A multiplicative thinning-based integer-valued GARCH model

Aknouche, Abdelhakim and Scotto, Manuel

Department of Mathematics, College of Science, Qassim University, Department of Mathematics and CEMAT, University of Lisbon

20 March 2022

A multiplicative thinning-based integer-valued GARCH model

Abdelhakim Aknouche

Department of Mathematics, College of Science, Qassim University

and

Manuel G. Scotto

Department of Mathematics and CEMAT, University of Lisbon

Abstract

In this paper we introduce a multiplicative integer-valued time series model, which is defined as the product of a unit-mean integer-valued independent and identically distributed (iid) sequence, and an integer-valued dependent process. The latter is defined as a binomial thinning operation of its own past and of the past of the observed process. Furthermore, it combines some features of the integer-valued GARCH (INGARCH), the autoregressive conditional duration (ACD), and the integer autoregression (INAR) processes. The proposed model is semi-parametric and is able to parsimoniously generate very high overdispersion, persistence, and heavy-tailedness. The dynamic probabilistic structure of the model is first analyzed. In addition, parameter estimation is considered by using a two-stage weighted least squares estimate (2SWLSE), consistency and asymptotic normality (CAN) of which are established under mild conditions. Applications of the proposed formulation to simulated and actual count time series data are provided.

Keywords: Integer-valued time series, INAR model, INGARCH model, multiplicative error model (MEM), ACD model, two-stage weighted least squares.

1 Introduction

The normal approximation to many usual discrete distributions such as binomial, Poisson and negative binomial, suggests that integer-valued data taking large enough values, uniformly, can be validly modeled by continuous-valued models. In this way, a wide spectrum of time series data with quite large integer values was approximated by continuous-valued models (e.g. Box et *al*, 1994; Brockwell and Davis, 1991). Notable examples are the celebrated Canadian lynx data and the sunspot data, see e.g. Box et *al*, (1994) and Wong and Li (2000) for details.

When integer-valued data are characterized by values that are too low for the normal approximation to be valid, it is necessary to resort to appropriate integer-valued models. This is why numerous integer-valued models have been introduced so far to highlight integer-valued traits frequently observed in practice, such as overdispersion, excess of zeros, asymmetry, and persistence (e.g. Scotto et *al*, 2015; Davis et *al*, 2016; Weiss, 2018). These models can broadly be grouped into two general classes: Equation-based models and distribution-based models. The first category mainly includes thinning-based stochastic equations where a special role is played by the INARMA model (e.g. McKenzie, 1985-2003; Al-Osh and Alzaid, 1987; Scotto et *al*, 2015). The second category encompasses integer-valued generalized ARMA models (Benjamin et *al*, 2003; Zheng et *al*, 2015), the best-known example of which is the integer-valued GARCH model (INGARCH in short, see e.g. Grunwald et *al*, 2000; Heinen, 2003; Ferland et *al*, 2006; Fokianos et *al*, 2009; Davis and Liu, 2016).

The class of thinning-based equations (e.g. McKenzie, 1985-2003; Al-Osh and Alzaid, 1987; Du and Li, 1991; Latour, 1998; Silva and Oliveira, 2004; Scotto et *al*, 2015; Weiss, 2018) has dominated the literature on count time series modeling due to various advantages. In particular, these models: i) Can be set semi-parametric without fully specifying the distribution of innovation, ii) their dynamic probability structures, and in particular, the marginal distribution, are relatively simple to unmask, iii) they can be easily estimated using least-squares and exponential family quasi-maximum likelihood methods due to the fact that the conditional mean in such models is generally observable, and finally, iv) they can be easily extended to multivariate settings. However, even with the most general INARMA model, persistence, long memory, and heavy-tailedness are not obviously reproduced.

On the contrary, INGARCH-type models are specified via a conditional distribution with a time-varying mean depending on past observations. Such models can exhibit several important features such as persistence and overdispersion, depending on the assumed conditional distribution. In addition, their estimation can be easily performed using (exact and quasi-) maximum likelihood and least squares-based methods. However, they have the disabilities: i) Of being generally fully parametric, ii) of having relatively complex probability structures because they are not defined through stochastic equations involving iid innovations, and iii) of being difficult to generalize to multivariate forms.

Besides uniformly low-valued time series which cannot be properly captured by continuous-valued models, there are also integer-valued series taking low values with large probabilities and also exhibiting sudden bursts of large values (e.g. Hall et *al*, 2010). These values, also known as outliers (e.g. Barczy et *al*, 2010, 2012; Wan and Chan, 2011; Silva et *al*, 2019), induce high overdispersion and entail distributions with fat tails. Standard INAR and INGARCH models are generally unable to account for this type of data, so well-adapted heavy-tailed models are needed. In general, heavy-tailedness can be introduced into a time series model in two general ways. The first one is by considering a heavy-tailed distribution, either through the conditional distribution when the model is distribution-based, or via the innovation distribution when the model is equation-based. Examples are the INGARCH model with beta negative binomial conditional distribution (Gorgi, 2020), and the INAR(1) equation with generalized Poisson inverse Gaussian distributed innovation (Qian et *al*, 2020). The second way to generate heavy-tailedness is via the mechanism of the conditional mean dynamics. The best-known examples are by considering nonlinear adapted forms (e.g. Bilinear INAR, Doukhan et *al*, 2006; Max-INAR(1), Scotto et *al*, 2018) or by including random parameters in the conditional mean dynamics such as independent mixture, Markov mixture, and threshold mixture integer-valued GARCH models (see Aknouche and Francq, 2022*b* and the references therein). The main drawback with the former approach, however, is that it involves specifying the full distribution, which is rarely known in practice. The latter approach has the drawback of requiring complicated formal settings or/and a lot of parameters which are not easily estimated.

The aim of this paper is to design a simple equation-based integer-valued time series model that allows for modeling persistence and high overdispersion. Moreover, it is well adapted to heavy-tailed data irrespective of the choice of the innovation distribution, and without resorting to complicated random coefficient equations. Inspired by the Autoregressive Conditional Duration (ACD) model of Engle and Russel (1998) (and more generally MEM forms, Engle, 2002; Engle and Gallo, 2006), the proposed integer-valued model results from the multiplication of a unit mean iid integer-valued innovation by an integer-valued scale, dependent process. The multiplicative form allows for a large variability because the resulting multiplicative process can oscillate between large and small values due to zero absorption. Alike to ACD and INGARCH models, the scale process will depend on its lagged values and those of the observed process. In order to ensure discreteness, the dependence on past values is rather generated by the binomial thinning operation of Steutel and Van Harn (1979), in the spirit of the INAR(p) model of Du and Li (1991).

We first study the dynamic structure of the model, namely the existence of stationary and ergodic solutions, the autocovariance structure, and its tail behavior. Moreover, parameter estimation will be performed using a two-stage weighted least squares estimator, consistency and asymptotic normality of which are established under general conditions.

The rest of this paper has the following structure. In Section 2, the proposed model is introduced. Section 3 set up the WLS estimation procedures. In Section 4, the WLSE is compared in finite samples with Poisson, exponential and negative binomial quasi-maximum likelihood estimates. Section 5 applies the model to two real-life examples. Finally, Section 6 is devoted to conclusions. The proofs of the main results are left in the appendix.

2 The model

In the sequel, $\mathbb{N} = \{0, 1, 2...\}$ and $\mathbb{Z} = \{..., -1, 0, 1, ...\}$ denote, respectively, the sets of nonnegative integers and the set of integers.

Let $\{\varepsilon_t, t \in \mathbb{Z}\}$ be a $\mathbb{N}$-valued sequence of independent and identically distributed (iid) random variables defined on a probability space $(\Omega, \mathcal{F}, P)$ with one-mean and finite variance, say σ_0^2. The multiplicative thinning-based INGARCH (MthINGARCH) model we deal with in this paper is an $\mathbb{N}$-valued sequence $\{Y_t, t \in \mathbb{Z}\}$ defined on $(\Omega, \mathcal{F}, P)$ as a solution to the following stochastic equation

$$\begin{aligned} Y_t &= \lambda_t \varepsilon_t && (2.1a) \\ \lambda_t &= 1 + \omega_0 \circ m + \sum_{i=1}^{q} \alpha_{0i} \circ Y_{t-i} + \sum_{j=1}^{p} \beta_{0j} \circ \lambda_{t-j}, && (2.2b) \end{aligned}$$

where the symbol $\circ$ stands for the binomial thinning operator (Steutel and Van Harn, 1979) defined by $\alpha \circ X := \sum_{i=1}^{X} W_i$, where $X > 0$, and the implied counting sequence $\{W_i, i \in \mathbb{N}\}$ is iid Bernoulli distributed with parameter $\alpha \in (0, 1)$, which is assumed to be independent of X. The coefficients in (2.2b) satisfy $0 \leq \omega_0 \leq 1$, $0 \leq \alpha_{0i} < 1$ and $0 \leq \beta_{0j} < 1$ ($i = 1, ..., q$, $j = 1, ..., p$), and m is a fixed positive integer number introduced for more flexibility. In real-world applications, one can set m as being the integer part of the sample mean, $\frac{1}{n} \sum_{i=1}^{n} y_i$, given an observed time series $y_1, ..., y_n$. The introduction of the term 1 in the right-hand side of (2.2b) ensures $\lambda_t \geq 1$ a.s. As in Du and Li (1991), it is assumed that the Bernoulli terms corresponding to the binomial variables $\omega_0 \circ m$, $\alpha_{0i} \circ Y_{t-i}$ and $\beta_{0j} \circ \lambda_{t-j}$ are mutually independent, and independent of the sequence $\{\varepsilon_t, t \in \mathbb{Z}\}$.

Note that when $\beta_{0j} = 0$ for all j, the model reduces to the following multiplicative INAR(q) equation

$$Y_t = \left(1 + \omega_0 \circ m + \sum_{i=1}^{q} \alpha_{0i} \circ Y_{t-i}\right) \varepsilon_t. \tag{2.2}$$

Further, denote by $\mathcal{F}_t^Y$ and $\mathcal{F}_{t-1}^{Y,\lambda}$ the σ-algebras generated by $\{Y_s, s \leq t\}$ and $\{(Y_s, \lambda_s),$ $s \leq t\}$, respectively, i.e. the observed and complete information history up to time t. Let $\mu_t := E\left(Y_t | \mathcal{F}_{t-1}^Y\right)$ and $\overline{\mu}_t := E\left(Y_t | \mathcal{F}_{t-1}^{Y,\lambda}\right)$ be the observed and complete conditional means of the process given $\mathcal{F}_{t-1}^{Y}$ and $\mathcal{F}_{t-1}^{Y,\lambda}$, respectively. Note that

$$\mu_t = E\left(\lambda_t | \mathcal{F}_{t-1}^{Y}\right) \quad \text{and} \quad \overline{\mu}_t = E\left(\lambda_t | \mathcal{F}_{t-1}^{Y,\lambda}\right). \tag{2.3}$$

It is important to point out that μ_t is observable whereas $\overline{\mu}_t$ is not. The expressions of these conditional means are given as follows.

Proposition 2.1 *For all* $t \in \mathbb{Z}$

$$\overline{\mu}_t = 1 + \omega_0 m + \sum_{i=1}^{q} \alpha_{0i} Y_{t-i} + \sum_{j=1}^{p} \beta_{0j} \lambda_{t-j} \tag{2.4}$$

$$\mu_t = 1 + \omega_0 m + \sum_{i=1}^{q} \alpha_{0i} Y_{t-i} + \sum_{j=1}^{p} \beta_{0j} \mu_{t-j}. \tag{2.5}$$

Regarding the observable conditional mean recursion (2.5), it turns out that model (2.1) has the same conditional mean dynamics as the INGARCH and ACD models. However, unlike the ACD and MEM models (e.g. Engle and Russell, 1998; Engle, 2002; Engle and Gallo, 2006), the unobserved scale term λ_t does not represent the conditional mean of the model. Nonetheless, it can be considered as an approximation of μ_t in the sense of (2.3).

We now study the existence of a causal strictly stationary and ergodic solution to equation (2.1). The finiteness of $E(\varepsilon_t)$ is immediately satisfied by model's definition.

Theorem 2.1 *There exists a non-anticipative strictly stationary and ergodic solution* $\{(Y_t, \lambda_t), t \in \mathbb{Z}\}$ *to* (2.1) *if*

$$\sum_{i=1}^{q} \alpha_{0i} + \sum_{j=1}^{p} \beta_{0j} < 1. \tag{2.6}$$

Conversely, if there is a stationary in mean solution to (2.1) *then* (2.6) *holds true.*

The unconditional mean of the process is given from (2.6) and $(A.4)$ (in the Appendix) by

$$\mu := E(Y_t) = E(\lambda_t) = \frac{1 + \omega_0 m}{1 - \left(\sum_{i=1}^{q} \alpha_{0i} + \sum_{j=1}^{p} \beta_{0j}\right)}. \tag{2.7}$$

We now derive the second-order structure of the model. Denote by $v_t := Var\left(\lambda_t | \mathcal{F}_{t-1}^{Y}\right)$ and $\overline{v}_t := Var\left(\lambda_t | \mathcal{F}_{t-1}^{Y,\lambda}\right)$ the observed, and complete conditional variances of λ_t with respect to $\mathcal{F}_{t-1}^{Y}$ and $\mathcal{F}_{t-1}^{Y,\lambda}$, respectively. Accordingly, let $\vartheta_t := Var\left(Y_t|\mathcal{F}_{t-1}^{Y}\right)$ and $\overline{\vartheta}_t := Var\left(Y_t|\mathcal{F}_{t-1}^{Y,\lambda}\right)$ be the observed and complete conditional variances of Y_t. We have the following results.

Proposition 2.2 *For all* $t \in \mathbb{Z}$

$$\overline{\upsilon}_t = \omega_0 (1-\omega_0) m + \sum_{i=1}^{q} \alpha_{0i} (1-\alpha_{0i}) Y_{t-i} + \sum_{j=1}^{p} \beta_{0j} \left(1-\beta_{0j}\right) \lambda_{t-j} \tag{2.8}$$

$$\overline{\vartheta}_t = \left(\sigma_0^2+1\right) \overline{\upsilon}_t + \sigma_0^2 \overline{\mu}_t^2 \tag{2.9}$$

$$\upsilon_t = \omega_0 (1-\omega_0) m + \sum_{i=1}^{q} \alpha_{0i} (1-\alpha_{0i}) Y_{t-i} + \sum_{j=1}^{p} \left(\beta_{0j} \left(1-\beta_{0j}\right) \mu_{t-j} + \beta_{0j}^2 \upsilon_{t-j}\right) \tag{2.10}$$

$$\vartheta_t = \left(\sigma_0^2+1\right) \upsilon_t + \sigma_0^2 \mu_t^2. \tag{2.11}$$

As can be seen from (2.10)-(2.11), the conditional variance-to-mean relationship has an unusual form, especially adapted to high overdispersion and strong persistence when $\beta_{0j} \neq 0$ for some j. In particular, compared to the INARMA and INGARCH processes, model (2.1) allows for a higher conditional overdispersion even for Poisson-distributed innovations. Such overdispersion is primarily generated from the model mechanism, and not only from the distribution of the innovation/model as use to be the case in INGARCH and INAR models. The following result provides a necessary and sufficient condition for the existence of a second-order stationary solution to (2.1). Set $r := \max(p, q)$.

Theorem 2.2 *Assume that* $E\left(\varepsilon_t^2\right) < \infty$. *There exists a non-anticipative stationary (and ergodic) solution* $\{(Y_t, \lambda_t), t \in \mathbb{Z}\}$ *to* (2.1) *with* $E\left(Y_t^2\right) < \infty$ *and* $E\left(\lambda_t^2\right) < \infty$ *if and only if*

$$\sum_{i=1}^{r} \left(\alpha_{0i} + \beta_{0i}\right)^2 + \sigma_0^2 \alpha_{0i}^2 < 1. \tag{2.12}$$

As expected, (2.12) implies (2.6), that is to say, the second-order stationarity domain is strictly included in the mean-stationarity region (cf. Figure 2.1), except for the degenerate case $\sigma_0^2 = 0$ with $p = q = 1$. Note that (2.12) implies the following conditional variance invertibility condition

$$\sum_{j=1}^{p} \beta_{0j}^2 < 1,$$

(cf. (2.10)), which is implied by the conditional mean invertibility condition

$$\sum_{j=1}^{p} \beta_{0j} < 1. \tag{2.13}$$

The two latter conditions will be imposed for the consistency of the weighted least squares estimator to be considered in Section 3.

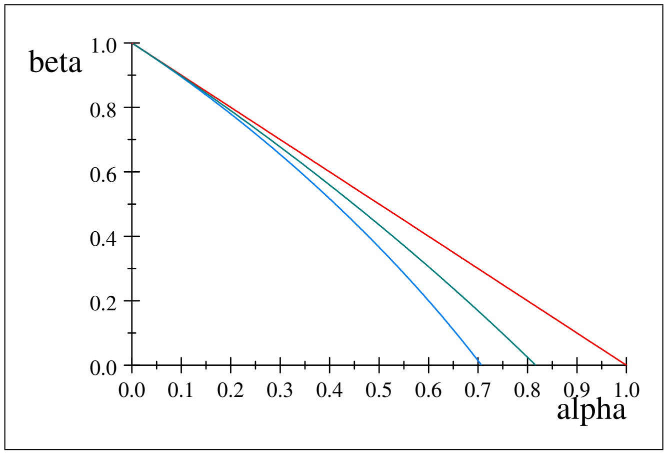


Figure 2.1 Stationarity domains for $p = q = 1$: Mean stationarity delimited by the red line. Variance stationarity delimited by the green line for $\sigma_0^2 = \frac{1}{2}$, and by the blue line for $\sigma_0^2 = 1$.

Let $\gamma_k := Cov\,(Y_t, Y_{t-k})$. Under (2.12), the unconditional variance and autocovariance function of the model (2.1) can be obtained from $(A.9)$ and $(A.10)$.

