$ model that we fit to the generated data points. Specifically, the $INGARCH\left(1,1\right)$ model in question is given by

$$\begin{cases}X=N_t\left(Z_t\lambda_t^I\right)\\ \lambda_t^I=\lambda_t^I\left(\mu\right)=\omega+\alpha X_{t-1}+\beta\lambda_{t-1}^I\end{cases},\qquad t\geq 1,$$

where λ_t^I denotes the intensity parameter, Z_t is degenerate in the Poisson case and is gamma distributed in the negative binomial case, $\mu=(\omega,\alpha,\beta)'$ and the starting value X_0 in the $INGARCH$ equation is set equal to $X_0=\lambda_0^I=\frac{\omega}{1-(\alpha+\beta)}$. The parameter μ is estimated using the Poisson quasi-maximum likelihood estimate ($P-QMLE$) or the two-stage negative binomial QMLE ($2SNB$-$QMLE$, see Aknouche et al., 2018), giving $\widehat{\mu}$. Hence, the initial intensity $\lambda^{(0)}$ is set to

$$\lambda^{(0)}=\lambda^I\left(\widehat{\mu}\right).$$

The starting value $\theta^{(0)}$ in the Gibbs samplers is taken to be the ordinary least-squares estimate of $\theta = \left(\phi_0, \phi_1, \sigma^2\right)'$, based on the series $\left(\log\left(\lambda_t^{(0)}\right)\right)_{1\leq t\leq n}$. For θ, we use the following conjugate priors

$$\phi \sim N\left((0,0)', diag\left(0.002, 0.01\right)\right), \quad \tfrac{5\times 0.2}{\sigma^2} \sim \chi_5^2. \tag{53}$$

These priors are informative, but reasonably flat. In the negative binomial case, the dispersion parameter τ is initialized, according to Aknouche et al., (2018), as follows:

$$\tau^{(0)} = \tfrac{\overline{Y}^2}{S^2 - \overline{Y}}, \tag{54}$$

where $\overline{Y}$ and S^2 are, respectively, the sample mean and sample variance of the generated series $Y_1, ..., Y_n$. The prior distribution of τ is gamma with hyperparameters

$$\tau \sim \mathcal{G}\left(0.1, 5\right). \tag{55}$$

Concerning the Griddy-Gibbs step, λ_t and τ are generated using 500 grid points and the ranges of λ_t and τ at the l-th Gibbs iteration are given in (43) and (48), respectively. Finally, the Gibbs sampler is run for 3000 iterations from which we discarded the first 300 cycles.

The simulation results for the $P\text{-}INSV$ and $NB\text{-}INSV$ models are summarized in Tables 1-3. It can be observed that in both models, the true parameters are well estimated with quite small $RMSEs$, implying efficiency in updating the estimated parameters. For the negative binomial $INSV$, it may be concluded that the scale mixture method (based on Algorithm 3) gives, in general, estimated parameters that are closer to their true values than the estimated parameters from the direct approach (based on Algorithm 2). Overall, the BGG samplers performed remarkably well.

5 Empirical examples

5.1 Transaction data

To illustrate the Bayesian methodology of this paper, we apply the $P\text{-}INSV$ and $NB\text{-}INSV$ models to a transaction data set that has been widely used by the relevant empirical literature (Fokianos et al, 2009; Davis and Liu, 2016; Christou and Fokianos, 2014; Aknouche et al, 2018). The time series in question consisting of $n = 460$ observations records the number of transactions per minute for the Ericsson B stock from 09:35 AM to 17:25 PM on July 05, 2002; see Figure 1a.

The transaction series with a sample mean of $\overline{Y}$=9.8239 and a sample variance of $S^2 = 23.7532$ is strongly overdispersed. It has a large frequency of zeros, an asymmetric marginal distribution and is characterized by a locally constant behavior; see Figure 1b.

Using the $2SNB\text{-}QMLE$ method (Aknouche et al, 2018), we first fitted a negative binomial $INGARCH(1,1)$ model to the data set in question, giving

$$\begin{aligned} &X_t/\mathcal{F}_{t-1} \sim \mathcal{NB}\left(\widehat{\tau}, \tfrac{\widehat{\tau}}{\widehat{\tau}+\widehat{\lambda}_t^I}\right), \ \widehat{\tau} = 7.8199 \\ &\begin{cases} \widehat{\lambda}_t^I = 0.7996 + 0.7928 Y_{t-1} + 0.1249\widehat{\lambda}_{t-1}^I, \quad 2 \leq t \leq 460 \\ \widehat{\lambda}_1^I = \overline{Y} = 9.8239. \end{cases} \end{aligned} \tag{56}$$

For the estimation of the $P\text{-}INSV$ parameters we applied Algorithm 1, while for the estimation of the $NB\text{-}INSV$ parameters we applied Algorithm 3, which, on average, delivered a better DIC value

than Algorithm 2. Regarding the Griddy-Gibbs sampling of λ_t and τ, we used grids of $g = 100$ points, with their ranges given by (43) and (48), respectively.

The number of Gibbs iterations was set equal to $M = 3500$ with a burn-in period of 500 updates. Furthermore, the initial values in the Gibbs sampler and the hyperparameters of the prior distributions for the $P\text{-}INSV$ and $NB\text{-}INSV$ models are similar to those used in the simulation study.

In particular, the initial intensity $\lambda^{(0)} = \widehat{\lambda}^I$ is taken to be the intensity generated by the negative binomial $INGARCH(1,1)$ model in (57). As an initial parameter vector $\psi^{(0)} = \left(\phi^{(0)\prime}, \sigma^{2(0)}\right)'$ we use the ordinary least-squares estimate of $\psi = \left(\phi', \sigma^2\right)'$, based on the series $\left(\log\left(\lambda_t^{(0)}\right)\right)_{1\leq t\leq 460}$. Finally, $\tau^{(0)}$ is calculated using $\tau^{(0)} = \frac{\overline{Y}^2}{S^2-\overline{Y}}$.

On the other hand, the prior distributions of ϕ and σ^2 are given, respectively, by

$$\phi \sim N\left((0,0)', diag\,(0.01, 0.01)\right), \quad \tfrac{5\times 0.2}{\sigma^2} \sim \chi^2_5,$$

while for τ we assume the uniform prior defined on the nonnegative region of the real line.

To estimate the standard error of the DIC we simply replicated its calculation $G = 100$ times and estimated the variance of DIC $[var(DIC)]$ by its sample variance. The estimated DIC*'s* across the two models and their corresponding standard deviations are reported in Table 4.

According to the DIC results, the best model is the $NB\text{-}INSV$, as it has the lowest DIC value (5835.9655). Therefore, we retain the $NB\text{-}INSV$ model in our analysis and report the posterior means and posterior standard deviations (Std) of its parameters in Table 5. For comparison purposes, Table 6 reports the same information concerning the $P\text{-}INSV$ parameters.

All the parameters are statistically significant in both models. In Figure 2, the posterior paths (middle panel) are stable and the posterior autocorrelations (top panel) decay rapidly, suggesting that the proposed MCMC algorithmic scheme for the $NB\text{-}INSV$ model is efficient. This is also confirmed by the reported measures of mixing and convergence (RNI, NSE, CD) in Table 5.

Furthermore, from the posterior histograms (middle panel of Figure 2) of the $NB\text{-}INSV$ parameters, we observe that $\widehat{\phi}_1$ lies in the stability domain, defined by expression (7), while $\widehat{\tau}$ is far away from zero, thus removing any doubt about the validity of the Poisson hypothesis.

Using expressions (14) and (19), the estimates of the mean $E\,(Y_t)$ and variance $var(Y_t)$ for the $NB\text{-}INSV$ model are, respectively, 8.0199 and 26.8621. These values are very close to the sample mean and sample variance of the series. This shows that the estimated model allows a good (two first) moment adjustments. For the $P\text{-}INSV$ case, we have $\widehat{E}\,(Y_t) = 9.1022$ and $\widehat{var}(Y_t)$= 20.3349.

