Random multiplication versus random sum: auto-regressive-like models with integer-valued random inputs

Aknouche, Abdelhakim and Gouveia, Sonia and Scotto, Manuel

Qassim University, Saudi Arabia, University of Aveiro, Portugal, University of Lisbon, Portugal

18 December 2023

Random multiplication versus random sum: auto-regressive-like models with integer-valued random inputs

Abdelhakim Aknouche

Department of Mathematics, College of Science

Qassim University, Al Qassim, Saudi Arabia

email: aknouche_ab@yahoo.com

Sónia Gouveia*

Institute of Electronics and Informatics Engineering of Aveiro (IEETA)

Department of Electronics, Telecommunications and Informatics (DETI)

University of Aveiro, Portugal

email: sonia.gouveia@ua.pt

Manuel G. Scotto

Center for Computational and Stochastic Mathematics (CEMAT)

Department of Mathematics, IST

University of Lisbon, Portugal

email: manuel.scotto@tecnico.ulisboa.pt

---

*corresponding author

Abstract

A common approach to analyze count time series is to fit models based on random sum operators. As an alternative, this paper introduces time series models based on a random multiplication operator, which is simply the multiplication of a variable operand by an integer-valued random coefficient, whose mean is the constant operand. Such operation is endowed into auto-regressive-like models with integer-valued random inputs, addressed as RMINAR. Two special variants are studied, namely the $\mathbb{N}_0$-valued random coefficient auto-regressive model and the $\mathbb{N}_0$-valued random coefficient multiplicative error model. Furthermore, $\mathbb{Z}$-valued extensions are considered. The dynamic structure of the proposed models is studied in detail. In particular, their corresponding solutions are everywhere strictly stationary and ergodic, a fact that is not common neither in the literature on integer-valued time series models nor real-valued random coefficient auto-regressive models. Therefore, the parameters of the RMINAR model are estimated using a four-stage weighted least squares estimator, with consistency and asymptotic normality established everywhere in the parameter space. Finally, the new RMINAR models are illustrated with some simulated and empirical examples.

Keywords: integer-valued random coefficient AR, random multiplication integer-valued auto-regression, random multiplication operator, RMINAR, WLS estimators.

MSC: 60G10, 62M10, 62M20, 62F12, 62F30.

1   Introduction

Modeling low integer-valued time series data is, nowadays, an ongoing concern in time series research. To this end, three common approaches are generally undertaken. The first assumes a discrete conditional distribution whose conditional mean is a parametric function of past observations. The resulting models are known as observation-driven in the terminology of Cox (1981). The most known examples are the integer-valued Generalized ARMA model (Zeger and Qaqish, 1988; Benjamin *et al.*, 2003; Zheng *et al.*, 2015) and, in particular, the integer-valued GARCH (INGARCH) model (e.g. Reydberg and Shephard, 2000; Heinen,

2003; Ferland *et al.*, 2006; Fokianos *et al.*, 2009; Zhu, 2011; Davis and Liu, 2016; Aknouche and Francq 2021). The primary feature of observation-driven models is that the likelihood function is explicit in terms of observations, which turns maximum likelihood estimation, inference, and forecasting quite easy to investigate. Other M-estimation methods such as quasi-maximum likelihood (QML) and weighted least squares (WLS) estimators are also straightforward to derive. However, observation-driven models suffer from some limitations, namely that they are often fully parametric and hence not robust to a distributional misspecification. Moreover, their probabilistic structures (e.g. ergodicity, tail behavior, extremal properties) are inherently complex since they are not defined through equations driven by independent and identically distributed (iid) innovation sequences.