Proposition 2.3 *Under* (2.12)

$$Var\,(\lambda_t) = \frac{\omega_0(1-\omega_0)m+\mu\sum_{i=1}^{r}(\alpha_{0i}(1-\alpha_{0i})+\beta_{0i}(1-\beta_{0i}))+\sigma_0^2\mu^2\sum_{i=1}^{q}\alpha_{0i}^2+2\sum_{i\neq j}\alpha_{0i}\beta_{0j}Cov(\lambda_{t-i},\lambda_{t-j})}{1-\sum_{i=1}^{r}\left((\alpha_{0i}+\beta_{0i})^2+\sigma_0^2\alpha_{0i}^2\right)} \tag{2.14}$$

$$Var\,(Y_t) = \frac{\left(\sigma_0^2+1\right)\left[\omega_0(1-\omega_0)m+\mu\sum_{i=1}^{r}(\alpha_{0i}(1-\alpha_{0i})+\beta_{0i}(1-\beta_{0i}))\right]+\sigma_0^2\mu^2\left(1-\sum_{i=1}^{r}\beta_{0i}^2+2\alpha_{0i}\beta_{0i}\right)}{1-\sum_{i=1}^{r}\left((\alpha_{0i}+\beta_{0i})^2+\sigma_0^2\alpha_{0i}^2\right)} + \frac{2\left(\sigma_0^2+1\right)\sum_{i\neq j}\alpha_{0i}\beta_{0j}Cov(\lambda_{t-i},\lambda_{t-j})}{1-\sum_{i=1}^{r}\left((\alpha_{0i}+\beta_{0i})^2+\sigma_0^2\alpha_{0i}^2\right)}, \tag{2.15}$$

$$\gamma_k = \omega_0\,(1-\omega_0)\,m + \sum_{i=1}^{r}\left(\alpha_{0i}+\beta_{0i}\right)\gamma_{k-i}, \quad k \geq 1. \tag{2.16}$$

For example, when $p = q = 1$, the unconditional variances and the autocovariance function take the forms

$$\begin{aligned} Var(\lambda_t) &= \frac{\omega_0(1-\omega_0)m+(\alpha_{01}(1-\alpha_{01})+\beta_{01}(1-\beta_{01}))\mu+\alpha_{01}^2\sigma_0^2\mu^2}{1-\left((\alpha_{01}+\beta_{01})^2+\sigma_0^2\alpha_{01}^2\right)} \\ Var(Y_t) &= \frac{\left(\sigma_0^2+1\right)[\omega_0(1-\omega_0)m+(\alpha_{01}(1-\alpha_{01})+\beta_{01}(1-\beta_{01}))\mu]+\sigma_0^2\mu^2\left(1-\left(\beta_{01}^2+2\alpha_{01}\beta_{01}\right)\right)}{1-\left((\alpha_{01}+\beta_{01})^2+\sigma_0^2\alpha_{01}^2\right)}, \end{aligned}$$

and

$$\gamma_k = \omega_0 (1-\omega_0) m + (\alpha_{01} + \beta_{01}) \gamma_{k-1}, \;\; k \geq 1.$$

From (2.14)-(2.15), it appears that the MthINGARCH model (2.1) is able to generate very high overdispersion. In the supplementary material to this paper, simulations experiments validate formulas (2.14)-(2.15) in the sense that the sample two (first) moments are very close to their theoretical counterparts. Moreover, the simulations also show that the generated MthINGARCH series can be highly overdispersed, with sample variances reaching 90 times the sample mean, under the second-order stationarity condition.

Denote by $e_t = Y_t - \mu_t$ the one-step ahead forecast error. Note that e_t can be decomposed into the sum $e_t = \eta_t + \zeta_t$, where $\eta_t = Y_t - \lambda_t = \lambda_t (\varepsilon_t - 1)$ and $\zeta_t = \lambda_t - \mu_t$ are terms of martingale difference sequences with respect to the filtration $\left\{\mathcal{F}_t^Y, t \in \mathbb{Z}\right\}$ where

$$Var\left(\eta_t | \mathcal{F}_{t-1}^Y\right) = \sigma_0^2 \left(v_t + \mu_t^2\right) \text{ and } Var\left(\zeta_t | \mathcal{F}_{t-1}^Y\right) = v_t.$$

Therefore, the process $\{Y_t, t \in \mathbb{Z}\}$ can be written in the following modified integer-valued ARMA (INARMA(q,p)) form

$$Y_t = 1 + \omega_0 \circ m + \sum_{i=1}^{q} \alpha_{0i} \circ Y_{t-i} + \eta_t + \sum_{j=1}^{p} \beta_{0j} \circ \left(Y_{t-j} - \eta_{t-j}\right), \tag{2.17}$$

where the white noise sequence $\{\eta_t, t \in \mathbb{Z}\}$ is generally not iid in contrast with INARMA models. In particular, when $\beta_{0j} = 0$ for all j, we obtain

$$Y_t = 1 + \omega_0 \circ m + \sum_{i=1}^{q} \alpha_{0i} \circ Y_{t-i} + \eta_t,$$

which is a "weak" INAR(q), rather driven by a martingale difference which can be seen as an extension of the INAR(p) of Du and Li (1991) and Latour (1998).

We finally examine the tail properties of the model (2.1) when $p = q = 1$, $\alpha_{01} \in (0,1)$ and $\beta_{01} \in (0,1)$. The multiplicative form (2.1a) of the model entails the regularly tail equivalence between ε_t and Y_t under a certain moment condition on λ_t. Indeed, assume that ε_t is regularly varying with tail index $\xi > 0$, i.e. $P(\varepsilon_t > n) = n^{-\xi}L(n)$, where $L(.)$ is a slowly varying function at ∞. Since ε_t and λ_t are independent for each t, Breiman's (1965) theorem thus implies that: If $E(\lambda_t^\delta) < \infty$ for some $\delta > \xi$ then

$$\lim_{n\to\infty} \frac{P(Y_t > n)}{P(\varepsilon_t > n)} = E(\lambda_t^\xi).$$

We can also use again Breiman's theorem in a converse sense. If we assume that λ_t is regularly varying with tail index $\xi > 0$ and if $E(\varepsilon_t^\tau) < \infty$ for some $\tau > \xi$ then

$$\lim_{n\to\infty} \frac{P(Y_t > n)}{P(\lambda_t > n)} = E(\varepsilon_t^\xi).$$

We now turn to the less obvious inverse problem: If Y_t is regularly varying with tail $\xi > 0$, what about $P(\lambda_t > n)$ for sufficiently large values of n? The next result shows that Y_t and λ_t are indeed tail equivalent at infinity.

Theorem 2.3 *Let* $\{(Y_t, \lambda_t), t \in \mathbb{Z}\}$ *be a stationary solution to* (2.1) *satisfying* (2.12) *with* $p = q = 1$. *Assume that* Y_t *is regularly varying with tail* $\xi > 0$. *Then*

$$P(\lambda_t > n) \sim \frac{\alpha_{01}^\xi}{1-\beta_{01}^\xi} P(Y_t > n) \text{ as } n \to \infty.$$

In view of the latter results, it can be concluded that the input ε_t can exhibit a tail sufficiently light for the output Y_t to have a regularly varying tail. It suffices that the moment of ε_t is finite for a certain positive order (see also Denisov and Zwart, 2007 and Jacobsen et *al*, 2009). Since many usual discrete distributions such as the Poisson and Negative Binomial have finite moments up to any order, we conclude that model (2.1) can generate outputs with fat tails even from light-tailed inputs, such as with the Poisson and negative binomial distributions. This is in contrast with the linear (random coefficient) INAR(1) model where fat-tailed outputs are generated from fat-tailed inputs (Roitershtein and Zhong, 2013). See also Embrechts et *al* (2013, Section 8.4) for a similar comparison between real-valued linear ARMA and multiplicative GARCH models.

3 Three-stage weighted least squares estimation

Let $Y_1, Y_2, ..., Y_n$ be a series generated from model (2.1) with true parameters $\theta_0 := (\omega_0, \alpha_{01}, ..., \alpha_{0q}, \beta_{01}, ..., \beta_{0p})'$ and σ_0^2. Assume that $\theta_0 \in \Theta \subset [0,1] \times [0,1)^{p+q}$ and $\sigma_0^2 \in \Delta \subset (0, \infty)$. The latter condition ensures the non-degeneracy of ε_t. From (2.5), the observed conditional mean μ_t has the same form as that of the standard INGARCH model, so the parameter θ_0 involved in the conditional mean can be consistently estimated using any exponential-family quasi-maximum likelihood estimate (QMLE). In particular, the Poisson QMLE (PQMLE, Christou and Fokianos, 2014; Ahmad and Francq, 2016), the negative binomial QMLE (NBQMLE, Aknouche et *al*, 2018), and the Exponential QMLE (EQMLE, Aknouche and Francq, 2021) are the most commonly used for INGARCH-type models. However, their asymptotic efficiency depends on the model variance-to-mean relationship, which is denoted by $\vartheta_t = g(\mu_t)$, where $g(.)$ is a positive real function. More precisely, each one of the PQMLE, NBQMLE, and EQMLE is asymptotically optimal whenever $g(x) = ax$, $g(x) = bx(1 + cx)$, and $g(x) = dx^2$, respectively, for all $a, b, c, d > 0$ (e.g. Gouriéroux et *al*, 1984; Wooldridge, 1999). Note that the form of the variance-to-mean function (2.10)-(2.11) is unusual, so the above three QMLEs cannot be asymptotically efficient. Therefore, an appropriate weighted least squares estimate, weighted by a consistent estimate of the conditional variance, would be asymptotically dominant (Aknouche and Francq, 2022a).

The purpose of this section is two-fold: Firstly, a three-stage weighted least squares estimate for the set of parameters $(\theta_0', \sigma_0^2)'$ is introduced and; secondly, its asymptotic properties are derived. The procedure is described as follows: i) Initially, the conditional mean parameter θ_0 is estimated using a WLSE with as weight the conditional variance arbitrarily evaluated. ii) Then, the variance of the innovation σ_0^2 is estimated based on the first-stage estimate. iii) Finally, a WLSE is again utilized to estimate θ_0, but with as weight, the estimated conditional variance. Note that the conditional variance as weight in the first stage allows to get consistency of the WLSE with minimal conditions on the moments, while in the third stage, the estimated conditional variance weight is set to improve the asymptotic efficiency of the WLSE.

For any generic $\theta := \left(\omega, \alpha_1, ..., \alpha_q, \beta_1, ..., \beta_p\right)' \in \Theta$, define the generic conditional mean $\mu_t(\theta)$ as

$$\mu_t(\theta) = 1 + \omega m + \sum_{i=1}^{q} \alpha_i Y_{t-i} + \sum_{j=1}^{p} \beta_j \mu_{t-j}(\theta), \; t \in \mathbb{Z}. \tag{3.1}$$

Note that $\mu_t(\theta_0) = \mu_t$, and the process $\{\mu_t(\theta), t \in \mathbb{Z}\}$ exists a.s. whenever

$$\sum_{j=1}^{p} \beta_j < 1, \tag{3.2}$$

for all $\theta \in \Theta$. Now for any fixed positive starting values $\widetilde{\mu}_0(\theta), ..., \widetilde{\mu}_{1-p}(\theta)$,$Y_0, ..., Y_{1-q}$, let $\widetilde{\mu}_t(\theta)$ be an observable counterpart of $\mu_t(\theta)$ given by the recursion

$$\widetilde{\mu}_t(\theta) = 1 + \omega m + \sum_{i=1}^{q} \alpha_i Y_{t-i} + \sum_{j=1}^{p} \beta_j \widetilde{\mu}_{t-j}(\theta), \; t \geq 0. \tag{3.3}$$

Recall that the Poisson QMLE $\widehat{\theta}_P$ (cf. Christou and Fokianos, 2014; Ahmad and Francq, 2016), the negative binomial QMLE $\widehat{\theta}_{NB}$ (Aknouche et *al*, 2018), and exponential QMLE $\widehat{\theta}_E$ (Aknouche and Francq, 2021), are respectively given by

$$\begin{aligned}
\widehat{\theta}_P &= \arg\max_{\theta \in \Theta} \tfrac{1}{n} \sum_{t=1}^{n} Y_t \log\left(\widetilde{\mu}_t(\theta)\right) - \widetilde{\mu}_t(\theta) \\
\widehat{\theta}_{NB} &= \arg\max_{\theta \in \Theta} \sum_{t=q+1}^{n} Y_t \log\left(\tfrac{\widetilde{\mu}_t(\theta)}{r + \widetilde{\mu}_t(\theta)}\right) - r \log\left(r + \widetilde{\mu}_t(\theta)\right), \text{ for some fixed } r > 0 \\
\widehat{\theta}_E &= \arg\min_{\theta \in \Theta} \sum_{t=q+1}^{n} \tfrac{Y_t}{\widetilde{\mu}_t(\theta)} + \log \widetilde{\mu}_t(\theta).
\end{aligned}$$

To describe the 2SWLSE, we need to define the conditional variance function, similarly to (3.1) and (3.3). For all $(\theta', \sigma^2)' \in \Theta \times \Delta$, define the generic conditional variance sequence $\{\vartheta_t(\theta, \sigma^2), t \in \mathbb{Z}\}$ as

$$\vartheta_t\left(\theta, \sigma^2\right) = \left(\sigma^2 + 1\right) v_t(\theta) + \sigma^2 \mu_t^2(\theta), \; t \in \mathbb{Z}, \tag{3.4a}$$

where

$$v_t(\theta) = \omega(1 - \omega) m + \sum_{i=1}^{q} \alpha_i (1 - \alpha_i) Y_{t-i} + \sum_{j=1}^{p} \beta_j \left(1 - \beta_j\right) \mu_{t-j}(\theta) + \sum_{j=1}^{p} \beta_j^2 v_{t-j}(\theta). \tag{3.4b}$$

The latter exists provided that for all $\theta \in \Theta$

$$\sum_{j=1}^{p} \beta_j^2 < 1. \tag{3.5}$$

Note that (3.5) is implied by (3.2), and $\vartheta_t = \vartheta_t(\theta_0, \sigma_0^2)$ and $v_t = v_t(\theta_0)$. Now, for fixed starting values $\widetilde{\mu}_0(\theta), ..., \widetilde{\mu}_{1-p}(\theta)$, $Y_0, ..., Y_{1-q}$, and $\widetilde{v}_0(\theta), ..., \widetilde{v}_{1-p}(\theta)$, let $\widetilde{\vartheta}_t(\theta, \sigma^2)$ and $\widetilde{v}_t(\theta)$ be observable proxies for $\vartheta_t(\theta, \sigma^2)$ and $v_t(\theta)$, respectively, which are defined for all $t \geq 0$ by

$$\widetilde{\vartheta}_t\left(\theta, \sigma^2\right) = \left(\sigma^2 + 1\right) \widetilde{v}_t(\theta) + \sigma^2 \widetilde{\mu}_t^2(\theta) \tag{3.6a}$$

$$\widetilde{v}_t(\theta) = \omega(1-\omega) m + \sum_{i=1}^{q} \alpha_i(1-\alpha_i) Y_{t-i} + \sum_{j=1}^{p} \beta_j\left(1-\beta_j\right) \widetilde{\mu}_{t-j}(\theta) + \sum_{j=1}^{p} \beta_j^2 \widetilde{v}_{t-j}(\theta). \tag{3.6b}$$

For an arbitrary fixed weighting point, say $(\theta_*', \sigma_*^2)' \in \Theta \times \Delta$, a first-stage weighted least square estimator (1WLSE) is defined as

$$\widehat{\theta}_{1W} := \arg\min_{\theta \in \Theta} \widetilde{L}_n(\theta, \widetilde{\vartheta}^*), \tag{3.7a}$$

where

$$\widetilde{L}_n(\theta, \widetilde{\vartheta}^*) = \tfrac{1}{n} \sum_{t=1}^{n} \widetilde{l}_t(\theta, \widetilde{\vartheta}_t^*) \quad \text{with} \quad \widetilde{l}_t(\theta, \widetilde{\vartheta}_t^*) = \tfrac{(Y_t - \widetilde{\mu}_t(\theta))^2}{\widetilde{\vartheta}_t(\theta_*, \sigma_*^2)}. \tag{3.7b}$$

Note that by replacing $\widetilde{\vartheta}_t^* := \widetilde{\vartheta}_t(\theta_*, \sigma_*^2)$ by one, the resulting unweighted least squares estimator coincides with the conventional Conditional Least Squares estimator (CLSE) of Klimko and Nelson (1978). Aknouche and Francq (2022*a*) proposed general regularity conditions for the consistency and asymptotic normality property of the WLSEs for a more general framework. Thus, we adapt here their general assumptions to model (2.1). Define the information matrices

$$I\left(\theta_0, \theta^*, \sigma_*^2\right) := E\left(\tfrac{\vartheta_t\left(\theta_0, \sigma_0^2\right)}{\vartheta_t^2(\theta^*, \sigma_*^2)} \tfrac{\partial \mu_t(\theta_0)}{\partial \theta} \tfrac{\partial \mu_t(\theta_0)}{\partial \theta'}\right) \text{ and } J\left(\theta_0, \theta^*, \sigma_*^2\right) := E\left(\tfrac{1}{\vartheta_t(\theta^*, \sigma_*^2)} \tfrac{\partial \mu_t(\theta_0)}{\partial \theta} \tfrac{\partial \mu_t(\theta_0)}{\partial \theta'}\right), \tag{3.8}$$

and consider the following assumptions:

A1 Θ is compact, and conditions (2.12) and (3.2) hold true.

A2 The polynomials $\alpha_0(z) = \sum_{i=1}^{q} \alpha_{0i} z^i$ and $\beta_0(z) = 1 - \sum_{j=1}^{p} \beta_{0j} z^j$ have no common root, $\alpha_0(1) \neq 0$, and $\alpha_{0q} + \beta_{0p} \neq 0$.

A3 θ_0 belongs to the interior of Θ.

Note that (2.12) entails (2.6) and $E\left(Y_t^2\right) < \infty$. Consistency and asymptotic normality of $\widehat{\theta}_{1W}$ are established in the next result.

Theorem 3.1 *Under the assumptions* **A1-A2**

$$\widehat{\theta}_{1W} \underset{n\to\infty}{\overset{a.s.}{\to}} \theta_0. \tag{3.9}$$

If, in addition, **A3** *is satisfied then*

$$\sqrt{n}\left(\widehat{\theta}_{1W} - \theta_0\right) \underset{n\to\infty}{\overset{\mathcal{D}}{\to}} \mathcal{N}\left(0, \Sigma\right), \qquad \Sigma = J^{-1}\left(\theta_0, \theta^*, \sigma_*^2\right) I\left(\theta_0, \theta^*, \sigma_*^2\right) J^{-1}\left(\theta_0, \theta^*, \sigma_*^2\right). \tag{3.10}$$

We now estimate the innovation variance σ_0^2. Let $\epsilon_t = \left(Y_t - \mu_t\right)^2 - \vartheta_t$. From (2.11) we obtain the following non-estimable regression

$$\left(Y_t - \mu_t\right)^2 = v_t + \left(v_t + \mu_t^2\right)\sigma_0^2 + \epsilon_t$$

or equivalently

$$\frac{(Y_t-\mu_t(\theta_0))^2 - v_t(\theta_0)}{v_t(\theta_0)+\mu_t^2(\theta_0)} = \sigma_0^2 + \tau_t, \tag{3.11}$$

where $\tau_t = \frac{1}{v_t+\mu_t^2}\epsilon_t$ is a term of a martingale difference. In order to obtain an estimable regression corresponding to (3.11), we replace $\mu_t\left(\theta_0\right)$ and $v_t\left(\theta_0\right)$ in (3.11) by their consistent estimates $\widehat{\mu}_t = \widetilde{\mu}_t(\widehat{\theta}_{1W})$ and $\widehat{v}_t = \widetilde{v}_t(\widehat{\theta}_{1W})$, respectively, where the latter is obtained recursively by

$$\widehat{v}_t = \widehat{\omega}(1-\widehat{\omega})m + \sum_{i=1}^{q} \widehat{\alpha}_i\left(1-\widehat{\alpha}_i\right)Y_{t-i} + \sum_{j=1}^{p}\left(\widehat{\beta}_j(1-\widehat{\beta}_j)\widehat{\mu}_{t-j} + \widehat{\beta}_j^2\widehat{v}_{t-j}\right), \ t \geq 1, \tag{3.12}$$

under the above starting values. From the resulting regression

$$\frac{(Y_t-\widehat{\mu}_t)^2 - \widehat{v}_t}{\widehat{v}_t+\widehat{\mu}_t^2} = \sigma_0^2 + \widehat{\tau}_t,$$

the (weighted) least squares estimate of σ_0^2 is

$$\widehat{\sigma}_n^2 = \frac{1}{n}\sum_{t=1}^{n} \frac{(Y_t-\widehat{\mu}_t)^2 - \widehat{v}_t}{\widehat{v}_t+\widehat{\mu}_t^2}. \tag{3.13}$$

Consistency and asymptotic normality of $\widehat{\sigma}_n^2$ are now established.