To examine the adequacy of the $NB\text{-}INSV$ model, we focused on its residuals, which are of two types. First, the Pearson conditional mean residual $(\widehat{\epsilon}_t)_{1\leq t\leq n}$ (Y-residuals in short) is defined by

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

where $\widehat{\lambda}_t$ is the BGG 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, the residuals $(\widehat{e}_t)_{1\leq t\leq n}$ (henceforth e-residuals) of the log-intensity equation are

$$\widehat{e}_t = \widehat{\sigma}^{-1}\left(\log\left(\widehat{\lambda}_t\right) - \widehat{\phi}_0 - \widehat{\phi}_1 \log\left(\widehat{\lambda}_{t-1}\right)\right), \quad 1 \leq t \leq n,$$

where $\widehat{\phi}_0, \widehat{\phi}_1$ and $\widehat{\sigma}^2$ are the BGG estimates of ϕ_0, ϕ_1 and σ^2, respectively.

As shown in Figures 3a (simple autocorrelations) and 3b (partial autocorrelations), the e-residuals

are uncorrelated, and according to Figures 3c (QQ plot) and 3d (kernel density), the normality assumption is acceptable. The autocorrelations of the squared and absolute e-residuals have the form of a white noise, which reinforces the independence assumption on the e-innovations; see Figure 4.

Concerning the analysis of Y-residuals, it can be seen from the simple and partial autocorrelation functions (Figures 5a and 5b, respectively) that there is no significant evidence of correlation within the residuals. However, a visual inspection of Figures 5c (QQ-plot) and 5d (kernel density) reveals that the normality assumption of the residuals is not tenable. It is natural the Y-residuals to not have a Gaussian shape because they are residuals from a discrete distribution.

To have a more meaningful conclusion about the BGG fitting we analysed the randomized quantile 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 residuals are especially useful to achieve continuous residuals, when the series is discrete, as in our case. In the $NB\text{-}INSV$ 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 $\widehat{\theta} = \left(\widehat{\phi}_0, \widehat{\phi}_1, \widehat{\sigma}^2, \widehat{\tau}\right)'$.

Figure 6 shows the simple and partial autocorrelations of the randomized residuals (panels (a) and (b), respectively), the QQ-plot of these residuals versus the standard normal distribution (panel (c)) and the kernel density (panel (d)). It can be observed that the residuals appear to be roughly a Gaussian white noise as expected, meaning that the distribution of the estimated model fits well to the negative binomial distribution. Similar conclusions can be extracted from the autocorrelation plots of the squared and absolute values of the randomized quantile residuals; see Figure 7.

Finally, Figure 8 displays the evolution of the estimated latent $\widehat{W_t}$, while Figure 9 shows the real time path of transaction data along with the estimated conditional mean $\widehat{\lambda}_t$ and the estimated conditional variance, given by $\widehat{\lambda}_t\left(1 + \frac{1}{\widehat{\tau}}\widehat{\lambda}_t\right)$. 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.

In summary, regarding the stability of the $NB\text{-}INSV$ model, the significance of its coefficients, the good mixing of the Gibbs draws and the residual analysis, it can be concluded that the estimated $NB\text{-}INSV$ model is valid for representing the polio data.

5.2 Polio data

We also applied the proposed Bayesian methodology to the Polio data set, repeating the same empirical analysis that we conducted for the transaction data. The Polio data set, which consists of $n = 168$ observations, refers to the monthly number of poliomyelitis cases in the United States from January 1971 to December 1983. The time series in question was originally modelled by Zeger (1988) and was later used by many authors (Zeger and Qaqish, 1988; Davis et al., 1999; Benjamin et al., 2003; Davis and Rodriguez-Yam, 2005; Davis and Wu, 2009; Zhu, 2011; Aknouche et al., 2018, among others).

The Polio time series (Figure 10a) with a sample mean of 1.3333 and a sample variance of 3.5050 is clearly overdispersed. The histogram of this data set is given in Figure 10b.

Following Aknouche et al., (2018), we fitted a negative binomial $INGARCH(1,1)$ model to the

Polio data, which was estimated using the Poisson quasi-maximum likelihood estimation method (P-$QMLE$), getting

$$X_t/\mathcal{F}_{t-1} \sim \mathcal{NB}\left(\widehat{\tau}, \frac{\widehat{\tau}}{\widehat{\tau}+\widehat{\lambda}_t^I}\right),\ \widehat{\tau} = 2.6023$$
$$\begin{cases} \widehat{\lambda}_t^I = 0.5808 + 0.1986 Y_{t-1} + 0.7445 \widehat{\lambda}_{t-1}^I, \quad 2 \leq t \leq 168 \\ \widehat{\lambda}_1^I = \overline{Y} = 1.3333 \end{cases}$$

The prior distributions of ϕ, σ^2 and τ are given, respectively, by

$$\phi \sim N\left((0.27, 0.60)', diag\,(0.001, 0.001)\right),\ \ \tfrac{5\times 0.2}{\sigma^2} \sim \chi_5^2,\ \ \tau \sim \mathcal{G}\,(0.1, 5)\,.$$

In Table 7, the model comparison results, based on the DIC criterion (Spiegelhalter et al, 2002) show that the NB-$INSV$ is preferred to the P-$INSV$. The estimation results for the NB-$INSV$ and P-$INSV$ models appear in Tables 8 and 9, respectively.

All the parameters are significant in both models. Comparing the two data sets, we notice that the estimated persistence, captured by ϕ_1, is smaller in the Polio data than in the transaction data for the NB-$INSV$ model. Furthermore, from the reported RNI, NSE and CD values (Table 8), as well as the plotted posterior paths and posterior autocorrelations (Figure 11), there are no mixing or convergence problems with the generated Markov chains of the posterior sampler for the dominant model.

Using expressions (14) and (19) of the main paper, the estimates of the mean $E\,(Y_t)$ and variance $var(Y_t)$ for the NB-$INSV$ model are, respectively, 1.4008 and 3.4966. These values are closer to the sample mean and sample variance than the corresponding estimated values obtained by Zhu (2011) and Aknouche et al., (2018). This shows that the NB-$INSV$ model allows for a good (two first) moment adjustments. For the P-$INSV$ case, $\widehat{E}\,(Y_t) = 1.6690$ and $\widehat{var}(Y_t) = 2.4475$.

The e-residual analysis in Figure 12 reveals uncorrelatedness (see also Figure 13) and non-normality. The non-normality is perhaps due to the small sample of the series or to lack of covariates (seasonal effects, etc), as suggested by Zegger (1998) and Davis and Rodriguez-Yam (2005). From the Y-residual analysis (Figure 14) and the randomized quantile residual analysis (Figures 15 and 16), the residuals are uncorrelated and resemble a Gaussian white noise process.

In addition, Figure 17 displays the path of the estimated latent $\widehat{W_t}$ variable. Figure 18 confirms that the data in question suffer for conditional overdispersion.

6 Conclusions

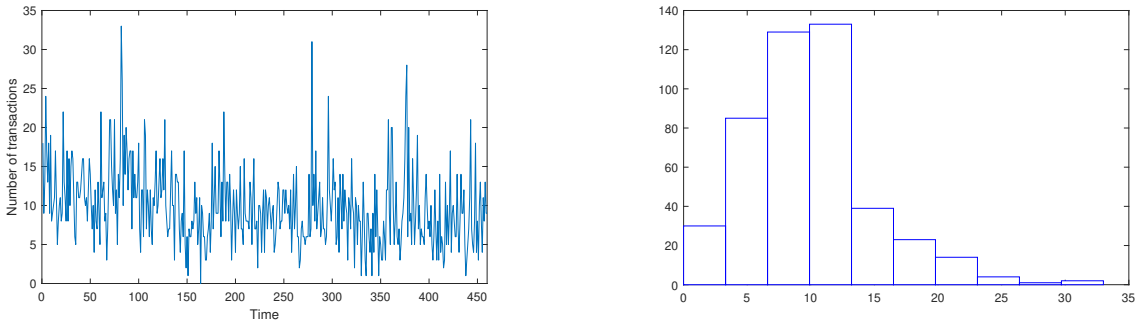
We proposed an integer-valued stochastic volatility ($INSV$) model that parallels the stochastic volatility model for real-valued time series. The proposed specification is a discrete-valued parameter-driven model, depending on a latent time-varying intensity parameter, the logarithm of which follows a drifted first-order autoregression. We focused on a rich class of Poisson mixture distributions that form a particular $INSV$-type model, the mixed Poisson $INSV$ model.

Unlike the standard stochastic volatility model, the mixed Poisson $INSV$ 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 two advantages. The first one, which is probabilistic, allows us to write the model as a non-linear stochastic difference equation with $i.i.d$

innovations, simplifying the study of its probabilistic properties. The second one, which is statistical, allows us to use the scale mixture representation of the conditional distribution of the model, a fact that simplifies the stability of MCMC computation, in particular in the presence of highly volatile (or overdispersed) series.