The second approach is addressed as parameter-driven (Cox, 1981) and shares with the first approach the fact that also requires the specification of a discrete conditional distribution for the data. Nonetheless, this distribution is rather conditioned on a latent process since the conditional mean has a proper dynamics in terms of its past (latent) values (Zeger, 1988; Davis and Rodriguez-Yam, 2005; Davis and Wu, 2009). Parameter-driven models have in general simple probability structures and are quite flexible to represent dynamic dependence of count data (e.g. Davis and Dunsmuir, 2016). In particular, the conditional mean depends on present shocks unlike INGARCH models for which the conditional intensity only depends on past observations. Due to the latent process, however, their estimation is rather difficult because the likelihood function cannot in general be obtained in a closed form and involves cumbersome multidimensional integration. Moreover, QML and WLS estimators are not simple to derive either, since the conditional mean is not explicit in terms of observations. This difficulty also arises in inference and prediction which explains why parameter-driven count models have received less attention than observation-driven ones.

The third approach considers appropriate stochastic difference equations driven by iid inputs whose solutions are integer-valued sequences. The main concern of these models is to handle integer-valued random operations on inputs to produce integer-valued outputs that have similar features to real-world integer-valued data. Random sum operations, aka

thinning operations, are the best-known examples. The rule is that, given a positive constant and an integer-valued random variable as operands, a random sum operation is the sum of iid discrete variables whose mean and number are the constant and variable operands, respectively. In particular, the binomial thinning operator (Steutel and van Harn, 1978) produces a binomial distributed variable that is bounded from above by the operand variable. Random sum operators not satisfying the latter feature are still called generalized thinning, e.g. Poisson and negative binomial random sums (see e.g. Scotto *et al.*, 2015).

The most elegant property of random sum-based equations is that the marginal distribution of the output sequence is readily known and depends on the operator and input distributions. For instance, the first order integer-valued auto-regressive process (INAR(1); McKenzie, 1985; Al-Osh and Alzaid, 1987) based on binomial thinning and driven by a Poisson iid innovation has a Poisson marginal. Since any random sum involves unobserved summands, the likelihood calculation is cumbersome and requires high dimensional summations, just like parameter-driven models. This is the reason why some authors classify thinning-based models as parameter-driven (e.g. Ahmad and Francq, 2016; Aknouche and Francq, 2023). Furthermore, the conditional mean and variance are parametric functions of past observations as in observation-driven models. In fact, the latent variables in a random sum only intervene in model's conditional distribution through their mean and variance as being unknown constant parameters. The same happens for random coefficient auto-regressive models (RCAR; Nicholls and Quinn, 1982). In this respect, thinning-based models are more similar to observation-driven models. In particular, QML and WLS estimators are quite easy to derive and analyze. Thus, thinning-based and RCAR models can be seen as semi-(or partially-) observation-driven models.

Numerous thinning-based count models have been emerged so far (e.g. Scotto *et al.*, 2015; Weiss, 2018). They differ fundamentally in the assumed distributions of the summands or/and the form of the stochastic equations. Note that the implicit form of the random sum in terms of the operand variable makes the study of the corresponding equation more complex than in conventional stochastic equations (e.g. ARMA, GARCH, RCAR).

Most existing thinning specifications are based on simple pure auto-regressions (INAR) or moving averages (INMA) with low orders. Only a few research works deal with integer-valued ARMA (INARMA) equations or their multivariate forms due to their inherent complexity. In particular, invertibility, tails behavior, asymptotic properties of maximum likelihood estimators, and forecasting remain an issue for INARMA-like models. Many other statistical aspects of INARMA models are not as developed as in standard continuous-valued ARMA models. Also, non-linear forms are not explored as much except in a few special cases. This is why most thinning-based models are unable to reproduce various interesting features such as high over-dispersion, multi-modality, persistence, etc. (e.g. Aknouche and Scotto, 2024).

As a simple alternative to random sum operators, this paper proposes a random multiplication operator and shows how to build on it simple and analytically tractable integer-valued time series models. An $\mathbb{N}_0$-valued random multiplication operator is in fact a random sum with identical summands and constitutes the direct multiplication of an operand variable by an integer-valued variable whose expected value (defined in $\mathbb{R}^+$) is precisely the constant operand. An extended $\mathbb{Z}$-valued random multiplication makes it possible to deal with the constant and variable operands defined in $\mathbb{R}$ and $\mathbb{Z}$, respectively. Compared to thinning operators, the random multiplication is analytically simpler and can produce variables with higher volatility even with Poisson multipliers. Actually, a random multiplication-based model is nothing but a model with integer-valued random inputs (coefficients and innovations). In particular, a random multiplication-based auto-regressive model with $\mathbb{N}_0$-valued inputs (henceforth RMINAR) is a special case of the RCAR model but with $\mathbb{N}_0$-valued random inputs. Likewise, a $\mathbb{Z}$-valued RMINAR model is an RCAR with $\mathbb{Z}$-valued random inputs.