Theorem 3.2 *Under* **A1-A3**

$$\widehat{\sigma}_n^2 \underset{n\to\infty}{\overset{a.s.}{\to}} \sigma_0^2. \tag{3.14}$$

If, in addition,

$$\Lambda := E\left(\frac{(Y_t-\mu_t(\theta_0))^2-\left(v_t(\theta_0)+\left(v_t(\theta_0)+\mu_t^2(\theta_0)\right)\sigma_0^2\right)}{v_t(\theta_0)+\mu_t^2(\theta_0)}\right)^2 < \infty, \tag{3.15}$$

then

$$\sqrt{n}\left(\widehat{\sigma}_n^2-\sigma_0^2\right) \underset{n\to\infty}{\overset{\mathcal{D}}{\to}} \mathcal{N}\left(0,\Lambda\right). \tag{3.16}$$

A consistent estimate of Λ is

$$\widehat{\Lambda}_n = \frac{1}{n}\sum_{t=1}^{n}\left(\frac{(Y_t-\widehat{\mu}_t)^2-\left(\widehat{v}_t+\left(\widehat{v}_t+\widehat{\mu}_t^2\right)\widehat{\sigma}_n^2\right)}{\widehat{v}_t+\widehat{\mu}_t^2}\right)^2.$$

Moreover, consistent estimates of $I\left(\theta_0,\theta^*,\sigma_*^2\right)$ and $J\left(\theta_0,\theta^*,\sigma_*^2\right)$ are respectively $\widehat{I}_n^*$ and $\widehat{J}_n^*$, where

$$\begin{aligned}\widehat{I}_n^* &= \frac{1}{n}\sum_{t=1}^{n}\frac{\widetilde{\vartheta}_t(\widehat{\theta}_{1W},\widehat{\sigma}_n^2)}{\widetilde{\vartheta}_t^2(\theta^*,\sigma_*^2)}\frac{\partial\widetilde{\mu}_t(\widehat{\theta}_{1W})}{\partial\theta}\frac{\partial\widetilde{\mu}_t(\widehat{\theta}_{1W})}{\partial\theta'}\\ \widehat{J}_n^* &= \frac{1}{n}\sum_{t=1}^{n}\frac{1}{\widetilde{\vartheta}_t(\theta^*,\sigma_*^2)}\frac{\partial\widetilde{\mu}_t(\widehat{\theta}_{1W})}{\partial\theta}\frac{\partial\widetilde{\mu}_t(\widehat{\theta}_{1W})}{\partial\theta'}.\end{aligned}$$

Having consistent estimates of θ_0 and σ_0^2, the optimal weight $\vartheta_t := \vartheta_t\left(\theta_0,\sigma_0^2\right)$ is estimated by $\widehat{\vartheta}_t := \widetilde{\vartheta}_t(\widehat{\theta}_{1W},\widehat{\sigma}_n^2)$ and a second stage WLS estimate (2WLSE) of θ_0 has the following form

$$\widehat{\theta}_{2W} := \arg\min_{\theta\in\Theta}\widetilde{L}_n(\theta,\widehat{\vartheta}), \tag{3.17}$$

where

$$\widetilde{L}_n(\theta,\widehat{\vartheta}) = \frac{1}{n}\sum_{t=1}^{n}\widetilde{l}_t(\theta,\widehat{\vartheta}_t) \quad \text{with} \quad \widetilde{l}_t(\theta,\widehat{\vartheta}_t) = \frac{(Y_t-\widetilde{\mu}_t(\theta))^2}{\widetilde{\vartheta}_t(\widehat{\theta}_{1W},\widehat{\sigma}_n^2)}.$$

The following result establishes the consistency and asymptotic normality of $\widehat{\theta}_{2W}$.

Theorem 3.3 *Under* **A1-A2**

$$\widehat{\theta}_{2W} \underset{n\to\infty}{\overset{a.s.}{\to}} \theta_0. \tag{3.18}$$

If, in addition, **A3** *is satisfied then*

$$\sqrt{n}\left(\widehat{\theta}_{2W}-\theta_0\right) \underset{n\to\infty}{\overset{\mathcal{D}}{\to}} \mathcal{N}\left(0,J^{-1}\left(\theta_0,\theta_0,\sigma_0^2\right)\right), \tag{3.19}$$

where

$$J\left(\theta_0,\theta_0,\sigma_0^2\right) := E\left(\frac{1}{\vartheta_t(\theta_0,\sigma_0^2)}\frac{\partial\mu_t(\theta_0)}{\partial\theta}\frac{\partial\mu_t(\theta_0)}{\partial\theta\prime}\right).$$

From (3.19), it turns out that the second-stage estimate $\widehat{\theta}_{2W}$ is asymptotically more efficient than $\widehat{\theta}_{1W}$. Moreover, $\widehat{\theta}_{2W}$ is also more efficient than the Poisson, negative binomial, and exponential QMLEs.

Finally, a consistent estimate of the optimal matrix $J\left(\theta_0,\theta_0,\sigma_0^2\right)$ involved in (3.19) is

$$\widehat{J}_n = \frac{1}{n}\sum_{t=1}^{n}\frac{1}{\widetilde{\vartheta}_t(\widehat{\theta}_{2W},\widehat{\sigma}_n^2)}\frac{\partial\widetilde{\mu}_t(\widehat{\theta}_{2W})}{\partial\theta}\frac{\partial\widetilde{\mu}_t(\widehat{\theta}_{2W})}{\partial\theta'}. \quad (3.20)$$

In the special MthINARCH(q) model corresponding to (2.1) with $\beta_{0j}=0$ $(1\leq j\leq p)$, the conditional mean function $\mu_t(\theta)=\theta'\chi_t$ is linear with respect to the parameter θ, where $\chi_t=(m,Y_{t-1},...,Y_{t-q})'$. Therefore, a numerical advantage of the WLSEs (given by (3.7) and (3.17)) over the QMLEs is that the formers can be given explicitly, without any optimization routine, as follows

$$\widehat{\theta}_{1W} = \left(\sum_{t=1}^{n}\frac{\chi_t\chi_t'}{\widetilde{\vartheta}_t(\theta_*,\sigma_*^2)}\right)^{-1}\sum_{t=1}^{n}\frac{\chi_t(Y_t-1)}{\widetilde{\vartheta}_t(\theta_*,\sigma_*^2)} \quad (3.21a)$$

$$\widehat{\theta}_{2W} = \left(\sum_{t=1}^{n}\frac{\chi_t\chi_t'}{\widetilde{\vartheta}_t(\widehat{\theta}_{1W},\widehat{\sigma}_n^2)}\right)^{-1}\sum_{t=1}^{n}\frac{\chi_t(Y_t-1)}{\widetilde{\vartheta}_t(\widehat{\theta}_{1W},\widehat{\sigma}_n^2)}. \quad (3.21b)$$

See also Aknouche (2012) and Aknouche and Francq (2022a) for similar models.

4 Simulation study

In this section, we assess the finite-sample properties of the 2SWLSE under the MthINGARCH model (2.1) via a simulation study. In particular, the empirical distribution of the 2WLSE over 1000 MthINGARCH(p,q) replications is compared with those of three QMLEs (PQMLE, NBQMLE, and EQMLE) and two WLSEs (the CLSE corresponding to $w=1$ and the first stage 1WLSE). The NBQMLE is computed using the profile dispersion parameter $r=1$. Two instances of (2.1) are considered, namely the MthINGARCH$(2,1)$ model with $m=4$, $\theta_0=(0.7,0.3,0.2,0.2)'$, and $\varepsilon_t\sim\mathcal{P}(1)$ so that $\sigma_0^2=1$ (cf. Tables 4.1-4.2), and the MthINGARCH$(1,1)$ model with $m=6$, $\theta_0=(0.3,0.4,0.2)'$, and negative binomial distributed innovation, $\varepsilon_t \sim \mathcal{NB}(1,\frac{1}{2})$, so $\sigma_0^2=2$ (cf. Table 4.3). A third exercise on the pure MthINARCH(2) ($p=0$) model with $m=10$, $\theta_0=(0.7,0.5,0.3)'$, and $\varepsilon_t \sim \mathcal{P}(1)$ is available in the supplementary material. The parameters are chosen so that the corresponding models are stationary with finite variance (i.e. the condition (2.12) is satisfied). For each scenario, the sample sizes, 600, and 1000, are considered. Additional simulations for the two aforementioned models with $n=300$ can be found in the supplementary document. In computing all the estimates, the initial parameter value for the optimization routines is $\theta^{(0)}=0.7\theta_0$ and the weighting parameters for the first stage WLSE are set to $\theta_w=0.3\theta_0$ and $\sigma_w^2=0.3\sigma_0^2$, respectively.

Tables 4.1-4.4 show the means, standard deviations (StD), asymptotic standard errors (ASE), and root mean square errors (RMSE) of the six estimates over the 1000 replications. The ASEs of the least squares based estimates (CLSE, 1WLSE, 2WLSE) are obtained from Theorems 3.1-3.3, while those from the QMLEs (Poisson, negative binomial, and exponential) are obtained from Ahmad and Francq (2016) for the PQMLE, from Aknouche et *al* (2018) for the NBQMLE, and from Aknouche and Francq (2021) for the the EQMLE. The RMSE of an estimate $\widehat{\theta}$ is obtained from the formula RMSE$=\sqrt{\text{Bias}^2+\text{StD}^2}$, where Bias represents the sample mean of $\widehat{\theta}-\theta_0$ over the 1000 replications.

n 600		PQMLE	NBQMLE	EQMLE	CLSE	1WLSE	2WLSE
	Mean	0.7130	0.7091	0.7092	0.7942	0.7045	0.7213
ω_0	StD	0.2239	0.2025	0.2037	0.3239	0.2069	0.1950
0.7	ASE	0.2100	0.1775	0.1775	0.3874	0.1849	0.1796
	RMSE	0.2243	0.2027	0.2039	0.3373	0.2070	0.1962
	Mean	0.3023	0.3036	0.3035	0.2851	0.3048	0.3030
α_{01}	StD	0.0572	0.0490	0.0490	0.0490	0.0857	0.0486
0.3	ASE	0.0602	0.0503	0.0504	0.0978	0.0520	0.0504
	RMSE	0.0573	0.0491	0.0491	0.0870	0.0505	0.0487
	Mean	0.1897	0.1961	0.1960	0.1712	0.1974	0.1969
α_{02}	StD	0.0837	0.0725	0.0724	0.1145	0.0751	0.0692
0.2	ASE	0.1559	0.1341	0.1342	0.2665	0.1380	0.1346
	RMSE	0.0843	0.0726	0.0725	0.1181	0.0751	0.0692
	Mean	0.1999	0.1939	0.1940	0.2108	0.1938	0.1888
β_{01}	StD	0.1318	0.1229	0.1236	0.1742	0.1225	0.1165
0.2	ASE	0.1152	0.0983	0.0983	0.2076	0.1014	0.0983
	RMSE	0.1319	0.1230	0.1237	0.1746	0.1226	0.1170
	Mean	1.0144	1.0073	1.0074	1.0366	1.0106	1.0084
σ_0^2	StD	0.0761	0.0747	0.0748	0.1352	0.0731	0.0746
1	ASE	0.0929	0.0910	0.0909	0.1020	0.0918	0.0909
	RMSE	0.0775	0.0751	0.0752	0.1401	0.0739	0.0750

Table 4.1. QML and WLS estimation results for the MthINGARCH$(2,1)$ series with $n = 600$, $m = 4$, $\theta_0 = (0.7, 0.3, 0.2, 0.2)'$, $\varepsilon_t \sim \mathcal{P}(1)$, and $\sigma_0^2 = 1$.

n 1000		PQMLE	NBQMLE	EQMLE	CLSE	1WLSE	2WLSE
	Mean	0.7307	0.7138	0.7143	0.8274	0.7093	0.7249
ω_0	StD	0.1998	0.1704	0.1705	0.2904	0.1858	0.1689
0.7	ASE	0.1669	0.1362	0.1364	0.3535	0.1418	0.1372
	RMSE	0.2021	0.1709	0.1711	0.3171	0.1860	0.1707
	Mean	0.3057	0.3077	0.3078	0.2935	0.3071	0.3074
α_{01}	StD	0.0524	0.0445	0.0444	0.0762	0.0455	0.0442
0.3	ASE	0.0471	0.0391	0.0392	0.0808	0.0403	0.0390
	RMSE	0.0527	0.0452	0.0451	0.0765	0.0460	0.0448
	Mean	0.1968	0.2016	0.2021	0.1883	0.2001	0.2029
α_{02}	StD	0.0619	0.0572	0.0571	0.0885	0.0598	0.0560
0.2	ASE	0.1195	0.0996	0.0996	0.2250	0.1038	0.1001
	RMSE	0.0620	0.0573	0.0572	0.0893	0.0598	0.0561
	Mean	0.1835	0.1837	0.1830	0.1734	0.1872	0.1784
β_{01}	StD	0.0997	0.0889	0.0891	0.1500	0.0965	0.0878
0.2	ASE	0.0855	0.0733	0.0732	0.1693	0.0758	0.0737
	RMSE	0.1011	0.0904	0.0907	0.1524	0.0973	0.0904
	Mean	1.0179	1.0126	1.0127	1.0400	1.0141	1.0141
σ_0^2	StD	0.0666	0.0655	0.0656	0.1158	0.0652	0.0655
1	ASE	0.0707	0.0701	0.0700	0.0789	0.0703	0.0701
	RMSE	0.0690	0.0667	0.0668	0.1225	0.0667	0.0670

Table 4.2. QML and WLS estimation results for the MthINGARCH$(2,1)$ series with $n = 1000$, $m = 4$, $\theta_0 = (0.7, 0.3, 0.2, 0.2)'$, $\varepsilon_t \sim \mathcal{P}(1)$, and $\sigma_0^2 = 1$.

n 600		PQMLE	NBQMLE	EQMLE	CLSE	1WLSE	2WLSE
	Mean	0.3213	0.3061	0.3057	0.3727	0.2997	0.3133
ω	StD	0.1180	0.1082	0.1085	0.1530	0.1143	0.1061
0.3	ASE	0.0971	0.0870	0.0872	0.1814	0.0906	0.0880
	RMSE	0.1199	0.1084	0.1086	0.1694	0.1143	0.1069
	Mean	0.3888	0.3981	0.3983	0.3540	0.3951	0.3967
α_1	StD	0.0863	0.0773	0.0775	0.1176	0.0787	0.0770
0.4	ASE	0.0818	0.0718	0.0720	0.1146	0.0735	0.0718
	RMSE	0.0870	0.0773	0.0775	0.1262	0.0788	0.0771
	Mean	0.1896	0.1971	0.1975	0.1810	0.2056	0.1910
β_1	StD	0.1232	0.1151	0.1155	0.1649	0.1229	0.1119
0.2	ASE	0.0980	0.0883	0.0886	0.1750	0.0917	0.0890
	RMSE	0.1237	0.1152	0.1156	0.1660	0.1230	0.1123
	Mean	2.0182	2.0030	2.0027	2.0529	2.0162	2.0046
σ_0^2	StD	0.2028	0.2015	0.2017	0.3586	0.2052	0.2016
2	ASE	0.3151	0.3134	0.3132	0.3353	0.3190	0.3130
	RMSE	0.2036	0.2016	0.2017	0.3625	0.2058	0.2016

Table 4.3. QML and WLS estimation results for MthINGARCH$(1,1)$ series with $n = 600$, $m = 6$, $\theta_0 = (0.3, 0.4, 0.2)'$, $\varepsilon_t \sim \mathcal{NB}(1, \frac{1}{2})$, and $\sigma_0^2 = 2$.

n 1000		PQMLE	NBQMLE	EQMLE	CLSE	1WLSE	2WLSE
	Mean	0.3099	0.2926	0.2915	0.3759	0.2997	0.2979
ω	StD	0.0858	0.0820	0.0826	0.1327	0.0843	0.0808
0.3	ASE	0.0757	0.0656	0.0657	0.1725	0.0730	0.0655
	RMSE	0.0863	0.0824	0.0831	0.1529	0.0843	0.0808
	Mean	0.3985	0.4029	0.4027	0.3692	0.4022	0.4023
α_1	StD	0.0614	0.0584	0.0587	0.0979	0.0575	0.0587
0.4	ASE	0.0684	0.0564	0.0564	0.1098	0.0583	0.0563
	RMSE	0.0614	0.0585	0.0587	0.1027	0.0575	0.0588
	Mean	0.1882	0.2022	0.2035	0.1603	0.1955	0.1971
β_1	StD	0.0817	0.0783	0.0788	0.1308	0.0822	0.0765
0.2	ASE	0.0773	0.0689	0.0691	0.1556	0.0752	0.0685
	RMSE	0.0826	0.0783	0.0789	0.1367	0.0823	0.0765
	Mean	2.0014	1.9920	1.9917	2.0560	1.9982	1.9933
σ_0^2	StD	0.1581	0.1567	0.1565	0.3865	0.1560	0.1563
2	ASE	0.2347	0.2330	0.2330	0.2825	0.2341	0.2329
	RMSE	0.1581	0.1569	0.1567	0.3906	0.1560	0.1564

Table 4.4. QML and WLS estimation results for MthINGARCH$(1,1)$ series with $n = 1000,\ m = 6$, $\theta_0 = (0.3, 0.4, 0.2)'$, $\varepsilon_t \sim \mathcal{NB}(1, \frac{1}{2})$, and $\sigma_0^2 = 2$.

Several conclusions can be drawn from the results in Tables 4.1-4.4. Firstly, the six estimation methods for θ_0 provide quite good results with low biases and dispersions. Secondly, except for the CLSE, which provides the worst results in terms of bias, ASE, and RMSE, the remaining five estimates are comparable, with a slight superiority of the 2SWLSE, especially for the RMSEs and ASEs. Furthermore, the results are consistent with the asymptotic theory, since the larger the sample size the smaller the RMSEs and ASEs. Moreover, the superiority of 2WLSE is more pronounced for large samples being the ASEs and StD quite close. Fourthly, the 1WLSE, although arbitrarily weighted, gives equally accurate results which are close to those of the EQMLE and NBQMLE. The latter are quite similar and are slightly better than the PQMLE. Finally, the estimates of σ_0^2 are good, which validate the conditional variance formulas (see proposition 2.2), on which is based $\widehat{\sigma}_n^2$ given by (3.13). Likewise, for $\widehat{\sigma}_n^2$, the StDs are overall close to the ASEs.