Under the Gaussianity assumption for the innovation of the log-intensity equation, we considered the Bayesian Griddy-Gibbs sampler in order to estimate the parameters of the mixed Poisson $INSV$ model for two particular conditional distributions; the Poisson and the negative binomial. In the negative binomial case we developed two estimation approaches. The first one is based on the direct representation of the negative binomial distribution, while the second one is an improved estimation method, based on the scale mixture representation of the negative binomial distribution. The proposed Bayesian methodology is not model-dependent and may be adapted to other important discrete distributions.

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



(a) Time series. (b) Histogram.

Figure 1: Transactions per minute for the Ericsson B stock on July 05, 2002.

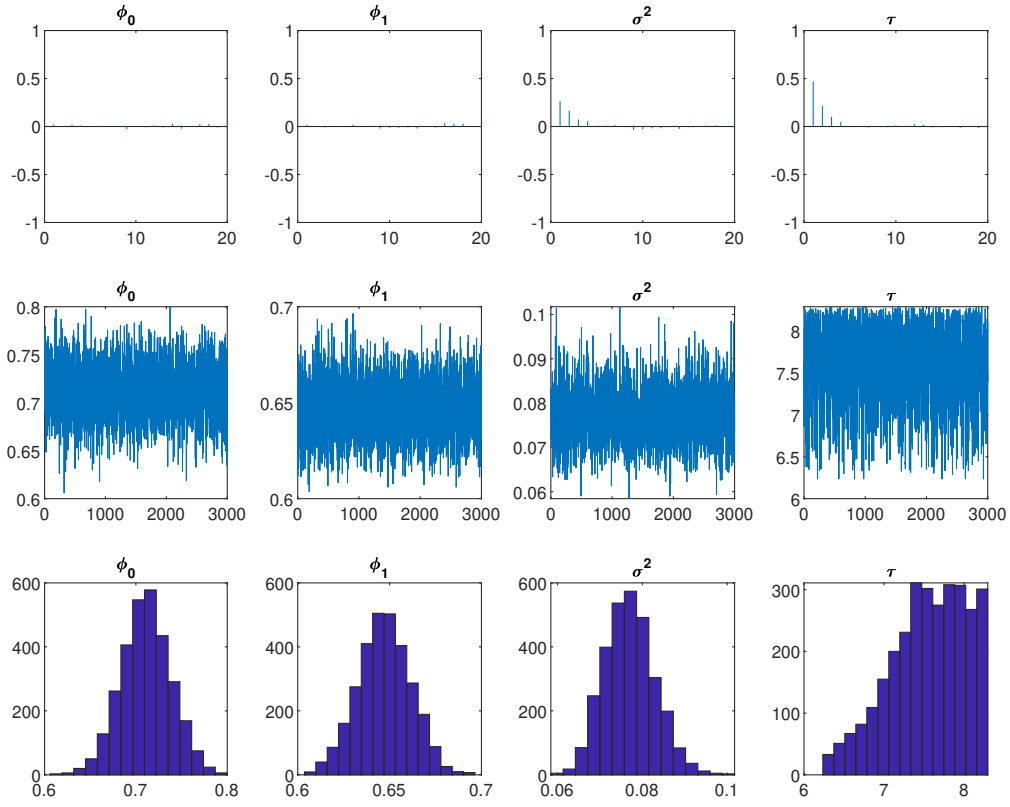
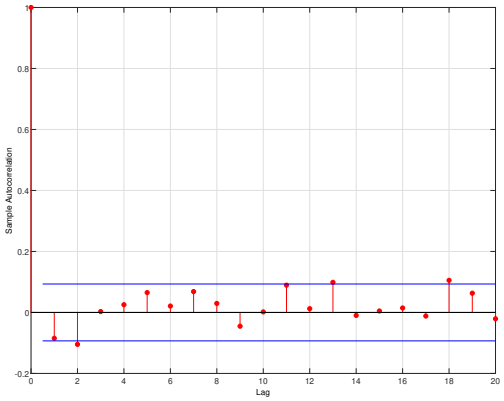
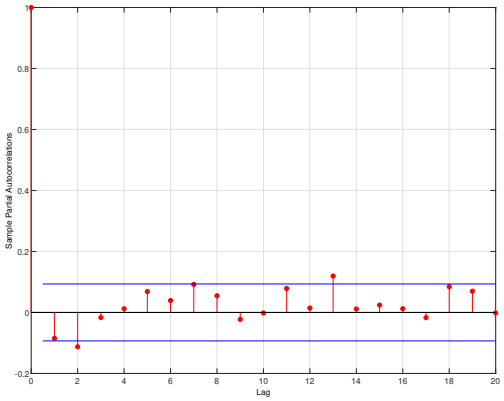


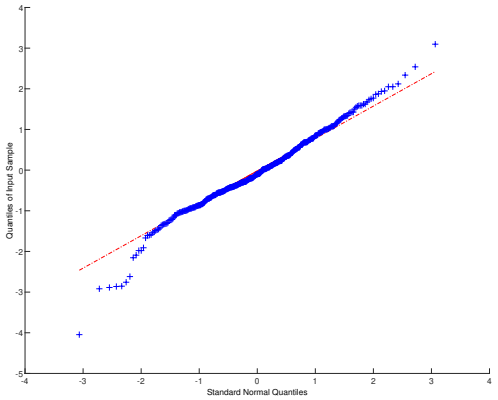
Figure 2: Empirical results (Transaction data). Posterior autocorrelations (top), posterior paths (middle) and posterior histograms (bottom) for the parameters of the $NB\text{-}INSV$ model.



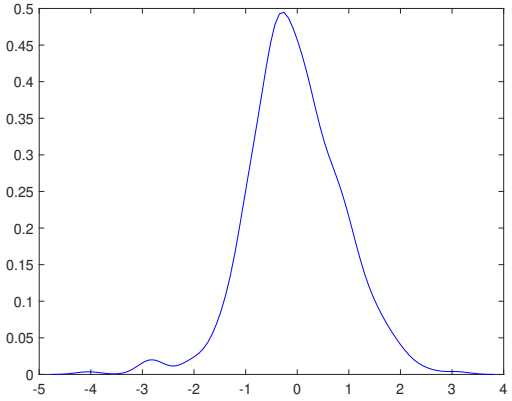
(a) Sample autocorrelations of the e-residuals.



(b) Partial autocorrelations of the e-residuals.

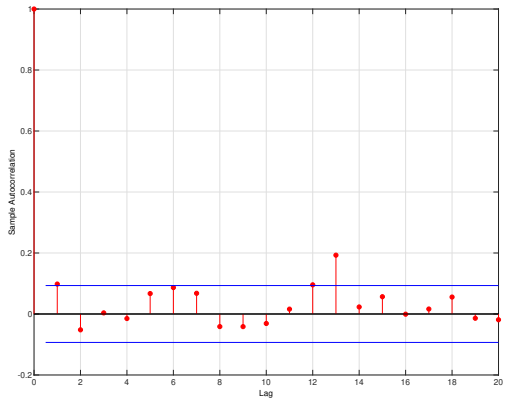


(c) QQ-plot of the e-residuals versus the standard normal distribution.

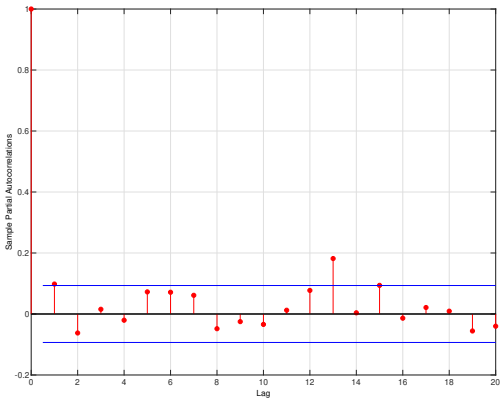


(d) Kernel density of the e-residuals.