Continuous-valued RCAR models have been widely studied since the late 1970s and most of their probabilistic and statistical properties are now well understood (e.g. Nicholls and Quinn, 1982; Tsay, 1987; Schick, 1996; Diaconis and Freedman, 1999; Aue *et al.*, 2006; Aknouche, 2013-2015; Aue and Horvath, 2019; Trapani, 2021, Regis *et al.*, 2022). Although the proposed RMINAR model belongs to the general class of RCAR models, it holds sur-

prising features that differ from those of the continuous-valued counterpart and of the afore-mentioned integer-valued models. In particular, any RMINAR solution is strictly stationary and ergodic regardless of its coefficients values, so it is useless testing strict stationary as in the continuous-valued case (Aue and Horvath, 2011; Aknouche, 2013-2015). In other words, RMINAR models are everywhere stationary and ergodic and thus can be strictly stationary with infinite means. As a consequence, multi-stage WLS and QML estimators are consistent and asymptotically Normal everywhere in the parameter space and for all parameter components.

The rest of this paper is organized as follows. Section 2 sets up the random multiplication operator and shows its main properties. Section 3 defines the RMINAR model for $\mathbb{N}_0$-valued data, a $\mathbb{Z}$-valued extension, and a multiplicative variant for $\mathbb{N}_0$-valued data. Section 4 proposes four-stage WLS estimators (4SWLSE) for the mean and variance of the random inputs in the three models. A simulation study and two real applications with a $\mathbb{N}_0$-valued and a $\mathbb{Z}$-valued time series are given in Section 5. Section 6 summarizes the conclusions of this work and the main proofs are left to an appendix.

2 Random operators: sum versus multiplication

This section overviews important properties of the random sum operator (RSO), denoted by $\circ_s$, and introduces the so-called random multiplication operator (RMO), denoted by $\odot_m$, highlighting fundamental differences between these operators. A few examples are given for the distribution of the random variable resulting from these random operations.

2.1 Random sum operation

Random sum operators also known as generalized thinning (Latour, 1998) are commonly used in branching and INAR-like processes. In its general form a RSO, denoted by $\circ_s$, is

defined for any positive constant α and any integer-valued random variable X by

$$\alpha \circ_s X := \sum_{i=1}^{X} \xi_i, \tag{1}$$

where the integer-valued sequence (ξ_i) is iid with mean $\mathbb{E}(\xi_i) = \alpha$ and variance $\sigma_\xi^2 := \mathbb{V}(\xi_i)$. It is further assumed that X and (ξ_i) are independent. The terms in (ξ_i) are called summands or counting series (or also offspring sequence) while α and X are the constant and variable operands, respectively. Thus, the operation $\alpha \circ_s X$ is a random sum up to the operand variable X where the common mean of the iid summands (ξ_i) is the constant operand α. The first two moments of $\alpha \circ_s X$ are

$$\mathbb{E}(\alpha \circ_s X) = \alpha \mathbb{E}(X) \quad \text{and} \quad \mathbb{E}(\alpha \circ_s X | X) = \alpha X, \tag{2}$$

$$\mathbb{V}(\alpha \circ_s X) = \sigma_\xi^2 \mathbb{E}(X) + \alpha^2 \mathbb{V}(X) \quad \text{and} \quad \mathbb{V}(\alpha \circ_s X | X) = \sigma_\xi^2 X. \tag{3}$$

An interesting property of the $\circ_s$ is that it generally inherits its distribution from those of (ξ_i) and X. Indeed, the probability generating function of $\alpha \circ_s X$ is

$$G_{\alpha \circ_s X}(z) := E\left(z^{\alpha \circ_s X}\right) = G_X\left(G_{\xi_1}(z)\right),$$

which shows that the probability law of $\alpha \circ_s X$ is uniquely determined from those of (ξ_i) and X. The most commonly used RSO are as follows.