5 Empirical examples

5.1 Ecoli data

The first application concerns the weekly number of Ecoli (Escherichia coli) diseases in the state of North Rhine-Westphalia (Germany). The data, which is taken from the tscount R package of Liboschik et *al* (2017) spans from January 2001 to May 2013, with a total of $n = 646$ observations (cf. Figure 5.1). The corresponding sample mean and sample variance are respectively 20.3344, and 88.7531, which shows a quite large overdispersion. In fact, the Ecoli series has been recently considered in several research papers as an example of a time series exhibiting simultaneously high overdispersion (Silva and Barreto-Souza, 2019), heavy-tailedness (Doukhan et *al*, 2021) and multimodality (Doukhan et *al*, 2021; Mao et *al*, 2020).

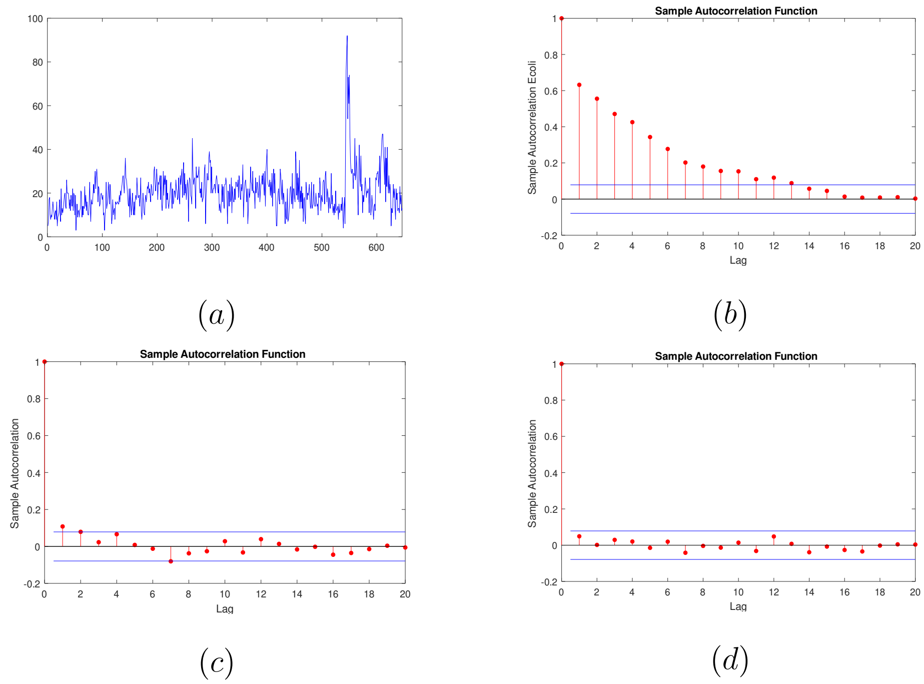


Figure 5.1. (a) The Ecoli series, (b) sample correlation of Ecoli, (c) sample correlation of the raw residuals, (d) sample correlation of the Pearson residuals.

The MthINGARCH$(1,1)$ is fitted through the six methods proposed in Section 4. The choice of $p = q = 1$ is made because when $p > 1$ or $q > 1$ the corresponding parameter estimates are not statistically significant. Moreover, the parameter m is fixed as being the upper integer part of the sample mean, that is $m = 21$. Note that m cannot be estimated because it cannot be identified. We also estimate MthINGARCH$(1,1)$ models for adjacent values of 21 ($m = 20$, $m = 22$), and it turns out that the results were almost identical. When m is set to be far from the integer mean (e.g. $m = 10$ or $m = 30$), the results are poor with high ASEs of estimates and significantly correlated residuals. For all estimates, the starting values are $\theta^{(0)} = (0.2, 0.3, 0.2)'$. Moreover, the weighting parameters are set to $\theta^{(w)} = \theta^{(0)}$ and $\sigma^{2(w)} = 1$. Parameter estimates $\widehat{\theta}_n = \left(\widehat{\omega}, \widehat{\alpha}_1, \widehat{\beta}_1\right)'$ and $\widehat{\sigma}_n^2$, and their ASEs in parenthesis are shown in Table 5.1 for the six methods. Table 5.1 also shows the corresponding estimated theoretic means $\widehat{\mu} = \frac{1+\widehat{\omega}m}{1-\left(\widehat{\alpha}_1+\widehat{\beta}_1\right)}$ and the estimated unconditional variances $\widehat{V}$ given by (2.15), where the true parameter θ_0 is replaced by its estimate $\widehat{\theta}$. Note that all parameter estimates are significant with quite small ASEs, and are inside the second-order stationarity domain. Moreover, all estimated means $\widehat{\mu}$ are very close to the sample mean, being the 2WLSE the nearest one with value 20.3364. The estimated variances remain reasonably close to the sample variance. Furthermore, the estimates $\widehat{\alpha}_1 + \widehat{\beta}_1$ indicate strong persistence as the corresponding values are close to 0.9. We then obtain the raw and Pearson residuals defined, respectively, by $Y_t - \widehat{\mu}_t$ and $\frac{Y_t - \widehat{\mu}_t}{\sqrt{\widehat{\vartheta}_t}}$, where $\widehat{\mu}_t = \widetilde{\mu}_t(\widehat{\theta}_{2W})$ and $\widehat{\vartheta}_t = \widetilde{\vartheta}_t(\widehat{\theta}_{2W})$. Figure 5.1 displays the sample autocorrelation of the raw and Pearson residuals computed only from the 2WLSE. The residuals obtained from the other methods are available upon request to the authors. It can be seen that the white noise assumption is tenable for the two series of residuals. Furthermore, there is no evidence of any significant correlation. For all six methods, Table 5.1 also displays the mean absolute residual (MAR) defined as MAR:$= \frac{1}{n}\sum_{t=1}^{n} |Y_t - \widehat{\mu}_t|$, and the mean square Pearson residual, MSPR:$= \frac{1}{n}\sum_{t=1}^{n} \frac{(Y_t - \widehat{\mu}_t)^2}{\widehat{\vartheta}_t}$. Note that all MSPRs are smaller than the ones in Doukhan et *al* (2021) (with value 1.13) and Mao et *al* (2020) (with value 1.01) which are obtained form mixture (Poisson) INARCH model and mixture (negative binomial) INGARCH model, respectively. For completeness, we also considered the Ecoli series without the two first observations as considered by Doukhan et *al* (2021) and the results are quite similar. Note that the MSPR corresponding to the 2WLSE is the closest value to one, and with respect to the MAR, the best value is the one given by the EQMLE.

	PQMLE	NBQMLE	EQMLE	CLSE	1WLSE	2WLSE
$\widehat{\omega}$	0.0804 (0.0349)	0.0709 (0.0328)	0.0705 (0.0328)	0.0853 (0.0438)	0.0674 (0.0295)	0.0746 (0.03392)
$\widehat{\alpha}_1$	0.3724 (0.0675)	0.3222 (0.0542)	0.3205 (0.0538)	0.4498 (0.0949)	0.3134 (0.0460)	0.3406 (0.0598)
$\widehat{\beta}_1$	0.4963 (0.0883)	0.5551 (0.0749)	0.5571 (0.0745)	0.4139 (0.1143)	0.5673 (0.0631)	0.5331 (0.0816)
$\widehat{\sigma}_n^2$	0.0722 (0.0001)	0.0705 (0.0001)	0.0704 (0.0001)	0.0786 (0.0001)	0.0704 (0.0001)	0.0710 (0.0001)
$\widehat{\mu}$	20.4786	20.2798	20.2737	20.4717	20.2526	20.3364
$\widehat{V}$	101.6643	95.3997	95.2038	115.9368	95.2746	97.3320
MAR	5.1662	5.1499	5.1498	5.2083	5.1597	5.1539
MSPR	0.9985	1.0039	1.0041	0.9892	1.0047	1.0019

Table 5.1. QML and WLS estimation results for the MthINGARCH$(1,1)$ on the Ecoli series with $m = 21$.

We finally recover the mean and variance of the scaled innovation (ε_t) using the randomized scaled residuals $(\widehat{\varepsilon}_t)$ obtained from the following equations

$$\widehat{\varepsilon}_t = \tfrac{Y_t}{\widehat{\lambda}_t}, \text{ and } \widehat{\lambda}_t = 1 + \widehat{\omega} \circ m + \widehat{\alpha} \circ Y_{t-1} + \widehat{\beta} \circ \widehat{\lambda}_{t-1}, \;\; t \geq 2, \tag{5.1}$$

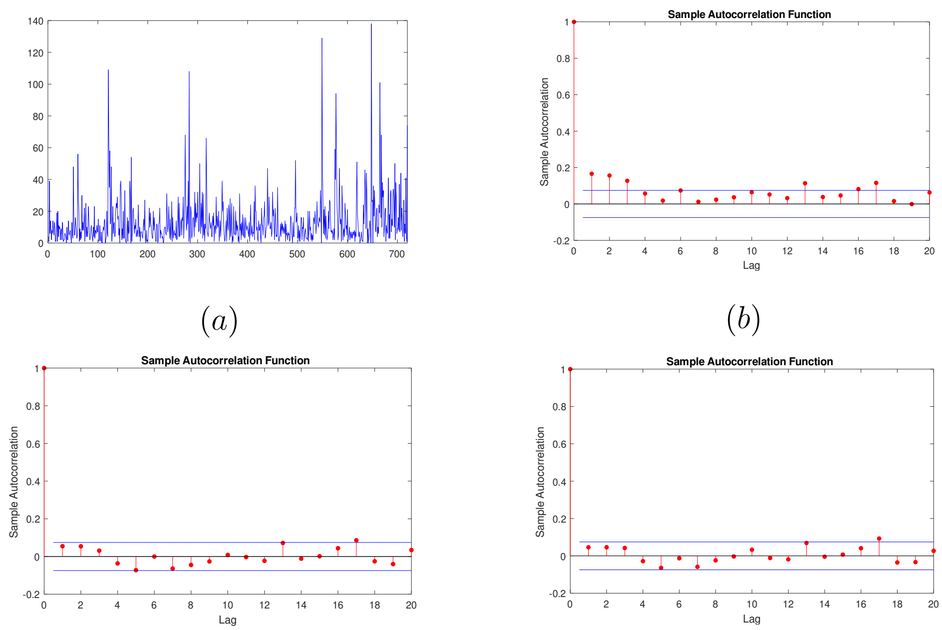
with $\widehat{\lambda}_1 = Y_1$ and $\widehat{\varepsilon}_1 = 1$. The term randomized is introduced because the random binomial variables $\widehat{\omega} \circ m$, $\widehat{\alpha} \circ Y_{t-1}$ and $\widehat{\beta} \circ \widehat{\lambda}_{t-1}$ change values at each simulation. So we generate 1000 replications from (5.1). For each replication, we compute the sample mean and sample variance of the simulated residuals $(\widehat{\varepsilon}_t)$. Then, the mean of the sample means (SMean$(\widehat{\varepsilon}_t)$) and the mean of the sample variances (SVar$(\widehat{\varepsilon}_t)$) over the 1000 replications are obtained as shown in Table 5.2. As expected, the sample means for the six methods are close to 1.

	PQMLE	NBQMLE	EQMLE	CLSE	1WLSE	2WLSE
SMean$(\widehat{\varepsilon}_t)$	1.0510	1.0536	1.0532	1.0536	1.0535	1.0530
SVar$(\widehat{\varepsilon}_t)$	0.2150	0.2176	0.2166	0.2158	0.2178	0.2162

Table 5.2. Estimated means and variances of the residuals $\widehat{\varepsilon}_t = Y_t/\widehat{\lambda}_t$ across the six estimates for the Ecoli series.

5.2 Euro-pound sterling exchange rate data

As a second application, we fit the MthINGARCH$(1,1)$ model to the number of tick changes by minute of the euro to British pound exchange rate (ExRate for short) on December 12th, from 9.00 a.m. to 9.00 p.m. This series has been taken from Gorgi (2020) to which the author fits the so-called Beta Negative Binomial (BNB) INGARCH$(1,1)$ model. The series comprises 720 observations with sample mean 13.2153 and sample variance 224.2498, showing a very high overdispersion (cf. Figure 5.2).



(a) (b)

(c) (d)

Figure 5.2. (a) The ExRate series, (b) sample correlation of ExRate, (c) sample correlation of the raw residuals, (d) sample correlation of the Pearson residuals.

For the same reasons as above, only the case $p = q = 1$ is considered and m is set to the value 14. The starting parameter values are $\theta^{(0)} = (0.9, 0.01, 0.75)'$, $\sigma^{2(w)} = 2$, and $\theta^{(w)} = \theta^{(0)}$. Table 5.2 displays means of the six estimates, their ASEs in parenthesis, the estimated theoretic means $\widehat{\mu}$, and the estimated unconditional variances $\widehat{V}$. Overall, all estimates are significant, with quite low ASEs, and lying in the second-order stationarity region. Moreover, the estimated means $\widehat{\mu}$ are quite close to the sample mean 13.2153, where the 2WLSE is the nearest one. The nearest estimated variance is the one provided by the 2WLSE with a value of 262.6562. Figure 5.2 shows the sample autocorrelation of the raw and Pearson residuals, which are consistent with the white noise assumption. It can also be observed from Table 5.3 that all mean absolute residuals (MARs) are smaller than those generated by all models considered by Gorgi (2020). In particular, the best MARs in Gorgi (2020) are 9.63 for the negative binomial GAS$(1,1)$ (Generalized Autoregressive Score) model and 9.66 for the INGARCH$(1,1)$ model. For the MthINGARCH$(1,1)$ model, the best MAR is reached by the PQMLE with a value of 9.2156, while the poorest is 9.2295 obtained from the CLSE. Table 5.2 also shows the MSPRs which are all near to 1. The best MSPR is the one generated by the NBQMLE and EQMLE. Note also that the estimates $\widehat{\omega}$ are mostly null or near to zero meaning that the value m does not influence the unconditional mean.

	PQMLE	NBQMLE	EQMLE	CLSE	1WLSE	2WLSE
$\widehat{\omega}$	0.0082 (0.0156)	0.0000 (0.0139)	0.0000 (0.0140)	0.0474 (0.0289)	0.0468 (0.0286)	0.0000 (0.0141)
$\widehat{\alpha}_1$	0.1048 (0.0334)	0.0937 (0.0317)	0.0936 (0.0318)	0.1468 (0.0410)	0.1463 (0.0399)	0.0804 (0.0310)
$\widehat{\beta}_1$	0.7958 (0.0731)	0.8320 (0.0667)	0.8320 (0.0672)	0.6357 (0.1155)	0.6380 (0.1147)	0.8444 (0.0677)
$\widehat{\sigma}_n^2$	1.1361 (0.0586)	1.1357 (0.0566)	1.1356 (0.0567)	1.1530 (0.0666)	1.1533 (0.0665)	1.1307 (0.0542)
$\widehat{\mu}$	13.3743	13.4574	13.4573	13.3188	13.3151	13.3063
$\widehat{V}$	275.6461	283.1743	283.1024	269.5907	269.6467	262.6562
MAR	9.2155	9.2198	9.2198	9.2295	9.2284	9.2163
MSPR	0.9972	0.9970	0.9970	0.9978	0.9978	0.9975

Table 5.2. QML and WLS estimation results for the MthINGARCH$(1,1)$ on the ExRate series with $m = 14$.

As described above, we obtain the randomized scaled residuals $(\widehat{\varepsilon}_t)$. The sample means and variances of the residuals oven 1000 replications of the estimates given by (5.1) are reported in Table 5.4. The sample means are quite close to 1, the theoretical mean of ε_t. The sample variances are, however, less closed to the estimated variance $\widehat{\sigma}^2 = 1$, which could be due to sampling fluctuations.

	PQMLE	NBQMLE	EQMLE	CLSE	1WLSE	2WLSE
SMean($\widehat{\varepsilon}_t$)	1.0889	1.0851	1.0892	1.0875	1.0863	1.0852
SVar($\widehat{\varepsilon}_t$)	1.8797	1.9013	1.9309	1.8579	1.8386	1.8909

Table 5.4. Estimated means and variances of the residuals $\widehat{\varepsilon}_t = Y_t/\widehat{\lambda}_t$ across the six estimates for the ExRate series.

6 Conclusion

This paper proposes a semi-parametric multiplicative thinning-based integer-valued GARCH model, driven by a sequence of iid integer-valued innovations. The model shares the same conditional mean structure as the standard INGARCH model, but its conditional variance has an unusual form, regardless of the distribution of the innovations. The MthINGARCH model is able to generate high overdispersion, high persistence, and heavy-tailedness even when the generating innovation is light-tailed and/or equidispersed. Parameter estimation was easily performed using distribution-free methods such as QMLEs and WLSEs. Furthermore, the second stage WLSE (2WLSE) is never less asymptotically efficient than any exponential family QMLE. Through two real-world applications, we have seen how the proposed model can sparsely generate interesting features without resorting to more complex models such as mixture and heavy-tailed INGARCH models.

The MthINGARCH model can be extended in several directions. Firstly, random coefficient extensions such as threshold, iid mixture, and Markov mixture can be introduced to model multimodality in count time series data. Secondly, a $\mathbb{Z}$-valued extension of the model could be introduced by assuming the innovation to be $\mathbb{Z}$-valued, while adapting the conditional mean dynamics. Thirdly, the model allows the inclusion of (integer-valued) covariates in the dynamics of λ_t. Moreover, revealing the extremal behavior of the model seems a promising area for future research. Fourthly, multivariate extensions could be easily derived. Finally, the model is unable to generate underdispersion, and certain modified forms of it might remedy this.