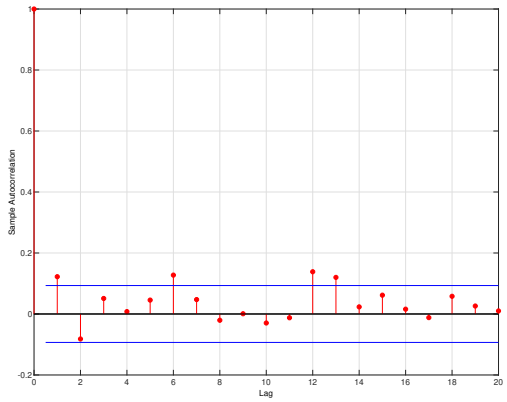
Figure 3: Empirical results (Transaction data). e-residual analysis for the $NB\text{-}INSV$ model.



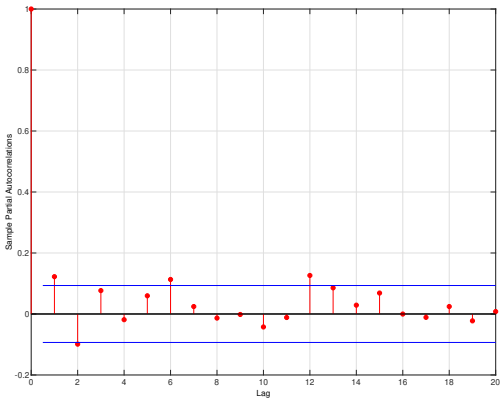
(a) Sample autocorrelations of the squared e-residuals.



(b) Partial autocorrelations of the squared e-residuals.

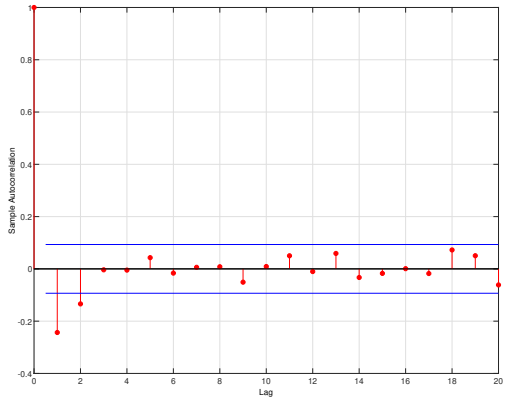


(c) Sample autocorrelations of the absolute e-residuals.

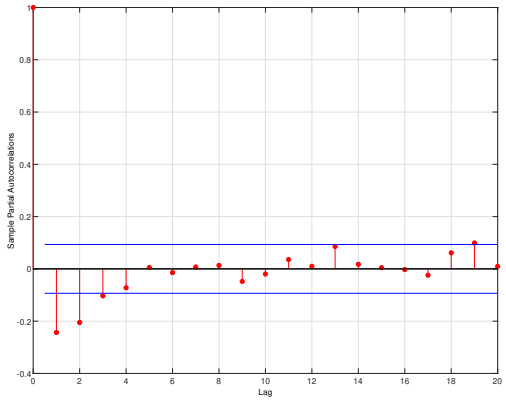


(d) Partial autocorrelations of the absolute e-residuals.

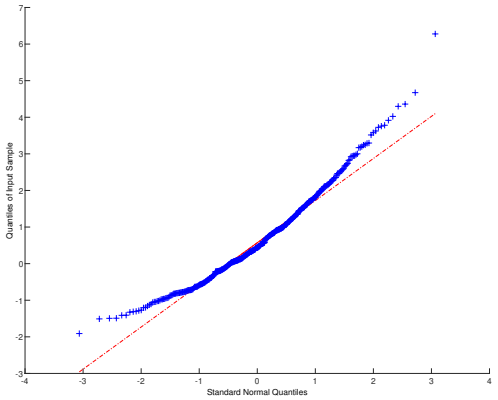
Figure 4: Empirical results (Transaction data). Squared and absolute e-residual analysis for the $NB\text{-}INSV$ model.



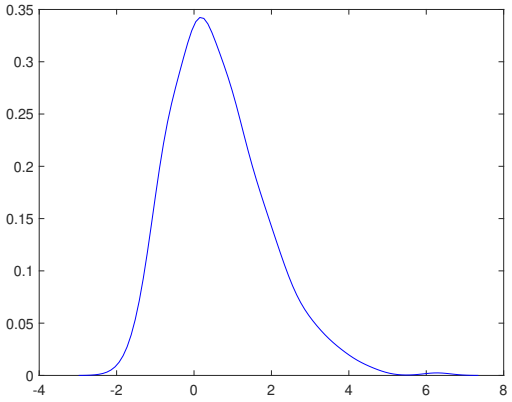
(a) Sample autocorrelations of the Pearson residuals.



(b) Partial autocorrelations of the Pearson residuals.

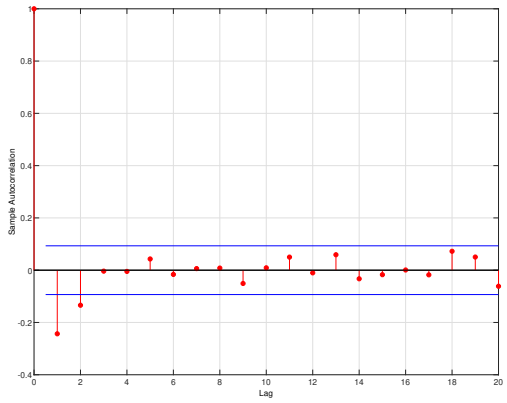


(c) QQ-plot of the Pearson residuals versus the standard normal distribution.

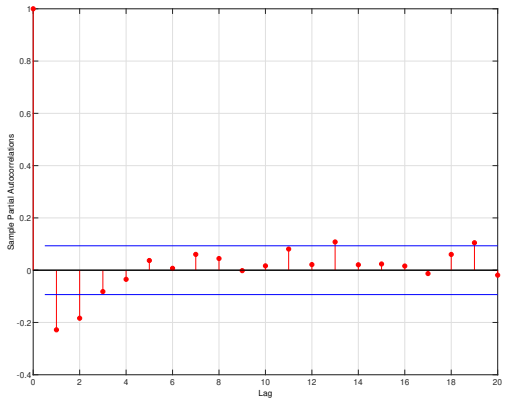


(d) Kernel density of the Pearson residuals.

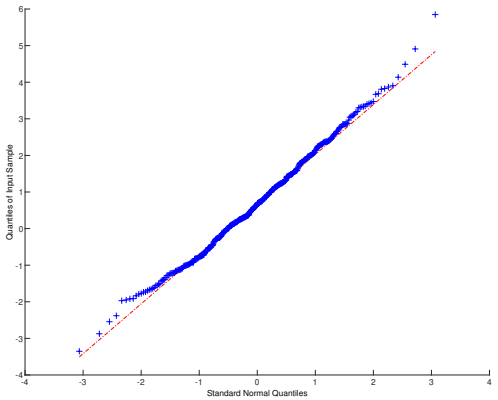
Figure 5: Empirical results (Transaction data). Pearson analysis for the $NB\text{-}INSV$ model.



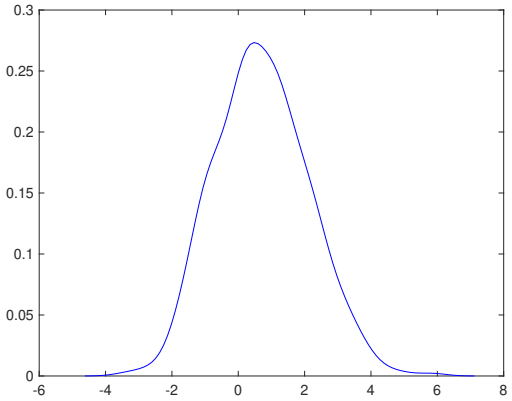
(a) Sample autocorrelations of the randomized quantile residuals.



(b) Partial autocorrelations of the randomized quantile residuals.

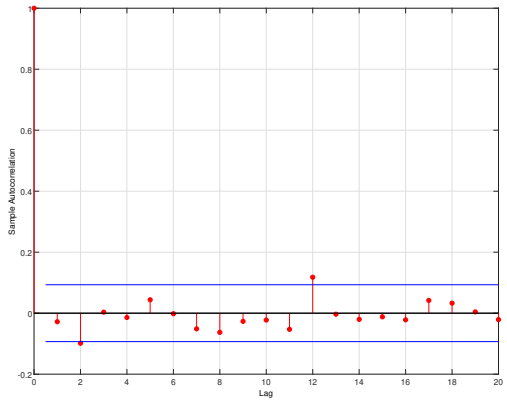


(c) QQ-plot of the randomized quantile residuals versus the standard normal distribution.