Example 2.1 i) *Binomial thinning*: The Bernoulli random sum, aka the binomial thinning operator, and being denoted by $\circ$ (Steutel and van Harn, 1978), assumes (ξ_i) to be Bernoulli distributed with mean $\alpha \in (0,1)$. This implies that $\alpha \circ X | X \sim B(X, \alpha)$ is conditionally binomial distributed and $\alpha \circ X$ is stochastically smaller than X, that is

$$\alpha \circ X \leq X, \tag{4}$$

hence the term thinning. For this operator,

$$\mathbb{V}(\alpha \circ X) = \alpha(1-\alpha)\mathbb{E}(X) + \alpha^2 \mathbb{V}(X) \quad \text{and} \quad \mathbb{V}(\alpha \circ X | X) = \alpha(1-\alpha)X.$$

ii) *Poisson random sum*: When (ξ_i) are Poisson distributed with parameter $\alpha > 0$, the random sum $\circ_s$ is known as the Poisson generalized thinning and satisfies $\alpha \circ_s X | X \sim \mathcal{P}(\alpha X)$

where $\mathcal{P}(\alpha X)$ stands for the Poisson distribution with mean αX. In this case, the range of $\alpha \circ_s X \in \mathbb{N}_0 := \{0, 1, \dots\}$ is the set of integers and the above thinning property (4) is no longer satisfied when $\alpha > 1$, so the term thinning makes no sense. In fact, as the range of $\alpha \circ_s X$ is larger than that of X, the Poisson sum operator cannot be seen as a thinning operator in the sense of (4). The variances in (3) reduce to

$$\mathbb{V}(\alpha \circ_s X) = \alpha\mathbb{E}(X) + \alpha^2\mathbb{V}(X) \quad \text{and} \quad \mathbb{V}(\alpha \circ_s X|X) = \alpha X.$$

iii) *Negative binomial random sum*: If (ξ_i) are geometric distributed with parameter $\frac{1}{1+\alpha}$, the geometric random sum $\circ_s$ is known as the negative binomial generalized thinning and $\alpha \circ_s X|X \sim \mathcal{NB}(X, \frac{1}{1+\alpha})$ is the negative binomial distribution with dispersion X and probability $\frac{1}{1+\alpha}$. As a consequence $\alpha \circ_s X \in \mathbb{N}_0$.

2.2 Random multiplication operation

Let Φ and X be two independent non-negative integer-valued random variables, and ϕ a positive constant representing $\mathbb{E}(\Phi) = \phi > 0$. The random multiplication operation (RMO) is defined as

$$\phi \odot_m X := \Phi X = \sum_{i=1}^{X} \Phi = \sum_{i=1}^{\Phi} X. \tag{5}$$

Equality (5) is satisfied for all $s \in \Omega$, assuming that all variables and sequences are defined on a probability space $(\Omega, \mathcal{F}, P))$, and

$$[\phi \odot_m X](s) = \left[\sum_{i=1}^{X} \Phi\right](s) = \left[\sum_{i=1}^{\Phi} X\right](s) = [\Phi X](s).$$

Naturally, these everywhere or sure equalities (in the sense of being satisfied $\forall s \in \Omega$) imply the almost sure equalities

$$\sum_{i=1}^{X} \Phi \stackrel{a.s.}{=} \sum_{i=1}^{\Phi} X \stackrel{a.s.}{=} \Phi X,$$

which in turn implies that the equalities also hold in distribution. The converse, however, is not true. The sure equalities in (5) are necessary to establish the almost sure convergence of the estimators in Section 4. The $\phi \odot_m X$ operation can also be seen as a random sum (1),