7 Proofs

Proof of Proposition 2.1 i) The result (2.4) is obvious noticing that $E\left(\lambda_{t-j}|\mathcal{F}_{t-1}^{Y,\lambda}\right) = \lambda_{t-j}$ for $j \geq 1$.

ii) Given $\mathcal{F}_{t-1}^{Y}$, the conditional mean μ_t writes as follows

$$\mu_t = 1 + \omega_0 m + \sum_{i=1}^{q} \alpha_{0i} Y_{t-i} + \sum_{j=1}^{p} \beta_{0j} E\left(\lambda_{t-j}|\mathcal{F}_{t-1}^{Y}\right). \qquad (A.1)$$

A recursion for calculating $E\left(\lambda_{t-j}|\mathcal{F}_{t-1}^{Y}\right)$ using $(2.1b)$ is as follows

$$\begin{aligned} E\left(\lambda_{t-j}|\mathcal{F}_{t-1}^{Y}\right) &= 1 + \omega_0 m + \sum_{i=1}^{q} \alpha_{0i} Y_{t-j-i} + \sum_{k=1}^{p} E\left(\beta_{0k} \circ \lambda_{t-j-k}|\mathcal{F}_{t-1}^{Y}\right) \\ &= 1 + \omega_0 m + \sum_{i=1}^{q} \alpha_{0i} Y_{t-j-i} + \sum_{k=1}^{p} \beta_{0k} E\left(\lambda_{t-j-k}|\mathcal{F}_{t-1}^{Y}\right). \qquad (A.2) \end{aligned}$$

In obtaining $(A.2)$, the Tower formula is used to get

$$\begin{aligned} E\left(\beta_{0k} \circ \lambda_{t-k}|\mathcal{F}_{t-1}^{Y}\right) &= E\left(E\left(Bin\left(\lambda_{t-k}, \beta_{0k}\right)|\mathcal{F}_{t-1}^{Y}\right)|\mathcal{F}_{t-1}^{Y}, \lambda_{t-k}\right) \\ &= E\left(E\left(Bin\left(\lambda_{t-k}, \beta_{0k}\right)|\mathcal{F}_{t-1}^{Y}, \lambda_{t-k}\right)|\mathcal{F}_{t-1}^{Y}\right) = \beta_{0k} E\left(\lambda_{t-k}|\mathcal{F}_{t-1}^{Y}\right). \end{aligned}$$

Letting $\mu_{t-h,t-1} := E\left(\lambda_{t-h}|\mathcal{F}_{t-1}^{Y}\right)$ then $\mu_{t,t-1} = \mu_t$, and $(A.1)$ can be rewritten

$$\mu_{t,t-1} = 1 + \omega_0 m + \sum_{i=1}^{q} \alpha_{0i} Y_{t-i} + \sum_{j=1}^{p} \beta_{0j} \mu_{t-j,t-1}, \qquad (A.3)$$

where

$$\mu_{t-j,t-1} = 1 + \omega_0 m + \sum_{i=1}^{q} \alpha_{0i} Y_{t-j-i} + \sum_{k=1}^{p} \beta_{0k} \mu_{t-j-k,t-1}.$$

Hence $(A.3)$ is in the form (2.5).

Proof of Theorem 2.1 We first show the necessity of (2.6). If there exists a stationary solution to (2.1) with $E(Y_t) < \infty$, then

$$\begin{aligned} E(Y_t) &= E(\lambda_t) \\ &= 1 + \omega_0 m + \left(\sum_{i=1}^{q} \alpha_{0i} + \sum_{j=1}^{p} \beta_{0j}\right) E(Y_t). \qquad (A.4) \end{aligned}$$

Therefore, by the non-negativity of $E(Y_t)$ and of the coefficients in (2.1b), condition (2.6) must be satisfied.

Now assume that (2.6) holds true. For all $t, k \in \mathbb{Z}$, let

$$Y_t^{(k)} = \begin{cases} \lambda_t^{(k)} \varepsilon_t & \text{if } k \geq 1 \\ 0 & \text{if } k \leq 0 \end{cases} \tag{$A.5$}$$

and

$$\lambda_t^{(k)} = \begin{cases} 1 + \omega_0 \circ m + \sum_{i=1}^{q} \alpha_{0i} \circ Y_{t-i}^{(k-i)} + \sum_{j=1}^{p} \beta_{0j} \circ \lambda_{t-j}^{(k-j)} & \text{if } k \geq 1 \\ 0 & \text{if } k \leq 0. \end{cases} \tag{$A.6$}$$

Note that $\lambda_t^{(k)}$ and $Y_t^{(k)}$ are measurable functions of terms of the iid counting sequences $(W_s)_s$, $(W_s^{(t-i,k-i)})_s$ $(W_s^{(t-j,k-j)})_s$ involved in $\omega_0 \circ m$, $\alpha_{0i} \circ Y_{t-i}^{(k-i)}$ and $\beta_{0j} \circ \lambda_{t-j}^{(k-j)}$, respectively, with obvious notation. Therefore, for any k, the sequences $\left(\lambda_t^{(k)}\right)_t$ and $\left(Y_t^{(k)}\right)_t$ are stationary and ergodic. The existence of a solution to (2.1) follows if the following intermediary result holds

$$\lambda_t = \lim_{k \to \infty} \lambda_t^{(k)} \text{ exists almost surely (a.s.) in } [0, +\infty). \tag{$A.7$}$$

Taking $k \to \infty$ in the equalities ($A.5$) and ($A.6$), and using ($A.7$), the solution of (2.1) can be written as

$$Y_t = \lim_{k \to \infty} Y_t^{(k)} = \lim_{k \to \infty} \lambda_t^{(k)} \varepsilon_t = \lambda_t \varepsilon_t \quad a.s.$$

Let us show ($A.7$). Using the same device in Latour (1998, p. 445) we have

$$\begin{aligned} E\left|\alpha_{0i} \circ Y_{t-i}^{(k-i)} - \alpha_{0i} \circ Y_{t-i}^{(k-i-1)}\right| &\leq \alpha_{0i} E\left|Y_{t-i}^{(k-i)} - Y_{t-i}^{(k-i-1)}\right| \\ &= \alpha_{0i} E\left|\lambda_{t-i}^{(k-i)} - \lambda_{t-i}^{(k-i-1)}\right|, \; 1 \leq i \leq q, \; k \geq 1 \end{aligned}$$

and

$$E\left|\beta_{0j} \circ \lambda_{t-j}^{(k-j)} - \beta_{0j} \circ \lambda_{t-j}^{(k-j-1)}\right| \leq \beta_{0j} E\left|\lambda_{t-j}^{(k-j)} - \lambda_{t-j}^{(k-j-1)}\right|, \; 1 \leq j \leq p, \; k \geq 1.$$

Since $E(\varepsilon_t) = 1$, it follows that

$$E\left|Y_t^{(k)} - Y_t^{(k-1)}\right| = E\left|\lambda_t^{(k)} - \lambda_t^{(k-1)}\right| \leq \sum_{i=1}^{r} (\alpha_{0i} + \beta_{0i}) E\left|\lambda_{t-i}^{(k-i)} - \lambda_{t-i}^{(k-i-1)}\right|. \tag{$A.8a$}$$

Letting $a^{(k)} := E\left|\lambda_t^{(k)} - \lambda_t^{(k-1)}\right|$ and $\mathbf{a}^{(k)} := \left(a^{(k)}, a^{(k-1)}, ..., a^{(k-r+1)}\right)'$, inequality $(A.8a)$ can be written in the following vector form

$$\mathbf{a}^{(k)} \leq A_0 \mathbf{a}^{(k-1)},$$

where

$$A_0 = \begin{pmatrix} \alpha_{01} + \beta_{01} & \alpha_{02} + \beta_{02} & \cdots & \alpha_{0,r-1} + \beta_{0,r-1} & \alpha_{0r} + \beta_{0r} \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}. \qquad (A.8b)$$

Denote by $\rho(A_0)$ the spectral radius of A_0, i.e. the maximum of eigenvalues of A_0 in modulus. If (2.6) is satisfied, which is equivalent to $\rho(A_0) < 1$ (e.g. Francq and Zakoian, 2019, Corollary 2.2), then in view of $(A.8b)$, $\mathbf{a}^{(k)} \to 0$ as $k \to \infty$. Therefore, $\left(\mathbf{a}^{(k)}\right)_k$ is a Cauchy sequence and thus converges in L^1 and a.s. $\square$

Proof of Proposition 2.2 i) Recursion (2.8) immediately follows from $(2.1b)$, the binomial thinning operation, and the independence of the counting series in $(2.1b)$.

ii) Equality (2.9) follows from direct calculations

$$\begin{aligned} \overline{\vartheta}_t &= E\left(\varepsilon_t^2\right) E\left(\lambda_t^2 | \mathcal{F}_{t-1}^{Y,\lambda}\right) - \overline{\mu}_t^2 = \left(\sigma_0^2 + 1\right)\left(Var\left(\lambda_t | \mathcal{F}_{t-1}^{Y,\lambda}\right) + \overline{\mu}_t^2\right) - \overline{\mu}_t^2 \\ &= \left(\sigma_0^2 + 1\right) \overline{v}_t + \sigma_0^2 \overline{\mu}_t^2. \end{aligned}$$

iii) To get a recursion for $Var\left(\lambda_t | \mathcal{F}_{t-1}^Y\right)$, we use the law of total conditional variance (cf. Bowsher and Swain, 2012)

$$Var(Y|X) = E(Var(Y|X,Z)|X) + Var(E(Y|X,Z)|X),$$

to obtain

$$\begin{aligned} Var\left(\beta_{0j} \circ \lambda_{t-j} | \mathcal{F}_{t-1}^Y\right) &= E\left(Var\left(\beta_{0j} \circ \lambda_{t-j} | \mathcal{F}_{t-1}^Y, \lambda_{t-j}\right) | \mathcal{F}_{t-1}^Y\right) + Var\left(E\left(\beta_{0j} \circ \lambda_{t-j} | \mathcal{F}_{t-1}^Y, \lambda_{t-j}\right) | \mathcal{F}_{t-1}^Y\right) \\ &= E\left(\beta_{0j}\left(1 - \beta_{0j}\right) \lambda_{t-j} | \mathcal{F}_{t-1}^Y\right) + Var\left(\beta_{0j} \lambda_{t-j} | \mathcal{F}_{t-1}^Y\right) \\ &= \beta_{0j}\left(1 - \beta_{0j}\right) E\left(\lambda_{t-j} | \mathcal{F}_{t-1}^Y\right) + \beta_{0j}^2 Var\left(\lambda_{t-j} | \mathcal{F}_{t-1}^Y\right). \end{aligned}$$

Therefore

$$\begin{aligned} v_t &= \omega_0 (1-\omega_0) m + \sum_{i=1}^{q} \alpha_{0i} (1-\alpha_{0i}) Y_{t-i} + \sum_{j=1}^{p} Var \left(\beta_{0j} \circ \lambda_{t-j} | \mathcal{F}_{t-1}^{Y} \right) \\ &= \omega_0 (1-\omega_0) m + \sum_{i=1}^{q} \alpha_{0i} (1-\alpha_{0i}) Y_{t-i} \\ &\quad + \sum_{j=1}^{p} \left(\beta_{0j} \left(1-\beta_{0j}\right) E \left(\lambda_{t-j} | \mathcal{F}_{t-1}^{Y} \right) + \beta_{0j}^2 Var \left(\lambda_{t-j} | \mathcal{F}_{t-1}^{Y} \right) \right). \end{aligned}$$

Setting $v_{t-j,t-1} := Var\left(\lambda_{t-j}|\mathcal{F}_{t-1}^{Y}\right)$, the latter equality becomes

$$v_{t,t-1} = \omega_0 (1-\omega_0) m + \sum_{i=1}^{q} \alpha_{0i} (1-\alpha_{0i}) Y_{t-i} + \sum_{j=1}^{p} \left(\beta_{0j} \left(1-\beta_{0j}\right) \mu_{t-j,t-1} + \beta_{0j}^2 v_{t-j,t-1} \right)$$

or more compactly as (2.10).

iv) Equality (2.11) is obtained in the same way as (2.9) while replacing $\mathcal{F}_{t-1}^{Y,\lambda}$ by $\mathcal{F}_{t-1}^{Y}$.

Proof of Theorem 2.2 We first show the necessity of (2.12). If (2.1) admits a stationary solution with $E\left(Y_t^2\right) < \infty$ and $E\left(\lambda_t^2\right) < \infty$ then by the independence of ε_t and λ_t we have

$$Var(Y_t) = \left(\sigma_0^2 + 1\right) Var(\lambda_t) + \sigma_0^2 (E(\lambda_t))^2, \tag{A.9}$$

and

$$\begin{aligned} Var(\lambda_t) &= Var[1 + \omega_0 \circ m + \sum_{i=1}^{q} \alpha_{0i} \circ Y_{t-i} + \sum_{j=1}^{p} \beta_{0j} \circ \lambda_{t-j}] \\ &= Var(\omega_0 \circ m) + \sum_{i=1}^{q} Var(\alpha_{0i} \circ Y_{t-i}) + \sum_{j=1}^{p} Var\left(\beta_{0j} \circ \lambda_{t-j}\right) + \\ &\quad 2 \sum_{i=1}^{q} \sum_{j=1}^{p} Cov\left(\alpha_{0i} \circ Y_{t-i}, \beta_{0j} \circ \lambda_{t-j}\right). \end{aligned}$$

Using the equality $Cov(X,Y) = E[Cov(X,Y|Z)] + Cov[E(X|Z), E(Y|Z)]$, it follows that

$$\begin{aligned} Cov(\alpha_{0i} \circ Y_{t-i}, \beta_{0j} \circ \lambda_{t-j}) &= E\left[Cov(\alpha_{0i} \circ Y_{t-i}, \beta_{0j} \circ \lambda_{t-j} | \mathcal{F}_{t-1}^{Y,\lambda})\right] + \\ &\quad Cov\left[E(\alpha_{0i} \circ Y_{t-i} | \mathcal{F}_{t-1}^{Y,\lambda}), E(\beta_{0j} \circ \lambda_{t-j} | \mathcal{F}_{t-1}^{Y,\lambda})\right] \\ &= \alpha_{0i} \beta_{0j} Cov(Y_{t-i}, \lambda_{t-j}). \end{aligned}$$

Moreover, in view of $(A.9)$, the equality $Var(\alpha \circ X) = \alpha(1-\alpha)E(X) + \alpha^2 Var(X)$, and the stationarity of $Var(\lambda_t)$, we obtain

$$\begin{aligned}
& Var(\lambda_t) \\
= & \sum_{i=1}^{q} \left(\alpha_{0i}(1-\alpha_{0i}) E(Y_{t-i}) + \alpha_{0i}^2 Var(Y_{t-i})\right) + \sum_{j=1}^{p} \left(\beta_{0j}\left(1-\beta_{0j}\right) E(\lambda_{t-j}) + \beta_{0j}^2 Var(\lambda_{t-j})\right) + \\
& 2\sum_{i=1}^{r} \alpha_{0i}\beta_{0i} Cov(Y_{t-i}, \lambda_{t-i}) + 2\sum_{i\neq j}^{r} \alpha_{0i}\beta_{0j} Cov(Y_{t-i}, \lambda_{t-j}) + \omega_0(1-\omega_0) m \\
= & Var(\lambda_t) \sum_{i=1}^{r} \left(\left(\sigma_0^2 + 1\right)\alpha_{0i}^2 + \beta_{0i}^2 + 2\alpha_{0i}\beta_{0i}\right) + 2\sum_{i\neq j}^{r} \alpha_{0i}\beta_{0j} Cov(\lambda_{t-i}, \lambda_{t-j}) + \\
& \mu \sum_{i=1}^{r} \left(\alpha_{0i}(1-\alpha_{0i}) + \beta_{0i}(1-\beta_{0i})\right) + \sigma_0^2 \mu^2 \sum_{i=1}^{r} \alpha_{0i}^2 + \omega_0(1-\omega_0) m. \qquad (A.10)
\end{aligned}$$

It is easy to show by induction that $Cov(\lambda_t, \lambda_{t-k}) \geq 0$ for all $k \geq 0$ (see $(A.16)$ below). Therefore, in view of $(A.10)$ and the positivity of model coefficients, it follows that (2.12) should be satisfied.

We now show the sufficiency of (2.12). Consider again the notations $(A.5)$ and $(A.6)$. The result will be established if we show that

$$E((\lambda_t^{(k)} - \lambda_t^{(k-1)})^2) \to 0 \text{ as } k \to \infty. \qquad (A.11)$$

For simplicity consider the case $p = q = 1$. The general case is established in the same line of reasoning. We have

$$\begin{aligned}
E((\lambda_t^{(k)} - \lambda_t^{(k-1)})^2) &= E[((\alpha_{01} \circ Y_{t-1}^{(k-1)} - \alpha_{01} \circ Y_{t-1}^{(k-2)}) + (\beta_{01} \circ \lambda_{t-1}^{(k-1)} - \beta_{01} \circ \lambda_{t-1}^{(k-2)}))^2] \\
&= E((\alpha_{01} \circ Y_{t-1}^{(k-1)} - \alpha_{01} \circ Y_{t-1}^{(k-2)})^2) + E((\beta_{01} \circ \lambda_{t-1}^{(k-1)} - \beta_{01} \circ \lambda_{t-1}^{(k-2)})^2) + \\
&\quad 2E(\alpha_{01} \circ Y_{t-1}^{(k-1)} - \alpha_{01} \circ Y_{t-1}^{(k-2)})(\beta_{01} \circ \lambda_{t-1}^{(k-1)} - \beta_{01} \circ \lambda_{t-1}^{(k-2)}) \\
&= E\left(\sum_{i=1}^{\left|Y_{t-1}^{(k-1)} - Y_{t-1}^{(k-2)}\right|} W_i^{t-1,k-1}\right)^2 + E\left(\sum_{j=1}^{\left|\lambda_{t-1}^{(k-1)} - \lambda_{t-1}^{(k-2)}\right|} Z_j^{t-1,k-1}\right)^2 + \\
&\quad 2E(\alpha_{01} \circ Y_{t-1}^{(k-1)} - \alpha_{01} \circ Y_{t-1}^{(k-2)})(\beta_{01} \circ \lambda_{t-1}^{(k-1)} - \beta_{01} \circ \lambda_{t-1}^{(k-2)}),
\end{aligned}$$

where $\left(W_i^{t-1,k-1}\right) \sim iid\mathcal{B}er(\alpha_{01})$ and $\left(Z_j^{t-1,k-1}\right) \sim iid\mathcal{B}er(\beta_{01})$. Using the fact that

$E(X^2) = np(1-p) + n^2p^2$ when $X \sim \mathcal{B}in(n,p)$, and the identity $E(XY) = Cov(X,Y) + E(X)E(Y)$, it follows that

$$\begin{aligned}
E((\lambda_t^{(k)} - \lambda_t^{(k-1)})^2) &= \alpha_{01}(1-\alpha_{01}) E\left|Y_{t-1}^{(k-1)} - Y_{t-1}^{(k-2)}\right| + \alpha_{01}^2 E((Y_{t-1}^{(k-1)} - Y_{t-1}^{(k-2)})^2) + \\
&\quad \beta_{01}(1-\beta_{01}) E\left|\lambda_{t-1}^{(k-1)} - \lambda_{t-1}^{(k-2)}\right| + \beta_{01}^2 E((\lambda_{t-1}^{(k-1)} - \lambda_{t-1}^{(k-2)})^2) + \\
&\quad 2Cov[\alpha_{01} \circ Y_{t-1}^{(k-1)} - \alpha_{01} \circ Y_{t-1}^{(k-2)}, \beta_{01} \circ \lambda_{t-1}^{(k-1)} - \beta_{01} \circ \lambda_{t-1}^{(k-2)}] \\
&\quad +2E(\alpha_{01} \circ Y_{t-1}^{(k-1)} - \alpha_{01} \circ Y_{t-1}^{(k-2)}) E(\beta_{01} \circ \lambda_{t-1}^{(k-1)} - \beta_{01} \circ \lambda_{t-1}^{(k-2)}) \\
&= \alpha_{01}(1-\alpha_{01}) E\left|Y_{t-1}^{(k-1)} - Y_{t-1}^{(k-2)}\right| + \alpha_{01}^2 E((Y_{t-1}^{(k-1)} - Y_{t-1}^{(k-2)})^2) + \\
&\quad \beta_{01}(1-\beta_{01}) E\left|\lambda_{t-1}^{(k-1)} - \lambda_{t-1}^{(k-2)}\right| + \beta_{01}^2 E((\lambda_{t-1}^{(k-1)} - \lambda_{t-1}^{(k-2)})^2) + \\
&\quad 2Cov(\alpha_{01} \circ Y_{t-1}^{(k-1)} - \alpha_{01} \circ Y_{t-1}^{(k-2)}, \beta_{01} \circ \lambda_{t-1}^{(k-1)} - \beta_{01} \circ \lambda_{t-1}^{(k-2)}) + \\
&\quad 2\alpha_{01}\beta_{01}(E(\lambda_{t-1}^{(k-1)} - \lambda_{t-1}^{(k-2)}))^2. \qquad (A.12)
\end{aligned}$$