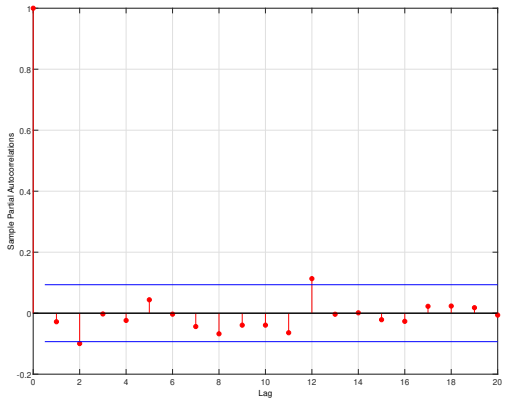


(d) Density of the randomized quantile residuals.

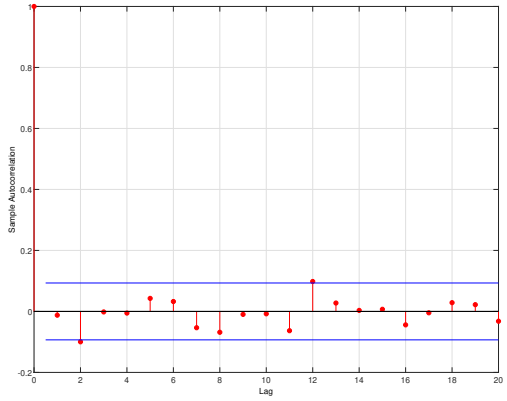
Figure 6: Empirical results (Transaction data). Randomized quantile residual analysis for the NB-$INSV$ model.



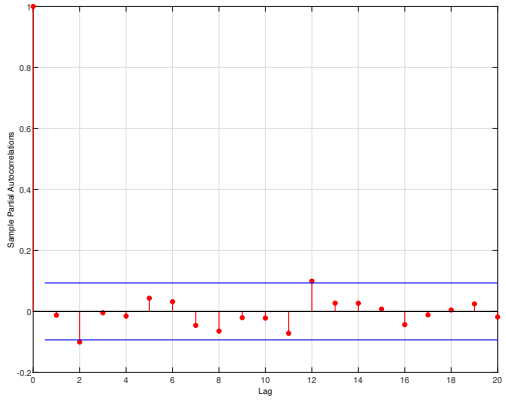
(a) Sample autocorrelations of the squared randomized quantile residuals.



(b) Partial autocorrelations of the squared randomized quantile residuals.



(c) Sample autocorrelations of the absolute randomized quantile residuals.



(d) Partial autocorrelations of the absolute randomized quantile residuals.

Figure 7: Empirical results (Transaction data). Squared and absolute randomized quantile residual analysis for the $NB\text{-}INSV$ model.

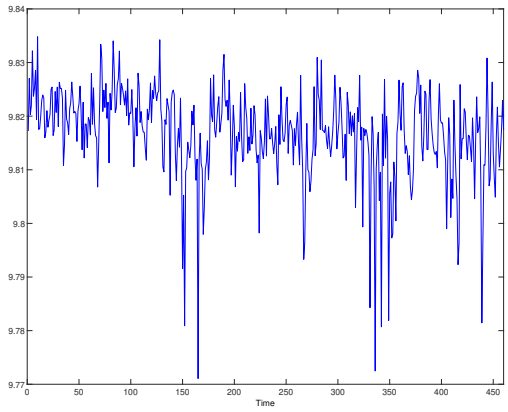


Figure 8: Empirical results (Transaction data). Estimated latent series $\widehat{W_t}$ for the $NB\text{-}INSV$ model.

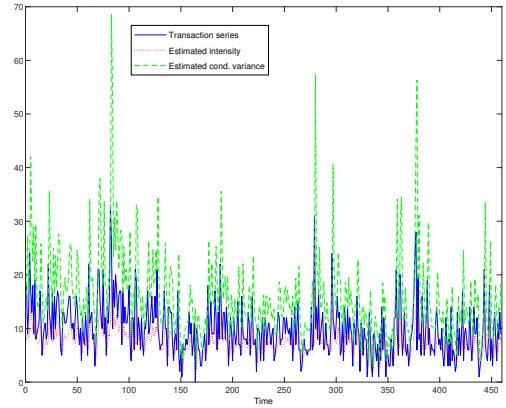
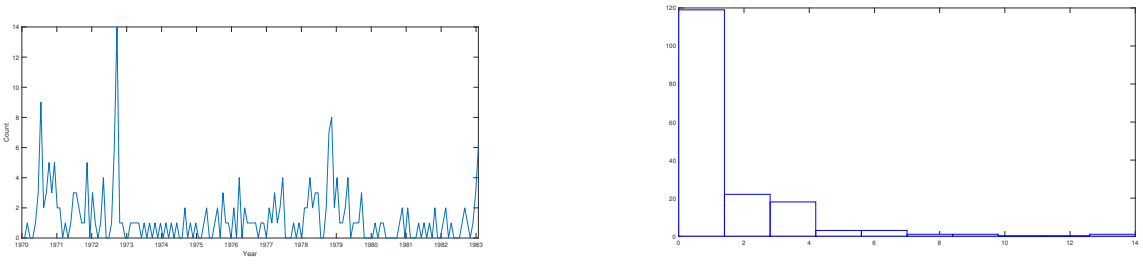


Figure 9: Empirical results (Transaction data). Real series and estimated intensities (conditional mean and variance) for the $NB\text{-}INSV$ model.



(a) Time series. (b) Histogram.

Figure 10: Monthly number of poliomyelitis cases in the United States from 1970 to 1983.

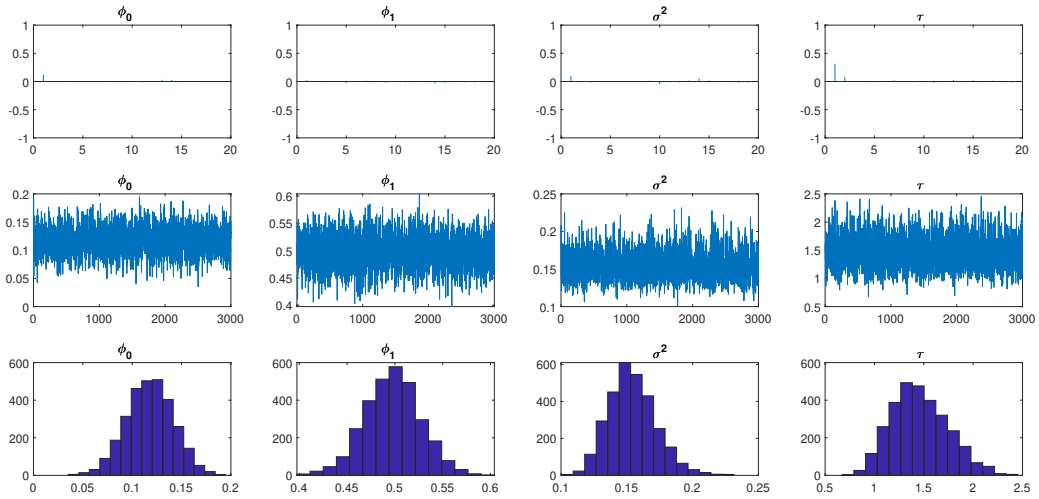
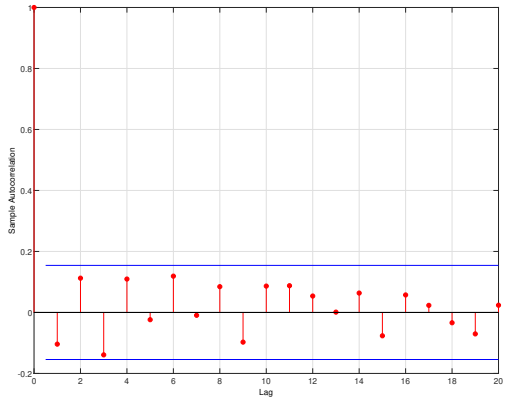
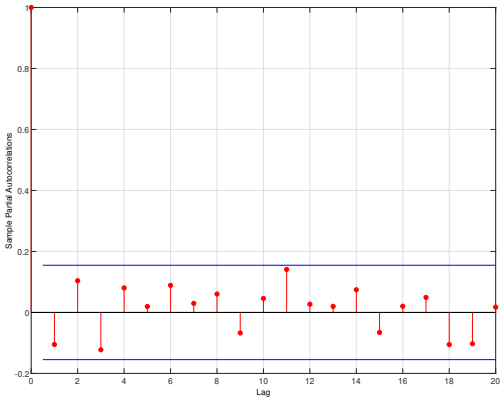


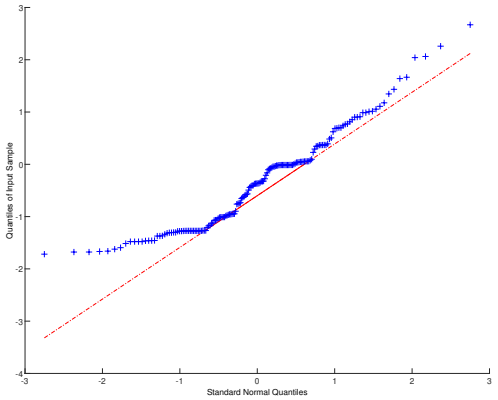
Figure 11: Empirical results (Polio data). Posterior autocorrelations (top), posterior paths (middle) and posterior histograms (bottom) for the parameters of the $NB\text{-}INSV$ model.