8

with the restriction of having identical summands, designated by Φ. The mean and variance of $\phi \odot_m X$ are given by

$$\mathbb{E}\left(\phi \odot_m X\right) = \phi\mathbb{E}\left(X\right) \quad \text{and} \quad \mathbb{E}\left(\phi \odot_m X | X\right) = \phi X, \tag{6}$$

$$\mathbb{V}\left(\phi \odot_m X\right) = \sigma_\Phi^2\mathbb{E}\left(X^2\right) + \phi^2\mathbb{V}\left(X\right) \quad \text{and} \quad \mathbb{V}\left(\phi \odot_m X | X\right) = \sigma_\Phi^2 X^2. \tag{7}$$

While the operations $\alpha \circ_s X$ and $\phi \odot_m X$ share the same conditional mean structure, the conditional variance is proportional to the operand variable X for the RSO (3) and is proportional to X^2 for the RMO (7). The variability implied by (7) is therefore allowed to be larger than that in (3). Moreover, the range of the random multiplication $\phi \odot_m X$ is

$$\phi \odot_m X | (\Phi, X) \in \{0, \dots, \Phi X\},$$

so that $\odot_m$ is not a proper thinning operation in the sense of (4).

The random multiplication (5) is much simpler and more tractable than the random sum (1) since it consists of a direct multiplication by an integer-valued random coefficient. However, what is gained in simplicity is lost in distributional reproducibility because the distribution of $\alpha \circ_s X$ is readily known while that of $\phi \odot_m X$ is in general not usual. In fact, although many well-known discrete distributions are stable under independent summation (e.g. binomial, Poisson and negative binomial), this is not the case for the multiplication operation. For example, the distribution of the product of two independent Poisson variables is not Poisson distributed. This makes the distribution of $\phi \odot_m X$ quite unusual except for special cases (e.g. the class of Bernoulli distributions is stable under multiplication). For other cases, the distributional properties of $\phi \odot_m X$ could easily be explored using direct methods where specific distributions for the random operand Φ in the $\phi \odot_m$ operation are of interest, as in the following examples.

Example 2.2 i) *Binomial random multiplication*: For a binomial distributed $\Phi \sim \mathcal{B}\left(r, \frac{\phi}{r}\right)$ with $r \geq 1$, the operation $\odot_m$ is called binomial multiplication. The variances in (7) translate into

$$\mathbb{V}\left(\phi \odot_m X\right) = \phi\left(1 - \tfrac{\phi}{r}\right)\mathbb{E}\left(X^2\right) + \phi^2\mathbb{V}\left(X\right) \quad \text{and} \quad \mathbb{V}\left(\phi \odot_m X | X\right) = \phi\left(1 - \tfrac{\phi}{r}\right)X^2.$$

ii) *Poisson random multiplication*: When $\Phi \sim \mathcal{P}(\phi)$ is Poisson distributed, the operation $\odot_m$ is called Poisson multiplication. In this case, the variances in (7) simplify to

$$\mathbb{V}(\phi \odot_m X) = \phi \mathbb{E}(X^2) + \phi^2 \mathbb{V}(X) \quad \text{and} \quad \mathbb{V}(\phi \odot_m X|X) = \phi X^2.$$

Note that if $X \sim \mathcal{P}(\lambda)$, then $\mathbb{V}(\phi \odot_m X) = \phi(\lambda + \lambda^2) + \phi^2 \lambda$ is allowed to be larger than $\mathbb{V}(\alpha \circ_s X) = \alpha\lambda + \alpha^2\lambda$ implied by (3) with $\alpha \circ_s X|X \sim \mathcal{P}(\alpha X)$ and $X \sim \mathcal{P}(\lambda)$.