Let $\mathcal{F}_{t-1,k-1}^{Y,\lambda}$ be the σ-algebra generated by $\left\{Y_{t-s}^{(k-j)}, \lambda_{t-1}^{(k-j)}, j \geq 1, s \geq 1\right\}$. Using the identity

$$Cov(X,Y) = E[Cov(X,Y|Z)] + Cov[E(X|Z), E(Y|Z)],$$

we have

$$\begin{aligned}
&Cov(\alpha_{01} \circ Y_{t-1}^{(k-1)} - \alpha_{01} \circ Y_{t-1}^{(k-2)}, \beta_{01} \circ \lambda_{t-1}^{(k-1)} - \beta_{01} \circ \lambda_{t-1}^{(k-2)}) \\
=\ & E(Cov(\alpha_{01} \circ Y_{t-1}^{(k-1)} - \alpha_{01} \circ Y_{t-1}^{(k-2)}, \beta_{01} \circ \lambda_{t-1}^{(k-1)} - \beta_{01} \circ \lambda_{t-1}^{(k-2)} | \mathcal{F}_{t-1,k-1}^{Y,\lambda})) + \\
& Cov(E(\alpha_{01} \circ Y_{t-1}^{(k-1)} - \alpha_{01} \circ Y_{t-1}^{(k-2)} | \mathcal{F}_{t-1,k-1}^{Y,\lambda}), E(\beta_{01} \circ \lambda_{t-1}^{(k-1)} - \beta_{01} \circ \lambda_{t-1}^{(k-2)} | \mathcal{F}_{t-1,k-1}^{Y,\lambda})) \\
=\ & Cov(\alpha_{01}(Y_{t-1}^{(k-1)} - Y_{t-1}^{(k-2)}), \beta_{01}(\lambda_{t-1}^{(k-1)} - \lambda_{t-1}^{(k-2)})) \\
=\ & \alpha_{01}\beta_{01}[E((\lambda_{t-1}^{(k-1)} - \lambda_{t-1}^{(k-2)})^2) - (E(\lambda_{t-1}^{(k-1)} - \lambda_{t-1}^{(k-2)}))^2]. \qquad (A.13)
\end{aligned}$$

Combining $(A.12)$ and $(A.13)$, and noting that $E\left|Y_{t-1}^{(k-1)} - Y_{t-1}^{(k-2)}\right| = E\left|\lambda_{t-1}^{(k-1)} - \lambda_{t-1}^{(k-2)}\right|$, and $E((Y_t^{(k)} - Y_t^{(k-1)})^2) = (\sigma_0^2 + 1) E((\lambda_t^{(k)} - \lambda_t^{(k-1)})^2)$, we obtain

$$\begin{aligned}
E[(\lambda_t^{(k)} - \lambda_t^{(k-1)})^2] &= \left((\sigma_0^2 + 1)\alpha_{01}^2 + (\beta_{01}^2 + 2\alpha_{01}\beta_{01})\right) E((\lambda_{t-1}^{(k-1)} - \lambda_{t-1}^{(k-2)})^2) + \\
&\quad (\alpha_{01}(1-\alpha_{01}) + \beta_{01}(1-\beta_{01})) E\left|\lambda_{t-1}^{(k-1)} - \lambda_{t-1}^{(k-2)}\right|. \qquad (A.14)
\end{aligned}$$

Note that when (2.12) is satisfied, the condition (2.6) immediately holds. Therefore, as shown in the proof of Theorem 2.1, $E\left|\lambda_{t-1}^{(k-1)} - \lambda_{t-1}^{(k-2)}\right| \to 0$ as $k \to \infty$. In view of $(A.14)$, it follows that $(A.11)$ holds true under (2.12). This concludes the proof. □

Proof of Proposition 2.3 Relationships (2.14) and (2.15) directly follows from $(A.9)$ and $(A.10)$ under (2.12). We now show (2.16). Note that

$$Cov(Y_t, Y_{t-k}) = Cov(\lambda_t \epsilon_t, \lambda_{t-k}\epsilon_{t-k}) = Cov(\lambda_t, \lambda_{t-k}).$$

Next

$$\begin{aligned} Cov(\lambda_t, \lambda_{t-k}) &= Cov(\omega_0 \circ m + \sum_{i=1}^{q} \alpha_{0i} \circ Y_{t-i} + \sum_{j=1}^{p} \beta_{0j} \circ \lambda_{t-j}, \lambda_{t-k}) \\ &= Cov\left(\omega_0 \circ m, \lambda_{t-k}\right) + \sum_{i=1}^{q} Cov\left(\alpha_{0i} \circ Y_{t-i}, \lambda_{t-k}\right) + \sum_{j=1}^{p} Cov\left(\beta_{0j} \circ \lambda_{t-j}, \lambda_{t-k}\right). \quad (A.15) \end{aligned}$$

By the independence of the counting series in $(2.1b)$ we have

$$\begin{aligned} Cov\left(\omega_0 \circ m, \lambda_{t-k}\right) &= Cov(\omega_0 \circ m, \omega_0 \circ m + \sum_{i=1}^{q} \alpha_{0i} \circ Y_{t-k-i} + \sum_{j=1}^{p} \beta_{0j} \circ \lambda_{t-k-j}) \\ &= Cov\left(\omega_0 \circ m, \omega_0 \circ m\right) = Var\left(\omega_0 \circ m\right) = \omega_0\left(1-\omega_0\right) m. \end{aligned}$$

Moreover,

$$\begin{aligned} Cov\left(\alpha_{0i} \circ Y_{t-i}, \lambda_{t-k}\right) &= ECov(\alpha_{0i} \circ Y_{t-i}, \lambda_{t-k} | \mathcal{F}_{t-1}^{Y,\lambda}) + Cov(E(\alpha_{0i} \circ Y_{t-i} | \mathcal{F}_{t-1}^{Y,\lambda}), E(\lambda_{t-k} | \mathcal{F}_{t-1}^{Y,\lambda})) \\ &= 0 + Cov\left(\alpha_{0i} Y_{t-i}, \lambda_{t-k}\right) = \alpha_{0i} Cov\left(\lambda_{t-i}, \lambda_{t-k}\right). \end{aligned}$$

Likewise

$$Cov\left(\beta_{0j} \circ \lambda_{t-j}, \lambda_{t-k}\right) = \beta_{0j} Cov\left(\lambda_{t-j}, \lambda_{t-k}\right).$$

Hence, $(A.15)$ becomes for $k \geq 1$

$$\gamma_k = \omega_0\left(1-\omega_0\right) m + \sum_{i=1}^{q} \alpha_{0i} Cov\left(\lambda_{t-i}, \lambda_{t-k}\right) + \sum_{j=1}^{p} \beta_{0j} Cov\left(\lambda_{t-j}, \lambda_{t-k}\right), \qquad (A.16)$$

which is exactly (2.16).

Proof of Theorem 2.3 By iterating backward the expression of λ_t we obtain

$$\lambda_t = \sum_{i=1}^{k-1} \left(\beta_{01}^i \circ 1 + \beta_{01}^i \omega_0 \circ m\right) + \sum_{j=1}^{k-1} \alpha_{01}\beta_{01}^j \circ Y_{t-j} + \beta_{01}^k \circ \lambda_{t-k}.$$

If (2.12) is satisfied then

$$\lim_{k\to\infty} E\left(\lambda_t - \sum_{i=1}^{k-1} \left(\beta_{01}^i \circ 1 + \beta_{01}^i \omega_0 \circ m\right) - \sum_{j=1}^{k-1} \alpha_{01}\beta_{01}^j \circ Y_{t-j}\right)^2 = \lim_{k\to\infty} E[\left(\beta_{01}^k \circ \lambda_{t-k}\right)^2]$$
$$= \lim_{k\to\infty} \left(\beta_{01}^k \left(1-\beta_{01}^k\right) E\left(\lambda_{t-k}\right) + \beta_{01}^{2k} Var\left(\lambda_{t-k}\right)\right) = 0,$$

and

$$\lambda_t = \sum_{i=1}^{\infty} \left(\beta_{01}^i \circ 1 + \beta_{01}^i \omega_0 \circ m\right) + \sum_{j=1}^{\infty} \alpha_{01}\beta_{01}^j \circ Y_{t-j},$$

where the series in the latter equality converge in mean square. Noticing that the contribution of the term $\sum_{i=1}^{\infty} \left(\beta_{01}^i \circ 1 + \beta_{01}^i \omega_0 \circ m\right)$, for large n, is negligible as compared with that of $\sum_{j=1}^{\infty} \alpha_{01}\beta_{01}^j \circ Y_{t-j}$, the limiting behavior of $P\left(\lambda_t > n\right)$ is equivalent to that of $P\left(\sum_{j=1}^{\infty} \alpha_{01}\beta_{01}^j \circ Y_{t-j} > n\right)$ as $n \to \infty$. By Theorem 3 in Hall (2001), it follows that, for all $j \geq 0$,

$$\frac{P\left(\alpha_{01}\beta_{01}^j \circ Y_{t-j} > n\right)}{P\left(Y_{t-j} > n\right)} \sim \left(\alpha_{01}\beta_{01}^j\right)^{\xi} \text{ as } n \to \infty.$$

Recall that $Y_t, Y_{t-1}, ...$ are not independent random variables. First, consider the partial sum $Y_t^{(\ell)} = \sum_{j=0}^{\ell} \alpha_{01}\beta_{01}^j \circ Y_{t-j}$, for some $\ell \geq 0$. As a consequence of Lemma 2.1 in Davis and Resnick (1996), it is sufficient to prove that for $0 \leq i_1 < i_2 \leq \ell$

$$\frac{P\left(\alpha_{01}\beta_{01}^{i_1} \circ Y_{t-i_1} > n, \alpha_{01}\beta_{01}^{i_2} \circ Y_{t-i_2} > n\right)}{P\left(Y_t > n\right)} \to 0 \text{ as } n \to \infty,$$

in order to conclude that

$$\frac{P\left(Y_t^{(\ell)} > n\right)}{P\left(Y_t > n\right)} \sim \sum_{j=0}^{\ell} \left(\alpha_{01}\beta_{01}^j\right)^{\xi} \text{ as } n \to \infty. \tag{A.17}$$

Now,

$$\frac{P\left(\alpha_{01}\beta_{01}^{i_1} \circ Y_{t-i_1} > n, \alpha_{01}\beta_{01}^{i_2} \circ Y_{t-i_2} > n\right)}{P\left(Y_t > n\right)} \leq \frac{P\left(Y_{t-i_1} > n, Y_{t-i_2} > n\right)}{P\left(Y_t > n\right)}$$
$$= P\left(Y_{t-i_1} > n | Y_{t-i_2} > n\right).$$

Note that $P\left(Y_{t-i_1}>n|Y_{t-i_2}>n\right)$ represents the upper tail dependence coefficient (Ledford and Tawn, 1996; Davis and Mikosch, 2009). According to Ledford and Tawn (1996) (see also Ledford and Tawn, 2003)

$$\lim_{n\to\infty} P\left(Y_{t-i_1}>n|Y_{t-i_2}>n\right)=0,$$

provided that $\xi>1$, or if $\xi=1$ and $L(n)\to 0$ as $n\to\infty$. The latter conditions are not very restrictive. In particular, the condition $\xi>1$ implies to assume that the marginal mean is finite, which immediately follows from (2.12). So, we can conclude that the result in $(A.17)$ holds under (2.12). Now, we extend the result above so that the number of summands can be infinite. By using the argument in Corollary 2.4 in Davis and Resnick (1996) we obtain

$$\frac{P\left(\sum_{j=0}^{\infty}\alpha_{01}\beta_{01}^{j}\circ Y_{t-j}>n\right)}{P\left(Y_t>n\right)}\sim\frac{\alpha_{01}^{\xi}}{1-\beta_{01}^{\xi}} \text{ as } n\to\infty,$$

establishing the result.

Proof of Theorem 3.1 *i) Proof of* (3.9). Let $L_n\left(\theta,\vartheta^*\right)$ and $l_t\left(\theta,\vartheta_t^*\right)$ be defined as $\widetilde{L}_n\left(\theta,\widetilde{\vartheta}^*\right)$ and $\widetilde{l}_t\left(\theta,\widetilde{\vartheta}_t^*\right)$, respectively, while replacing in (3.7b) $\widetilde{\mu}_t(\theta)$ by $\mu_t(\theta)$, and $\widetilde{\vartheta}_t\left(\theta_*,\sigma_*^2\right)$ by $\vartheta_t\left(\theta_*,\sigma_*^2\right)$. We first show that

$$\sup_{\theta\in\Theta}\left|\widetilde{L}_n\left(\theta,\widetilde{\vartheta}^*\right)-L_n\left(\theta,\vartheta^*\right)\right|\underset{n\to\infty}{\overset{a.s.}{\to}}0. \qquad (A.18)$$

Let $a_t:=\sup_{\theta\in\Theta}\left|\widetilde{\mu}_t(\theta)-\mu_t(\theta)\right|$, $b_t:=\sup_{\theta\in\Theta}\left|\widetilde{v}_t(\theta)-v_t(\theta)\right|$ and $c_t^*:=\sup_{\theta\in\Theta}\left|\widetilde{\vartheta}_t\left(\theta,\sigma_*^2\right)-\vartheta_t\left(\theta,\sigma_*^2\right)\right|$. Standard arguments (cf. Aknouche and Francq, 2022a, Corollary 2.2) show that under (3.2) and (2.6), $a_t(1+Y_t+\sup_{\theta\in\Theta}\mu_t(\theta))\underset{t\to\infty}{\overset{a.s.}{\to}}0$. Likewise, (3.5) and (2.12) entail $b_t(1+Y_t^2+\sup_{\theta\in\Theta}\mu_t^2(\theta))\underset{t\to\infty}{\overset{a.s.}{\to}}0$, and thus $c_t^*(1+Y_t^2+\sup_{\theta\in\Theta}\mu_t^2(\theta))\underset{t\to\infty}{\overset{a.s.}{\to}}0$ for all $\sigma_*^2>0$. Therefore, noting that $\widetilde{\vartheta}_t\left(\theta^*,\sigma_*^2\right)\geq\widetilde{\vartheta}$ and $\vartheta_t\left(\theta^*,\sigma_*^2\right)\geq\vartheta$, are bounded from below, we obtain the inequalities

$$\begin{aligned}\sup_{\theta\in\Theta}\left|L_n\left(\theta,\vartheta^*\right)-\widetilde{L}_n\left(\theta,\widetilde{\vartheta}^*\right)\right| &\leq \frac{1}{n}\sum_{t=1}^{n}\sup_{\theta\in\Theta}\frac{|\widetilde{\mu}_t(\theta)-\mu_t(\theta)||\widetilde{\mu}_t(\theta)+\mu_t(\theta)-2Y_t|}{\widetilde{\vartheta}_t(\theta^*,\sigma_*^2)}+\frac{\left|\widetilde{\vartheta}_t\left(\theta^*,\sigma_*^2\right)-\vartheta_t\left(\theta^*,\sigma_*^2\right)\right||Y_t-\mu_t(\theta)|^2}{\widetilde{\vartheta}_t(\theta^*,\sigma_*^2)\vartheta_t(\theta^*,\sigma_*^2)}\\ &\leq \frac{1}{n}\sum_{t=1}^{n}\frac{2}{\widetilde{\vartheta}}a_t[1+Y_t+\sup_{\theta\in\Theta}\mu_t(\theta)]+\frac{2}{\widetilde{\vartheta}\vartheta}c_t^*[Y_t^2+\sup_{\theta\in\Theta}\mu_t^2(\theta)],\end{aligned}$$

from which $(A.18)$ follows by Césaro's lemma.

Now, in view of (2.6), (2.12), (3.5), and (3.2), the process $\{(v_t(\theta), \vartheta_t(\theta,\sigma^2), \mu_t(\theta), Y_t), t\in\mathbb{Z}\}$ is stationary and ergodic. Hence $E(l_t(\theta,\vartheta_t^*)) = \lim_{n\to\infty} L_n(\theta,\vartheta^*) \in [0,\infty]$ a.s. In addition, for all $\theta^*\in\Theta$ and $\sigma_*^2>0$,

$$E(l_t(\theta_0,\vartheta_t^*)) = E\left(\frac{(Y_t-\mu_t(\theta_0))^2}{\vartheta_t(\theta^*,\sigma_*^2)}\right) = E\left(E\left(\frac{(Y_t-\mu_t(\theta_0))^2}{\vartheta_t(\theta^*,\sigma_*^2)}|\mathcal{F}_{t-1}^Y\right)\right) = E\left(\frac{\vartheta_t\left(\theta_0,\sigma_0^2\right)}{\vartheta_t(\theta^*,\sigma_*^2)}\right) < \infty.$$

Thus $E(l_t(\theta_0,\vartheta_t^*)) \le E(l_t(\theta,\vartheta_t^*))$ for all $\theta\in\Theta$ with equality if and only if $\mu_t(\theta) = \mu_t(\theta_0)$, and by **A2** if and only if $\theta=\theta_0$. Finally, in addition to the above arguments, the consistency result (3.9) follows from standard compactness arguments under **A1** (see e.g. Ahmad and Francq, 2016, Theorem 2.1).

ii) *Proof of* (3.10). Let $d_t := \sup_{\theta\in\Theta}\left|\frac{\partial\mu_t(\theta)}{\partial\theta} - \frac{\partial\widetilde{\mu}_t(\theta)}{\partial\theta}\right|$. Due to the linear form of $\mu_t(\theta)$ and under (3.2), which entails (3.5), we have for some $\tau > \frac{1}{2}$ (see also Aknouche and Francq, 2022a; Ahmad and Francq, 2016)

$$t^\tau a_t \sup_{\theta\in\Theta}\left\|\frac{\mu_t(\theta)}{\partial\theta}\right\| \overset{a.s.}{=} O(1),\ t^\tau d_t(Y_t + \sup_{\theta\in\Theta}\mu_t(\theta)) \overset{a.s.}{=} O(1),\text{ and}$$
$$t^\tau\left|\widetilde{\vartheta}_t\left(\theta^*,\sigma_*^2\right) - \vartheta_t\left(\theta^*,\sigma_*^2\right)\right|\sup_{\theta\in\Theta}\left\|\frac{\partial\mu_t(\theta)}{\partial\theta}\right\|(Y_t+\sup_{\theta\in\Theta}\mu_t(\theta)) \overset{a.s.}{=} O(1).$$

Therefore

$$\begin{aligned}&\sqrt{n}\sup_{\theta\in\Theta}\left|\frac{\partial\widetilde{L}_n(\theta,\widetilde{\vartheta}^*)}{\partial\theta} - \frac{\partial L_n(\theta,\vartheta^*)}{\partial\theta}\right|\\ \le\ & \frac{2}{\sqrt{n}}\sum_{t=1}^n\Bigg[a_t + \left|\widetilde{\vartheta}_t\left(\theta^*,\sigma_*^2\right) - \vartheta_t\left(\theta^*,\sigma_*^2\right)\right|(Y_t + a_t + \sup_{\theta\in\Theta}\mu_t(\theta))\sup_{\theta\in\Theta}\left\|\frac{\mu_t(\theta)}{\partial\theta}\right\|\\ & + d_t(Y_t + a_t + \sup_{\theta\in\Theta}\mu_t(\theta))\Bigg] \underset{n\to\infty}{\overset{a.s.}{\to}} 0.\end{aligned}$$

Next, under $E(Y_t^2) < \infty$, which is ensured by **A1**, the matrices $I(\theta_0,\theta^*,\sigma_*^2)$ and $J(\theta_0,\theta^*,\sigma_*^2)$ are finite for all $\theta^*\in\Theta$ and $\sigma_*^2>0$. Furthermore, the invertibility of $J(\theta_0,\theta^*,\sigma_*^2)$ follows from the identifiability assumption **A2** and the non-degeneracy of ε_t (i.e. $\sigma_0^2>0$) by the same argument in Remark 2.3 of Ahmad and Francq (2016) (see also Francq and Zakoian, 2019). Now, in view of **A1**, the martingale central limit theorem yields

$$\sqrt{n}\frac{\partial L_n(\theta_0,\vartheta^*)}{\partial\theta} = \frac{1}{\sqrt{n}}\sum_{t=1}^n\frac{Y_t-\mu_t(\theta_0)}{\vartheta_t(\theta^*,\sigma_*^2)}\frac{\partial\mu_t(\theta_0)}{\partial\theta} \underset{n\to\infty}{\overset{\mathcal{D}}{\to}} \mathcal{N}\left(0, 4I\left(\theta_0,\theta^*,\sigma_*^2\right)\right).$$

Moreover, the ergodic theorem, the dominated convergence theorem, the continuity of $\frac{\partial^2 l_t(\theta,\vartheta_t^*)}{\partial\theta\partial\theta'}$, and the finiteness of $E\left(Y_t^2\right)$ imply that

$$\frac{\partial^2 L_n(\widehat{\theta}_{1W},\vartheta^*)}{\partial\theta\partial\theta'} \underset{n\to\infty}{\overset{a.s.}{\to}} 2J\left(\theta_0,\theta^*,\sigma_*^2\right).$$

Finally, the asymptotic normality result (3.10) follows from Taylor expansions, **A3** and standard arguments (see also Aknouche and Francq, 2022a, Theorem 2.1).