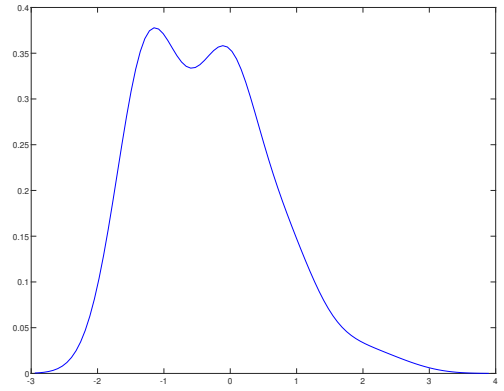
(a) Sample autocorrelations of the e-residuals.



(b) Partial autocorrelations of the e-residuals.

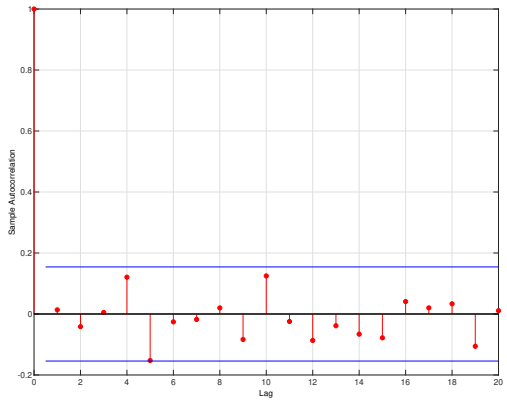


(c) QQ-plot of the e-residuals versus the standard normal distribution.

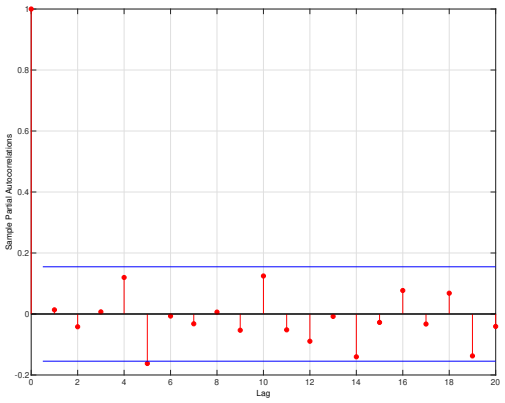


(d) Kernel density of the e-residuals.

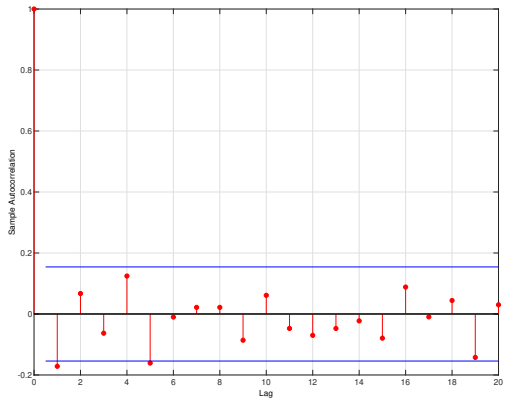
Figure 12: Empirical results (Polio data). e-residual analysis for the $NB\text{-}INSV$ model.



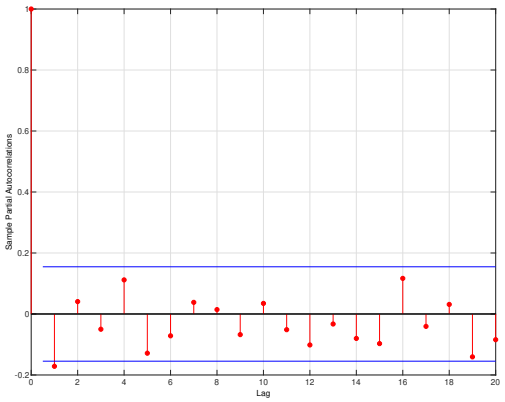
(a) Sample autocorrelations of the squared e-residuals.



(b) Partial autocorrelations of the squared e-residuals.

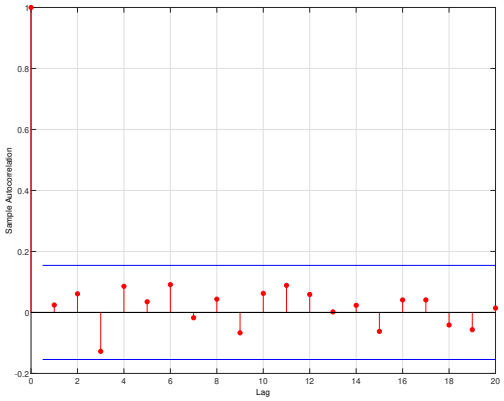


(c) Sample autocorrelations of the absolute e-residuals.

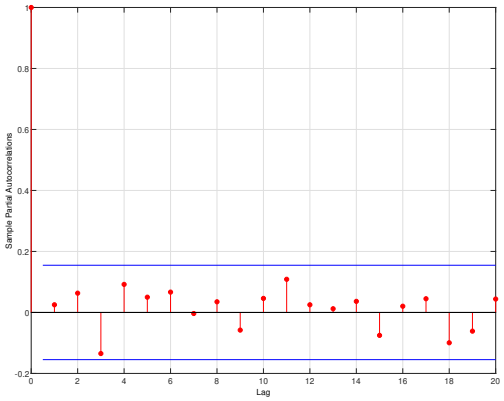


(d) Partial autocorrelations of the absolute e-residuals.

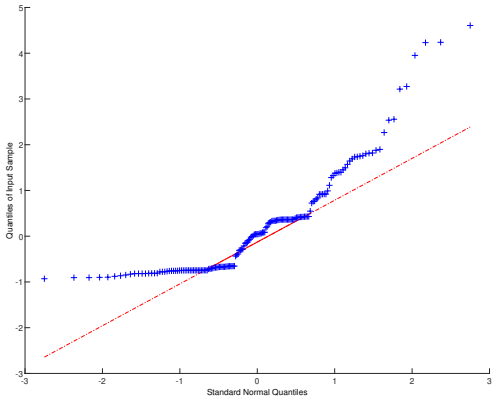
Figure 13: Empirical results (Polio data). Squared and absolute e-residual analysis for the $NB\text{-}INSV$ model.



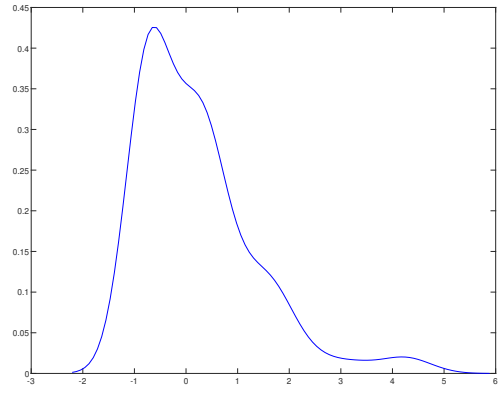
(a) Sample autocorrelations of the Pearson residuals.



(b) Partial autocorrelations of the Pearson residuals.

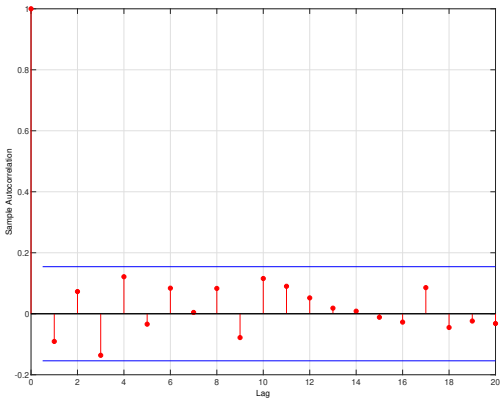


(c) QQ-plot of the Pearson residuals versus the standard normal distribution.

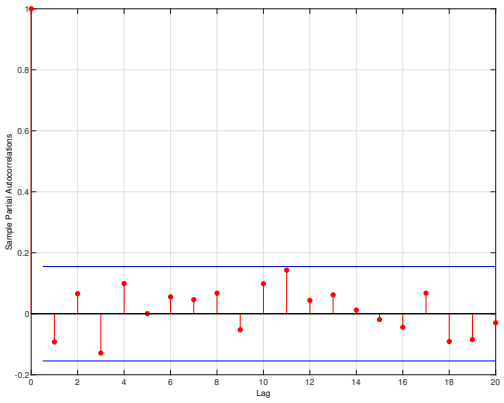


(d) Kernel density of the Pearson residuals.

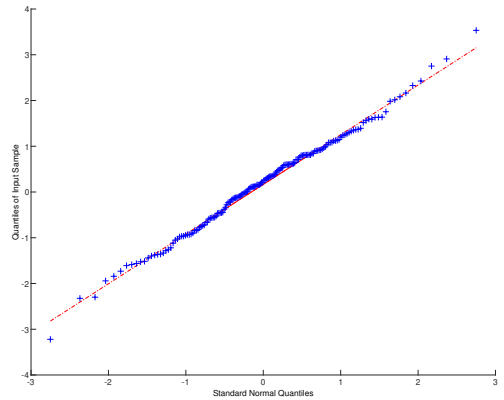
Figure 14: Empirical results (Polio data). Pearson analysis for the $NB\text{-}INSV$ model.



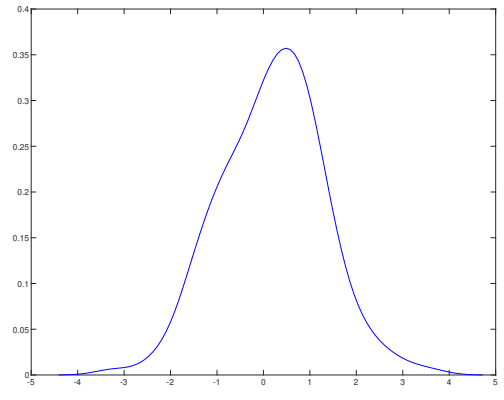
(a) Sample autocorrelations of the randomized quantile residuals.



(b) Partial autocorrelations of the randomized quantile residuals.

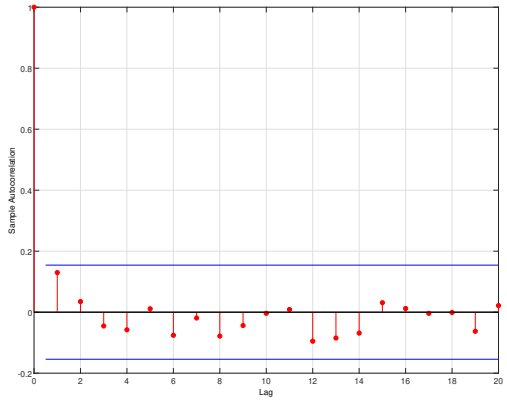


(c) QQ-plot of the randomized quantile residuals versus the standard normal distribution.

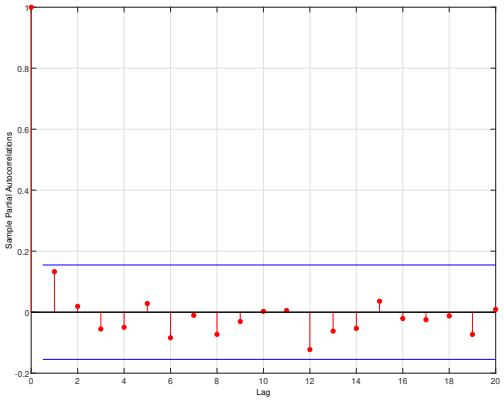


(d) Kernel density of the randomized quantile residuals.

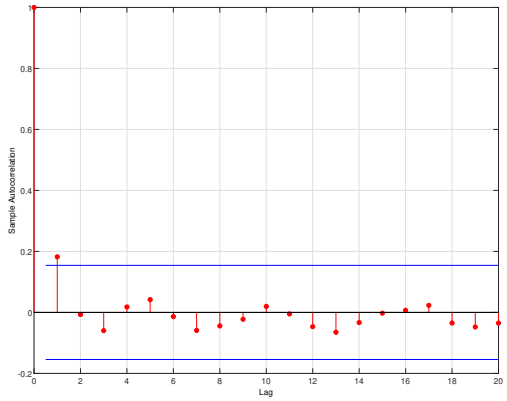
Figure 15: Empirical results (Polio data). Randomized quantile residual analysis for the $NB\text{-}INSV$ model.



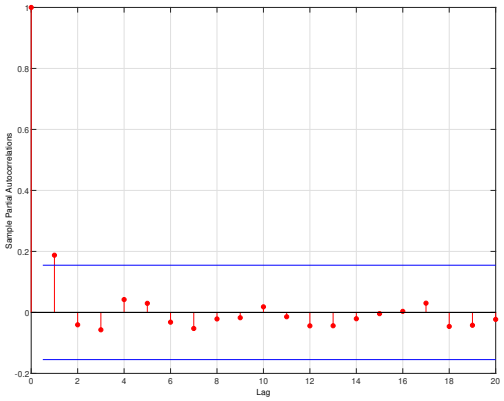
(a) Sample autocorrelations of the squared randomized quantile residuals.



(b) Partial autocorrelations of the squared randomized quantile residuals.



(c) Sample autocorrelations of the absolute randomized quantile residuals.



(d) Partial autocorrelations of the absolute randomized quantile residuals.

Figure 16: Empirical results (Polio data). Squared and absolute randomized quantile residual analysis for the $NB\text{-}INSV$ model.

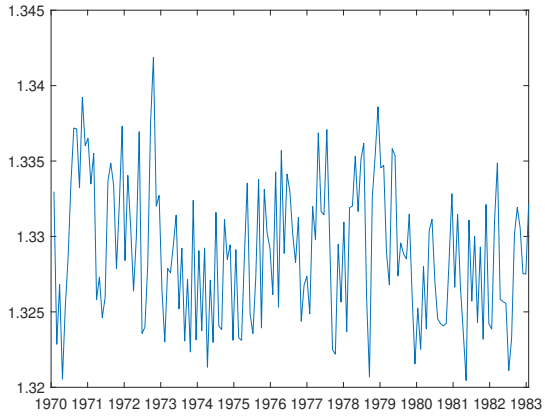


Figure 17: Empirical results (Polio data). Estimated latent series $\widehat{W}_t$ for the $NB\text{-}INSV$ model.

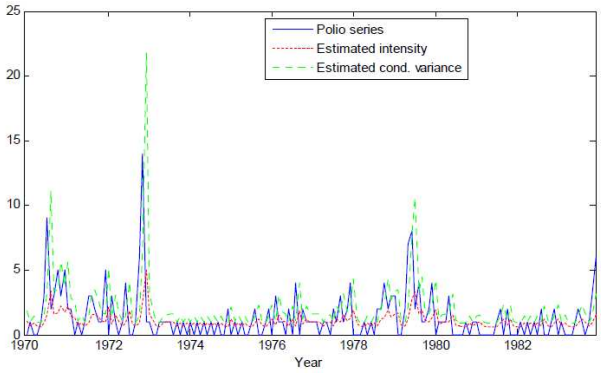


Figure 18: Empirical results (Polio data). Real series and estimated intensities (conditional mean and variance) for the $NB\text{-}INSV$ model.

Table 1: Simulation results for the $P\text{-}INSV$ model (Algorithm 1).