Proof of Theorem 3.2 Let $U_t(\theta) = \frac{(Y_t-\mu_t(\theta))^2-v_t(\theta)}{v_t(\theta)+\mu_t^2(\theta)}$, and denote by $o_{a.s.}(1)$ a term converging almost surely to 0 as $n\to\infty$. A Taylor expansion of $U_t(\widehat{\theta}_{1W})$ around θ_0 gives

$$\widehat{\sigma}_n^2 = \frac{1}{n}\sum_{t=1}^n U_t(\widehat{\theta}_{1W}) = \frac{1}{n}\sum_{t=1}^n U_t(\theta_0) + (\widehat{\theta}_{1W}-\theta_0)\frac{1}{n}\sum_{t=1}^n \frac{\partial U_t(\overline{\theta})}{\partial\theta}, \tag{A.19}$$

where $\overline{\theta}$ is between $\widehat{\theta}_{1W}$ and θ_0, and

$$\frac{\partial U_t(\theta)}{\partial\theta} = \frac{2(\mu_t(\theta)-Y_t)\left(v_t(\theta)+\mu_t^2(\theta)\right)-2\mu_t(\theta)\left((Y_t-\mu_t(\theta))^2-v_t(\theta)\right)}{\left(v_t(\theta)+\mu_t^2(\theta)\right)^2}\frac{\partial\mu_t(\theta)}{\partial\theta} - \frac{\left(v_t(\theta)+\mu_t^2(\theta)\right)+\left((Y_t-\mu_t(\theta))^2-v_t(\theta)\right)}{\left(v_t(\theta)+\mu_t^2(\theta)\right)^2}\frac{\partial v_t(\theta)}{\partial\theta}.$$

Therefore, a similar argument to the one in Francq and Zakoian (2019, p. 197) shows that $E\left(\sup_{\theta\in\mathcal{V}(\theta_0)}\left\|\frac{\partial U_t(\theta)}{\partial\theta}\right\|\right) < \infty$ for some neighborhood $\mathcal{V}(\theta_0)$ of θ_0. Hence, the ergodic theorem and the consistency of $\widehat{\theta}_{1W}$ entail

$$\limsup_{n\to\infty}\left\|\frac{1}{n}\sum_{t=1}^n\frac{\partial U_t(\overline{\theta})}{\partial\theta}\right\| \le \limsup_{n\to\infty}\frac{1}{n}\sum_{t=1}^n\sup_{\theta\in\mathcal{V}(\theta_0)}\left\|\frac{\partial U_t(\theta)}{\partial\theta}\right\| = E\left(\sup_{\theta\in\mathcal{V}(\theta_0)}\left\|\frac{\partial U_t(\theta)}{\partial\theta}\right\|\right),$$

and

$$(\widehat{\theta}_{1W}-\theta_0)\frac{1}{n}\sum_{t=1}^n\frac{\partial U_t(\overline{\theta})}{\partial\theta} = o_{a.s.}(1).$$

In view of $(A.19)$, we obtain

$$\widehat{\sigma}_n^2 - \sigma_0^2 = \frac{1}{n}\sum_{t=1}^n\frac{(Y_t-\mu_t(\theta_0))^2-\vartheta_t(\theta_0)}{v_t(\theta_0)+\mu_t^2(\theta_0)} + o_{a.s.}(1).$$

Hence, the consistency result (3.14) follows from the ergodic theorem and the fact that $E\left(\frac{(Y_t-\mu_t(\theta_0))^2-\vartheta_t(\theta_0)}{v_t(\theta_0)+\mu_t^2(\theta_0)}\right) = 0$. Similar standard arguments show (3.16).

Proof of Theorem 3.3 To lighten the notation, let $\phi = (\theta',\sigma^2)' \in \Phi = \Theta\times\Delta$. From the invertibility conditions (3.2) and (3.5) we have

$$\sup_{\phi\in\mathcal{V}(\phi_0)}\left|\widetilde{\vartheta}_t(\phi)-\vartheta_t(\phi)\right| \le K\rho^t, \tag{A.20}$$

where $\mathcal{V}(\phi_0)$ is a neighborhood of $\phi_0 = (\theta_0', \sigma_0^2)'$, and $K > 0$ and $\rho \in [0,1]$ are generic constants. Moreover, noticing that $\widehat{\vartheta}_t := \widetilde{\vartheta}_t(\widehat{\phi})$ and $\vartheta_t := \vartheta_t(\phi_0)$, a Taylor expansion and $(A.20)$ yield

$$\begin{aligned}\left|\widehat{\vartheta}_t - \vartheta_t\right| &= \left|\widetilde{\vartheta}_t(\widehat{\phi}) - \vartheta_t(\phi_0)\right| \leq \left|\widetilde{\vartheta}_t(\widehat{\phi}) - \vartheta_t(\widehat{\phi})\right| + \left|\vartheta_t(\widehat{\phi}) - \vartheta_t(\phi_0)\right| \\ &\leq K\rho^t + \left\|\widehat{\phi} - \phi_0\right\| \sup_{\phi\in\mathcal{V}(\phi_0)} \left\|\frac{\partial \vartheta_t(\phi)}{\partial \phi}\right\|,\end{aligned}$$

where $\widehat{\phi} = (\widehat{\theta}_{1W}', \widehat{\sigma}_n^2)'$. Hence under **A1**, the consistency of $\widehat{\phi}$, and the finiteness of

$$\sum_{t=1}^{\infty} \rho^t E \sup_{\theta\in\mathcal{V}(\theta_0)} (Y_t - \mu_t(\theta))^2,$$

which follows from **A1**, we obtain using again Césaro's lemma

$$\begin{aligned}&\sup_{\theta\in\Theta}\left|\frac{\partial L_n(\theta,\vartheta)}{\partial\theta} - \frac{\partial \widetilde{L}_n(\theta,\widehat{\vartheta})}{\partial\theta}\right| = \sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{t=1}^{n}(l_t(\theta,\vartheta_t) - \widetilde{l}_t(\theta,\widehat{\vartheta}_t))\right| \\ &\leq \frac{1}{n}\sum_{t=1}^{n}\Bigg[2a_t(1 + Y_t + \sup_{\theta\in\Theta}\mu_t(\theta)) + \\ &\qquad 2\sup_{\theta\in\Theta}(Y_t - \mu_t(\theta))^2\left(K\rho^t + \left\|\widehat{\phi} - \phi_0\right\| \sup_{\phi\in\mathcal{V}(\phi_0)}\left\|\frac{\partial\vartheta_t(\phi)}{\partial\phi}\right\|\right)\Bigg] \underset{n\to\infty}{\overset{a.s.}{\to}} 0.\end{aligned}$$

The rest of the proof of (3.18) follows similarly to the proof of Theorem 3.1.

Turn now to the asymptotic normality result (3.19). Using the same arguments as in the proof of Theorem 3.1, **A1** entails

$$\sqrt{n}\sup_{\theta\in\Theta}\left(\frac{\partial L_n(\theta,\vartheta)}{\partial\theta} - \frac{\partial\widetilde{L}_n(\theta,\widehat{\vartheta})}{\partial\theta}\right) = \frac{1}{\sqrt{n}}\sup_{\theta\in\Theta}\sum_{t=1}^{n}\left(l_t(\theta,\vartheta_t) - \widetilde{l}_t(\theta,\widehat{\vartheta})\right) = o_p(1).$$

Hence, in view of **A1** and **A3**, the consistency of $\widehat{\phi}$, which implies $\sqrt{n}\left(\widehat{\phi} - \phi_0\right) = O_p(1)$, and using Taylor expansions following the same lines of the proof of Aknouche and Francq (2022a, Theorem 2.1), we finally obtain

$$0 = \sqrt{n}\frac{\partial\widetilde{L}_n(\theta,\widehat{\vartheta})}{\partial\theta} = \sqrt{n}\frac{\partial L_n(\theta,\widehat{\vartheta})}{\partial\theta} + o_p(1) = \sqrt{n}\frac{\partial L_n(\theta_0,\vartheta)}{\partial\theta} + o_p(1).$$

Under **A1**, the ergodic theorem entails $\frac{1}{n}\sum_{t=1}^{n}\frac{1}{\vartheta_t(\theta_0,\sigma_0^2)}\frac{\partial\mu_t(\theta_0)}{\partial\theta} \underset{n\to\infty}{\overset{a.s.}{\to}} J(\theta_0,\theta_0,\sigma_0^2)$. Since $\{Y_t - \mu_t, t\in\mathbb{Z}\}$ is an $\left(\mathcal{F}_t^Y\right)$-martingale difference, the asymptotic normality result (3.19) thus follows from the following convergence

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{Y_t-\mu_t(\theta_0)}{\vartheta_t(\theta_0,\sigma_0^2)}\frac{\partial\mu_t(\theta_0)}{\partial\theta}\underset{n\to\infty}{\overset{\mathcal{D}}{\to}}\mathcal{N}\left(0,J\left(\theta_0,\theta_0,\sigma_0^2\right)\right),$$

which, in turn, is a consequence of the central limit theorem for stationary and ergodic (square integrable) martingale differences.

Acknowledgements Manuel Scotto acknowledges the Project UIDB/04621/2020 (CEMAT/UL).

Data availability statement The Ecoli dataset is available from the tscount R package of Liboschik et *al* (2017). The Exchange rate (ExRate) series is taken from Gorgi (2020).

References

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

[2] Aknouche, A. (2012). Multistage weighted least squares estimation of ARCH processes in the stable and unstable cases. *Statistical Inference for Stochastic Processes* 15, 241–256.

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

[4] Aknouche, A. and Francq, C. (2021). Count and duration time series with equal conditional stochastic and mean orders. *Econometric Theory* 37, 248–280.

[5] Aknouche, A. and Francq, C. (2022a). Two-stage weighted least squares estimator of the conditional mean of observation-driven time series models. *Journal of Econometrics*, forthcoming, https://doi.org/10.1016/j.jeconom.2021.09.002.

[6] Aknouche, A. and Francq, C. (2022b). Stationarity and ergodicity of Markov-Switching positive conditional mean models. *Journal of Time Series Analysis*, forthcoming, https://doi.org/10.1111/jtsa.12621.

[7] Al-Osh, M.A. and Alzaid, A.A. (1987). First-order integer-valued autoregressive (INAR(1)) process. *Journal of Time Series Analysis* 8**,** 261–275.

[8] Barczy, M., Ispany, M., Pap, G., Scotto, M. and Silva, M.E. (2010). Innovational outliers in INAR(1) models. *Communications in Statistics- Theory Methods* 39, 3343–3362.

[9] Barczy, M., Ispany, M., Pap, G., Scotto, M. and Silva, M.E. (2012). Additive outliers in INAR(1) models. *Statistical Papers* 53, 935–949.

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

[11] Bowsher, C.G. and Swain. P.S. (2012). Identifying sources of variation and the flow of information in biochemical networks. *Proceedings of the National Academy of Sciences (PNAS)* 109, E1320-E1328.

[12] Breiman, L. (1965). On some limit theorems similar to the arc-sin law. *Theory of Probability and its Applications* 10, 323–331.

[13] Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. (1994) *Time Series Analysis: Forecasting and Control*. 3rd Edition, Prentice Hall, Englewood Cliff, New Jersey.

[14] Brockwell, P.J. and Davis, R.A. (1991). *Time series: Theory and methods*. Springer Science & Business Media.

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

[16] Davis, R.A. and Mikosch, T. (2009). The extremogram: a correlogram for extreme events. *Bernoulli* 15, 977–1009.

[17] Davis, R.A. and Resnick, S.I. (1996). Limit theory for bilinear processes with heavy-tailed noise. *Annals of Applied Probability* 6, 1191–1201.

[18] Davis, R.A. and 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.

[19] Davis, R.A., Holan, S.H., Lund, R. and Ravishanker, N. (2016). *Handbook of Discrete-valued Time Series*. Boca Raton: CRC Press.

[20] Denisov, D. and Zwart, B. (2007). On a theorem of Breiman and a class of random difference equations. *Journal of Applied Probability* 44, 1031–1046.

[21] Doukhan, P., Latour, A. and Oraichi, D. (2006). A simple integer-valued bilinear time series model. *Advances in Applied Probability* 38, 559–78.

[22] Doukhan, P., Fokianos, K. and Rynkiewicz, J. (2021). Mixtures of nonlinear Poisson autoregressions. *Journal of Time Series Analysis* 42, 107–135.

[23] Du, J.-G. and Li, Y. (1991). The integer-valued autoregressive (INAR(p)) model. *Journal of Time Series Analysis* 12, 130–142.

[24] Embrechts, P., Kluppelberg, C. and Mikosch, T. (2013). *Modelling Extremal Events for Insurance and Finance*. Berlin: Springer-Verlag, Second Edition.

[25] Engle, R. and Russell, J. (1998). Autoregressive conditional duration: A new model for irregular spaced transaction data. *Econometrica* 66, 1127–1162.

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

[27] Engle, R.F. and Gallo, G.M. (2006). A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics* 131, 3–27.

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

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

[30] Francq, C. and Zakoian, J.-M. (2019). *GARCH Models: Structure, Statistical Inference and Financial Applications*, 2nd edition, John Wiley & Sons.

[31] Gorgi, P. (2020). Beta–negative binomial auto-regressions for modelling integer-valued time series with extreme observations. *Journal of the Royal Statistical Society* B 82, 1325-1347.

[32] Gouriéroux, C., Monfort, A. and Trognon, A. (1984). Pseudo maximum likelihood methods: Theory. *Econometrica* 52, 681–700.

[33] Grunwald, G., Hyndman, R., Tedesco, L. and Tweedie, R. (2000). Non-Gaussian conditional linear AR(1) models. *Australian and New Zealand Journal of Statistics* 42, 479–495.

[34] Hall, A. (2001). Extremes of integer-valued moving averages models with regularly varying tails. *Extremes* 4, 219–239.

[35] Hall, A., Scotto, M.G. and Cruz, J.P. (2010). Extremes of integer-valued moving average sequences. *TEST* 19, 359–74.

[36] Heinen, A. (2003). Modelling time series count data: an autoregressive conditional Poisson model. *CORE Discussion Paper 2003/62.*

[37] Jacobsen, M., Mikosch, T., Rosinski, J. and Samorodnitsky, G. (2009). Inverse problems for regular variation of linear filters, a cancellation property for σ-finite measures and identification of stable laws. *Annals of Applied Probability* 19, 210–242.

[38] Klimko, L.A. and Nelson, P.I. (1978). On conditional least squares estimation for stochastic process. *Annals of Statistics* 6, 629–642.

[39] Latour, A. (1998). Existence and stochastic structure of a non-negative integer-valued autoregressive processes. *Journal of Time Series Analysis* 4, 439–55.

[40] Ledford, A.W. and Tawn, J.A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika* 83, 169–187.

[41] Ledford, A.W. and Tawn, J.A. (2003). Diagnostics for dependence within time series extremes. *Journal of the Royal Statistical Society* B 65, 521–543.

[42] Liboschik, T., Fokianos, K. and Fried, R. (2017). tscount: An R Package for Analysis of Count Time Series Following Generalized Linear Models. *Journal of Statistical Software* 82, 1–51.

[43] Mao, H., Zhu, F. and Cui, Y. (2020). A generalized mixture integer-valued GARCH model. *Statistical Methods & Applications* 29, 527–552.

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

[45] McKenzie, E. (2003). *Discrete variate time series*, in Handbook of statistics. Amsterdam: Elsevier Science.

[46] Roitershtein, A. and Zhong, Z. (2013). On random coefficient INAR(1) processes. *Science China Mathematics* 56, 177–200.

[47] Qian, L., Li, Q. and Zhu, F. (2020). Modelling heavy-tailedness in count time series. *Applied Mathematical Modelling* 82, 766–784.

[48] Scotto, M.G. Weiss, C.H. and Gouveia, S. (2015). Thinning-based models in the analysis of integer-valued time series: a review. *Statistical Modelling*, 15, 590–618.

[49] Scotto, M.G., Weiss, C.H., Moller, T.A. and Gouveia, S. (2018). The max-INAR(1) model for count processes. *TEST* 27, 850–870.

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

[51] Silva, M.E. and Oliveira, V.L. (2004). Difference equations for the higher order moments and cumulants of the INAR(1) model. *Journal of Time Series Analysis* 25, 317–333.

[52] Steutel, F.W. and varn Harn, K. (1979). Discrete analogues of self-decomposability and stability. *Annals of Probability* 7, 893–899.