		ϕ_0	ϕ_1	σ^2
MCE-1	$True$	0.4	0.8	0.5
	$Mean$	0.4273	0.7932	0.5074
	Std	0.0449	0.0189	0.0231
	$RMSE$	0.0525	0.0201	0.0243
MCE-2	$True$	1.2	0.9	0.2
	$Mean$	1.2031	0.9001	0.2056
	Std	0.1670	0.0138	0.0092
	$RMSE$	0.1750	0.0143	0.0107
MCE-3	$True$	-0.1	0.3	0.1
	$Mean$	-0.0825	0.2901	0.1116
	Std	0.0111	0.0306	0.0051
	$RMSE$	0.0207	0.0321	0.0127

Table 2: Simulation results for the $NB\text{-}INSV$ model (Algorithm 2).

		ϕ_0	ϕ_1	σ^2	τ
MCE-1	$True$	1	0.6	0.2	7
	$Mean$	1.0675	0.6299	0.1998	7.1166
	Std	0.0620	0.0241	0.0090	0.0555
	$RMSE$	0.0949	0.0384	0.0100	0.1113
MCE-2	$True$	1.5	0.8	0.4	6
	$Mean$	1.4736	0.8047	0.4111	6.1000
	Std	0.1380	0.0180	0.0182	0.0476
	$RMSE$	0.2210	0.0306	0.0213	0.0954
MCE-3	$True$	2	0.5	0.1	5
	$Mean$	2.0386	0.4958	0.0886	5.0833
	Std	0.1063	0.0260	0.0040	0.0396
	$RMSE$	0.2540	0.0700	0.0121	0.0795

Table 3: Simulation results for the $NB\text{-}INSV$ model (Algorithm 3).

		ϕ_0	ϕ_1	σ^2	τ
MCE-1	$True$	1	0.6	0.2	7
	$Mean$	1.0414	0.5814	0.1966	7.0386
	Std	0.0651	0.0255	0.0088	0.0183
	$RMSE$	0.0772	0.0316	0.0094	0.0366
MCE-2	$True$	1.5	0.8	0.4	6
	$Mean$	1.6680	0.7731	0.4081	6.0331
	Std	0.1515	0.0204	0.0180	0.0157
	$RMSE$	0.2262	0.0337	0.0198	0.0314
MCE-3	$True$	2	0.5	0.1	5
	$Mean$	2.0236	0.5166	0.0994	5.0276
	Std	0.1098	0.0271	0.0044	0.0131
	$RMSE$	0.1191	0.0317	0.0046	0.0261

Table 4: Empirical results. Estimated DIC values for the proposed models (Transaction data).

	$P\text{-}INSV$	$NB\text{-}INSV$ (Algorithm 3)
DIC	5851.9099	5835.9655
$\left(\sqrt{var(DIC)}\right)$	(1.15012)	(1.2309)
Rank	2	1

Table 5: Empirical results. BGG estimates for the $NB\text{-}INSV$ model (Transaction data).

Parameters	Mean	Std	RNI	NSE	CD
ϕ_0	0.7112*	0.0273	0.7751	0.0004	0.9672
ϕ_1	0.6472*	0.0144	0.8260	0.0002	-0.7989
σ^2	0.0770*	0.0060	1.3943	0.0001	0.0964
τ	7.5495*	0.4790	1.5827	0.0102	-0.5877

*Significant based on the 95% highest posterior density interval.

Table 6: Empirical results. *BGG* estimates for the *P-INSV* model (Transaction data).

Parameters	Mean	*Std*	*RNI*	*NSE*	*CD*
ϕ_0	1.6190*	0.0285	0.2727	0.0002	0.5142
ϕ_1	0.2452*	0.0153	0.3785	0.0001	0.5228
σ^2	0.1195*	0.0118	2.5212	0.0003	1.6496

*Significant based on the 95% highest posterior density interval.

Table 7: Empirical results. Estimated *DIC* values for the proposed models (Polio data).

	P-INSV	*NB-INSV* (Algorithm 3)
DIC $\left(\sqrt{var(DIC)}\right)$	-107.8545 (0.9002)	-120.7663 (0.9895)
Rank	2	1

Table 8: Empirical results. *BGG* estimates for the *NB-INSV* model (Polio data).

Parameters	Mean	*Std*	*RNI*	*NSE*	*CD*
ϕ_0	0.1178*	0.0237	0.7025	0.0003	-0.7304
ϕ_1	0.4976*	0.0292	0.8479	0.0005	-0.0165
σ^2	0.1543*	0.0181	1.4121	0.0004	-2.1833
τ	1.4605*	0.2837	1.5985	0.0061	-0.2562

*Significant based on the 95% highest posterior density interval.

Table 9: Empirical results. *BGG* estimates for the *P-INSV* model (Polio data).

Parameters	Mean	*Std*	*RNI*	*NSE*	*CD*
ϕ_0	0.1994*	0.0173	0.3447	0.0014	0.9678
ϕ_1	0.4874*	0.0260	0.2721	0.0019	-0.9325
σ^2	0.1879*	0.0265	0.2120	0.0017	0.8572

*Significant based on the 95% highest posterior density interval.

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 $\{(e_t, Z_t), t \in \mathbb{Z}\}$. Hence, $\{Y_t, t \in \mathbb{Z}\}$ is strictly stationary and ergodic. If $|\phi| \geq 1$, then clearly there does not exist a nonanticipative strictly stationary solution of (1)-(2), like expression (9). □

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 $\{Y_t, a \leq t \leq b\}$. [3]

□

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

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

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).

□

[3] 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, $e_t \sim N\left(0,1\right)$, 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 $h > 0$

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

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

If $e_t \sim N\left(0,1\right)$, 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)

$$E\left(Y_t^s/\lambda_t\right) = \sum_{i=0}^{r} \left\{ {s \atop i} \right\} \lambda_t^i.$$

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

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

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

Abanto-Valle, C.A., Migon, H.S., Lachos V.H. (2011). Stochastic volatility in mean models with scale mixtures of normal distributions and correlated errors: A Bayesian approach. *Journal of Statistical Planning and Inference* 141: 1875–1887.

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., Demouche, N. (2019). Ergodicity conditions for a double mixed Poisson autoregression. *Statistics and Probability Letters, forthcoming.*

Aknouche, A., Francq, C. (2018). Count and duration time series with equal conditional stochastic and mean orders. *MPRA paper 90838.*

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

Box, G.E.P., Tiao, G.C. (1973). *Bayesian Inference in Statistical Analysis.* Addison-Wesley: Reading, MA.

Bauwens, L., Hautsch, N. (2006). Stochastic conditional intensity processes. *Journal of Financial Econometrics* 4: 450–493.

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

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.

Cameron, A., Trivedi, P. (2013). *Regression Analysis of Count Data.* Cambridge University Press.

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

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.

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., Wang, Y. (2000). On autocorrelation in a Poisson regression model. *Biometrika* 87: 491-505.

Davis, R.A., Dunsmuir, W.T. (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.

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.

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

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.

Geweke, J. (1989). Bayesian inference in econometric models using Monte Carlo integration. *Econometrica* 57: 1317–1339.

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.

Geyer, C.J. (1992). Practical Markov chain Monte Carlo. *Statistical Science* 7: 473–483.

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., 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.

Jacquier, E., Polson, N.G., Rossi, P.E. (1994). Bayesian analysis of stochastic volatility models. *Journal of Business & Economic Statistics* 12: 371–389.

Jacquier, E., Polson, N.G., Rossi, P.E. (2004). Bayesian analysis of fat-tailed stochastic volatility models with correlated errors. *Journal of Econometrics* 122: 185–212.

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.

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.

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

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

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.

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

Ritter, C., Tanner, M.A. (1992). Facilitating the Gibbs sampler: the Gibbs stopper and the Griddy-Gibbs sampler. *Journal of the American Statistical Association* 87: 861-868.

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*.

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

Spiegelhalter, D., Best, N., Carlin, B., Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64:

583-639.

Taylor, S. (1986). *Modelling Financial Time Series*. Chichester: Wiley.

Tsay, R.S. (2010). *Analysis of Financial Time Series*. Wiley.

Tsiakas, I. (2006). Periodic stochastic volatility and fat tails. *Journal of Financial Econometrics* 4: 90–135.

Wang, J.J., Chana, J.S.K., Choy, S.T.B. (2011). Stochastic volatility models with leverage and heavy-tailed distributions: A Bayesian approach using scale mixtures. *Computational Statistics & Data Analysis* 55: 852–862.

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.