[53] Wan, W.Y. and Chan, J.S-K. (2011). Bayesian analysis of robust Poisson geometric process model using heavy-tailed distributions. *Computational Statistics and Data Analysis* 55, 687–702.

[54] Weiss, C.H. (2018). *An introduction to discrete-valued time series*. Wiley.

[55] Wong, C. S. and Li, W.K. (2000). On a mixture autoregressive model. *Journal of the Royal Statistical Society*, B 62, 95–115.

[56] Wooldridge, J.M. (1999). Quasi-likelihood methods for count data. In M.H. Pesaran & P. Schmidt (eds.), *Handbook of Applied Econometrics, Volume 2: Microeconomics*, pp. 35–406. Blackwell.

[57] Zheng, T., Xiao, H. and Chen, R. (2015). Generalized ARMA with martingale difference errors. *Journal of Econometrics* 189, 492–506.

Supplementary material to "A multiplicative thinning-based integer-valued GARCH model"

By

Abdelhakim Aknouche and Manuel G. Scotto

1 Validating the unconditional mean and variance formulas

We generate 1000 replications of the MthINGARCH(1,1) model with a sample size $n = 500$, a Poisson(1) distributed innovation sequence, and parameters m, $\theta_0 = (\omega_0, \alpha_{01}, \beta_{01})'$ and σ_0^2 (cf. Table S.1). For all cases, the parameters θ_0 and σ_0^2 are such that the MthINGARCH(1,1) model is stationary with finite variance (i.e. $(\alpha_{01} + \beta_{01})^2 + \alpha_{01}^2\sigma_0^2 < 1$). The means of the sample means ($m(\overline{Y})$, $\overline{Y} = \frac{1}{n}\sum_{t=1}^{n} Y_t$) and sample variances ($m(S_Y^2)$, $S_Y^2 = \frac{1}{n}\sum_{t=1}^{n}(Y_t - \overline{Y})^2$) over the 1000 replications are compared with the theoretical moments $E(Y_t)$ and $Var(Y_t)$ obtained from (2.7) and (2.15), respectively. From Table S.1 it can be observed that the generated sample moments are very close to their theoretical counterparts, which suggests that (2.7) and (2.15) are plausible. When $\alpha_{01} + \beta_{01}$ is near to 1, the generated series becomes highly (unconditionally) overdispersed. In fact, the sample variance can reaches 90 times the sample mean (cf. Figures S.1-S.4), which cannot be reproduced with the standard stationary

INAR and INGARCH models.

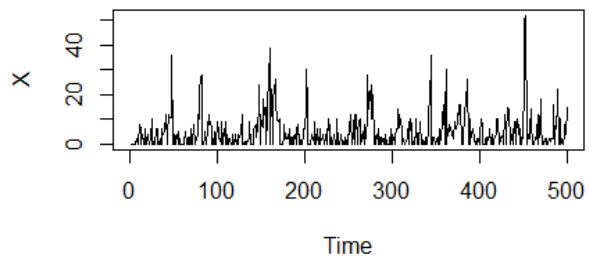


Figure S.1. An MthINGARCH$(1,1)$ series with $n = 500$, $m = 1$, $\theta_0 = (0.5, 0.4, 0.3)'$, $\varepsilon_t \sim \mathcal{P}(1)$, $\overline{Y} = 13.908$, and $S_Y^2 = 218.5286$.

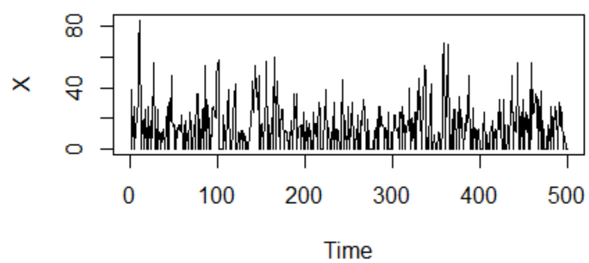


Figure S.2. An MthINGARCH$(1,1)$ series with $n = 500$, $m = 5$, $\theta_0 = (0.7, 0.2, 0.5)'$, $\varepsilon_t \sim \mathcal{P}(1)$, $\overline{Y} = 4.972$, $S_Y^2 = 49.815$.

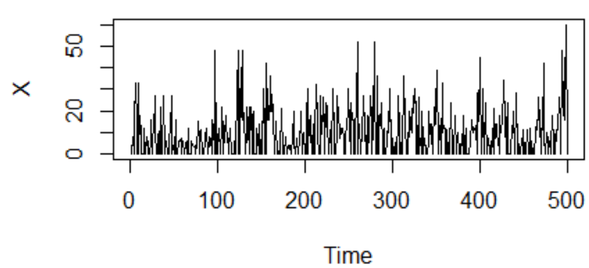


Figure S.3. An MthINGARCH$(1,1)$ series with $n = 500$, $m = 4$, $\theta_0 = (0.2, 0.1, 0.7)'$, $\varepsilon_t \sim \mathcal{P}(1)$, $\overline{Y} = 9.798$, $S_Y^2 = 121.288$.

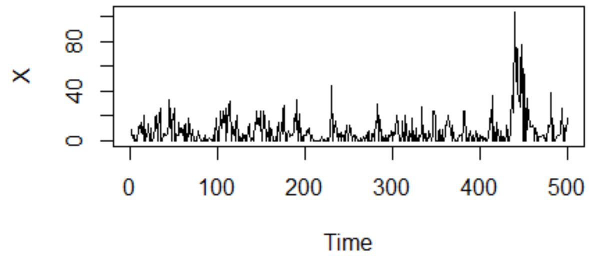


Figure S.4. An MthINGARCH$(1,1)$ series with $n = 500$, $m = 2$, $\theta_0 = (0.3, 0.3, 0.5)'$, $\varepsilon_t \sim \mathcal{P}(1)$, $\overline{Y} = 8.602$, $S_Y^2 = 135.290$.

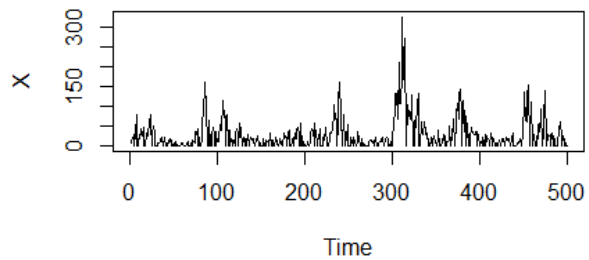


Figure S.5. An MthINGARCH$(1,1)$ series with $n = 500$, $m = 9$, $\theta_0 = (0.1, 0.3, 0.6)'$, $\varepsilon_t \sim \mathcal{P}(1)$, $\overline{Y} = 27.584$, $S_Y^2 = 1561.233$.

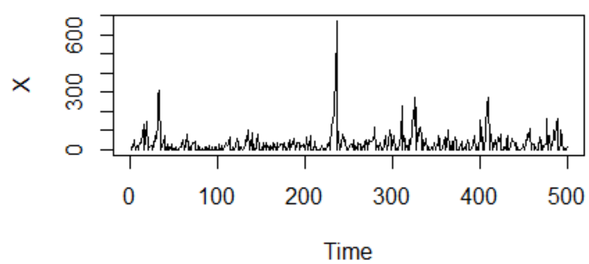


Figure S.6. An MthINGARCH$(1,1)$ series with $n = 500$, $m = 9$, $\theta_0 = (0.9, 0.5, 0.2)'$, $\varepsilon_t \sim \mathcal{P}(1)$, $\overline{Y} = 32.598$, $S_Y^2 = 3063.435$.

	Parameters (m,θ_0',σ_0^2)					Means		Variances		$\frac{Var(Y_t)}{E(Y_t)}$
1	m	ω_0	α_{01}	β_{01}	σ_0^2	$E(Y_t)$	5.000	$Var(Y_t)$	62.143	12.428
	1	0.5	0.4	0.2	1	$m(\overline{Y})$	5.009	$m(S_Y^2)$	62.905	
2	m	ω_0	α_{01}	β_{01}	σ_0^2	$E(Y_t)$	15.000	$Var(Y_t)$	293.936	19.596
	5	0.7	0.2	0.5	1	$m(\overline{Y})$	15.021	$m(S_Y^2)$	292.588	
3	m	ω_0	α_{01}	β_{01}	σ_0^2	$E(Y_t)$	9.000	$Var(Y_t)$	104.714	11.635
	4	0.2	0.1	0.7	1	$m(\overline{Y})$	8.999	$m(S_Y^2)$	104.773	
4	m	ω_0	α_{01}	β_{01}	σ_0^2	$E(Y_t)$	8.000	$Var(Y_t)$	137.037	17.130
	2	0.3	0.3	0.5	1	$\overline{Y}$	7.944	S_Y^2	133.361	
5	m	ω_0	α_{01}	β_{01}	σ_0^2	$E(Y_t)$	19.000	$Var(Y_t)$	1198.00	63.053
	9	0.1	0.3	0.6	1	$\overline{Y}$	19.134	S_Y^2	1169.05	
6	m	ω_0	α_{01}	β_{01}	σ_0^2	$E(Y_t)$	30.333	$Var(Y_t)$	2791.45	92.026
	9	0.9	0.5	0.2	1	$\overline{Y}$	30.274	S_Y^2	2857.81	

Table S.1. Sample and theoretical means and variances for 1000 series generated from the MthINGARCH$(1,1)$ model with (m,θ_0',σ_0^2), $n=500$ and $\varepsilon_t \sim \mathcal{P}(1)$.

2 Additional simulation results

2.1 Simulation results for the MthINARCH(2) model

n 300		PQMLE	NBQMLE	EQMLE	CLSE	1WLSE	2WLSE
	Mean	0.5416	0.5098	0.5087	0.6790	0.5108	0.5105
ω_0	StD	0.0840	0.0717	0.0718	0.1600	0.0695	0.0702
0.5	ASE	0.0854	0.0661	0.0660	0.1939	0.0668	0.0661
	RMSE	0.0937	0.0723	0.0723	0.2399	0.0704	0.0709
	Mean	0.3758	0.3946	0.3954	0.3343	0.3935	0.3937
α_{01}	StD	0.0908	0.0794	0.0796	0.1233	0.0775	0.0776
0.4	ASE	0.0875	0.0718	0.0719	0.1276	0.0721	0.0717
	RMSE	0.0940	0.0795	0.0797	0.1397	0.0778	0.0778
	Mean	0.2748	0.2902	0.2908	0.2381	0.2898	0.2899
α_{02}	StD	0.0854	0.0726	0.0727	0.1228	0.0699	0.0699
0.3	ASE	0.0742	0.0636	0.0637	0.1081	0.0639	0.0636
	RMSE	0.0890	0.0733	0.0732	0.1375	0.0707	0.0707
	Mean	0.9948	0.9847	0.9847	1.0542	0.9873	0.9850
σ_0^2	StD	0.1230	0.1246	0.1248	0.4923	0.1236	0.1248
1	ASE	0.1234	0.1224	0.1225	0.4127	0.1227	0.1224
	RMSE	0.1231	0.1255	0.1258	0.4953	0.1243	0.1257

Table S.2. QML and WLS estimation results for MthINARCH(2) series with $n = 300,\ m = 10,\ \theta_0 = (0.5, 0.4, 0.3)'$, $\varepsilon_t \sim \mathcal{P}(1)$, and $\sigma_0^2 = 1$.

n 600		PQMLE	NBQMLE	EQMLE	CLSE	1WLSE	2WLSE
	Mean	0.5226	0.5036	0.5032	0.6392	0.5041	0.5040
ω_0	StD	0.0617	0.0495	0.0495	0.1175	0.0492	0.0492
0.5	ASE	0.0633	0.0469	0.0469	0.1751	0.0473	0.0469
	RMSE	0.0657	0.0496	0.0496	0.1822	0.0494	0.0493
	Mean	0.3874	0.3995	0.3998	0.3493	0.3989	0.3991
α_{01}	StD	0.0657	0.0540	0.0540	0.1069	0.0533	0.0534
0.4	ASE	0.0661	0.0511	0.0512	0.1133	0.0514	0.0511
	RMSE	0.0669	0.0539	0.0540	0.1183	0.0534	0.0533
	Mean	0.2866	0.2950	0.2952	0.2558	0.2949	0.2947
α_{02}	StD	0.0602	0.0468	0.0469	0.1080	0.0459	0.0462
0.3	ASE	0.0570	0.0453	0.0453	0.0952	0.0456	0.0452
	RMSE	0.0616	0.0471	0.0471	0.1167	0.0461	0.0464
	Mean	1.0095	1.0015	1.0015	1.0199	1.0027	1.0017
σ_0^2	StD	0.0822	0.0817	0.08167	0.2202	0.0823	0.0818
1	ASE	0.0903	0.0892	0.0891	0.1069	0.0892	0.0890
	RMSE	0.0827	0.0818	0.0817	0.2211	0.0824	0.0818

Table S.3. QML and WLS estimation results for MthINARCH(2) series with $n = 600,\ m = 10,\ \theta_0 = (0.5, 0.4, 0.3)'$, $\varepsilon_t \sim \mathcal{P}(1)$, and $\sigma_0^2 = 1$.

n 1000		PQMLE	NBQMLE	EQMLE	CLSE	1WLSE	2WLSE
	Mean	0.5174	0.5038	0.5035	0.6173	0.5043	0.5042
ω_0	StD	0.0467	0.0372	0.0373	0.1055	0.0370	0.0370
0.5	ASE	0.0502	0.0364	0.0365	0.1589	0.0367	0.0364
	RMSE	0.0498	0.0375	0.0375	0.1578	0.0371	0.0373
	Mean	0.3917	0.3986	0.3988	0.3617	0.3984	0.3984
α_{01}	StD	0.0536	0.0406	0.0406	0.0974	0.0405	0.0403
0.4	ASE	0.0526	0.0398	0.0399	0.1004	0.0400	0.0397
	RMSE	0.0542	0.0406	0.0406	0.1046	0.0405	0.0403
	Mean	0.2887	0.2967	0.2970	0.2587	0.2964	0.2965
α_{02}	StD	0.0471	0.0358	0.0358	0.0947	0.0355	0.0355
0.3	ASE	0.0450	0.0356	0.0355	0.0838	0.0355	0.0353
	RMSE	0.0485	0.0359	0.0360	0.1033	0.0357	0.0356
	Mean	1.0040	0.9975	0.9975	1.0399	0.9986	1.0031
σ_0^2	StD	0.0683	0.0658	0.0654	0.2783	0.0655	0.0653
1	ASE	0.0689	0.0678	0.0678	0.1425	0.0681	0.0678
	RMSE	0.0684	0.0654	0.0654	0.2812	0.0656	0.0654

Table S.4. QML and WLS estimation results for MthINARCH(2) series with $n = 1000$, $m = 10$, $\theta_0 = (0.5, 0.4, 0.3)'$, $\varepsilon_t \sim \mathcal{P}(1)$, and $\sigma_0^2 = 1$.

2.2 Simulation results for the MthINGARCH$(2,1)$ with Poisson innovation

n 300		PQMLE	NBQMLE	EQMLE	CLSE	1WLSE	2WLSE
	Mean	0.7163	0.6958	0.6934	0.7548	0.6993	0.7050
ω	StD	0.1864	0.1777	0.1779	0.2232	0.1753	0.1696
0.7	ASE	0.2095	0.1839	0.1832	0.3073	0.1865	0.1859
	RMSE	0.1871	0.1778	0.1781	0.2300	0.1754	0.1697
	Mean	0.2860	0.2959	0.2967	0.2675	0.2952	0.2958
α_1	StD	0.0839	0.0775	0.0775	0.1054	0.0766	0.0750
0.3	ASE	0.0769	0.0705	0.0707	0.0978	0.0713	0.0706
	RMSE	0.0851	0.0776	0.0776	0.1103	0.0768	0.0751
	Mean	0.1737	0.1832	0.1843	0.1558	0.1839	0.1856
α_2	StD	0.0933	0.0939	0.0946	0.1044	0.0900	0.0897
0.2	ASE	0.0949	0.0849	0.0849	0.1298	0.0849	0.0848
	RMSE	0.0970	0.0954	0.0959	0.1134	0.0914	0.0908
	Mean	0.1863	0.1887	0.1893	0.1881	0.1885	0.1840
β_1	StD	0.1568	0.1509	0.1511	0.1720	0.1503	0.1405
0.2	ASE	0.1750	0.1475	0.1465	0.2639	0.1496	0.1461
	RMSE	0.1583	0.1522	0.1524	0.1757	0.1513	0.1409
	Mean	0.9930	0.9816	0.9813	1.0215	0.9866	0.9836
σ_0^2	StD	0.1164	0.1158	0.1159	0.1365	0.1158	0.1151
1	ASE	0.1327	0.1294	0.1292	0.1410	0.1307	0.1293
	RMSE	0.1166	0.1173	0.1174	0.1382	0.1166	0.1163

Table S.5. QML and WLS estimation results for MthINGARCH$(2,1)$ series with $n = 300,\ m = 4,\ \theta_0 = (0.7, 0.3, 0.2, 0.2)'$, $\varepsilon_t \sim \mathcal{P}(1)$, and $\sigma_0^2 = 1$.

2.3 Simulation results for the MthINGARCH$(1,1)$ with Negative binomial innovation

n 300		PQMLE	NBQMLE	EQMLE	CLSE	1WLSE	2WLSE
	Mean	0.3193	0.3105	0.3082	0.3743	0.2970	0.3051
ω	StD	0.0895	0.0874	0.0885	0.1328	0.0862	0.0843
0.3	ASE	0.0914	0.0808	0.0806	0.1590	0.0836	0.0819
	RMSE	0.0852	0.0871	0.0876	0.1396	0.0882	0.0858
	Mean	0.3646	0.3947	0.3971	0.3052	0.3883	0.3918
α_1	StD	0.1114	0.1132	0.1138	0.1192	0.1117	0.1117
0.4	ASE	0.1119	0.1041	0.1050	0.1337	0.1066	0.1040
	RMSE	0.1169	0.1133	0.1138	0.1523	0.1123	0.1120
	Mean	0.1896	0.1977	0.2002	0.1886	0.1915	0.1848
β_1	StD	0.1522	0.1437	0.1448	0.1944	0.1468	0.1404
0.2	ASE	0.1454	0.1288	0.1289	0.2405	0.1310	0.1292
	RMSE	0.1526	0.1437	0.1448	0.1948	0.1471	0.1412
	Mean	1.9691	1.9435	1.9418	1.9860	1.9561	1.9485
σ_0^2	StD	0.3623	0.3465	0.3452	0.4808	0.3566	0.3481
2	ASE	0.4665	0.4572	0.4571	0.5049	0.4655	0.4554
	RMSE	0.3636	0.3511	0.3501	0.4810	0.3593	0.3519

Table S.6. QML and WLS estimation results for MthINGARCH$(1,1)$ series with $n = 300,\ m = 6,\ \theta_0 = (0.3, 0.4, 0.2)'$, $\varepsilon_t \sim \mathcal{NB}(1, \frac{1}{2})$, and $\sigma_0^2 = 2$.