iii) *Negative binomial I random multiplication*: Assuming that $\Phi \sim \mathcal{NB}\left(r\phi, \frac{r}{r+1}\right)$ is negative binomial distributed (denoted by NB1, see Aknouche *et al.*, 2018), the operation $\odot_m$ is called NB1 multiplication. The variances in (7) are thus

$$\mathbb{V}(\phi \odot_m X) = \phi\left(1 + \tfrac{1}{r}\right)\mathbb{E}(X^2) + \phi^2\mathbb{V}(X) \quad \text{and} \quad \mathbb{V}(\phi \odot_m X|X) = \phi\left(1 + \tfrac{1}{r}\right)X^2.$$

iv) *Negative binomial II random multiplication*: If $\Phi \sim \mathcal{NB}\left(r, \frac{r}{r+\phi}\right)$ $(r > 0)$ has a negative binomial distribution (usually denoted by NB2, see Aknouche *et al.*, 2018; Aknouche and Francq 2021), then the operation $\odot_m$ is called NB2 multiplication. In this case, the variances in (7) become

$$\mathbb{V}(\phi \odot_m X) = \phi\left(1 + \tfrac{1}{r}\phi\right)\mathbb{E}(X^2) + \phi^2\mathbb{V}(X) \quad \text{and} \quad \mathbb{V}(\phi \odot_m X|X) = \phi\left(1 + \tfrac{1}{r}\phi\right)X^2.$$

Other constructions or extensions of the RMO can be considered including e.g. a $\mathbb{Z}$-valued extension of the random multiplication in (5) simply by assuming that X and Φ are $\mathbb{Z}$-valued random variables and $\phi \in \mathbb{R}$.

3 Random multiplication based auto-regressive models

This section presents a few examples of auto-regressive-like models constructed from the new RMO. These models will be addressed as random multiplication INAR of order p and designated as RMINAR(p).

3.1 $\mathbb{N}_0$-valued random multiplication AR (RMINAR) model

Let $\{\Phi_{it}, \ t \in \mathbb{Z}\}$ $(i = 1, \ldots, p)$ and $\{\varepsilon_t, \ t \in \mathbb{Z}\}$ be mutually independent $\mathbb{N}_0$-valued iid sequences with $\phi_i := \mathbb{E}(\Phi_{it}) > 0$, $\sigma_{\phi_i}^2 := \mathbb{V}(\Phi_{it}) > 0$ $(i = 1, \ldots, p)$, $\mu_\varepsilon := \mathbb{E}(\varepsilon_t) > 0$, and

$\sigma_\varepsilon^2 := \mathbb{V}(\varepsilon_t) > 0$. A $\mathbb{N}_0$-valued process $\{Y_t, \ t \in \mathbb{Z}\}$ is said to be an integer-valued random multiplication AR model, in short RMINAR(p), if Y_t admits the representation

$$Y_t = \sum_{i=1}^p \Phi_{it} Y_{t-i} + \varepsilon_t, \ \ t \in \mathbb{Z}. \tag{8}$$

Note that the RMINAR(p) model (8) can be rewritten as follows

$$Y_t = \sum_{i=1}^p \phi_i \odot_m Y_{t-i} + \varepsilon_t, \ \ t \in \mathbb{Z}.$$

The distribution of the input sequences $\{\Phi_{it}, \ t \in \mathbb{Z}\}$ $(i = 1, \ldots, p)$ and $\{\varepsilon_t, \ t \in \mathbb{Z}\}$ can be specified as in Example 2.2 (binomial, Poisson, negative binomial, etc.). Thus, (8) can be seen as a random coefficient AR model (RCAR) in the sense of Nicholls and Quinn (1982) but with $\mathbb{N}_0$-valued random inputs $\{\Phi_{it}, \ t \in \mathbb{Z}\}$ $(i = 1, \ldots, p)$ and $\{\varepsilon_t, \ t \in \mathbb{Z}\}$. The conditional mean and variance of (8) are given by

$$\mathbb{E}\left(Y_t | \mathcal{F}_{t-1}^Y\right) = \sum_{i=1}^p \phi_i Y_{t-i} + \mu_\varepsilon \ \text{ and } \ \mathbb{V}\left(Y_t | \mathcal{F}_{t-1}^Y\right) = \sum_{i=1}^p \sigma_{\phi_i}^2 Y_{t-i}^2 + \sigma_\varepsilon^2, \tag{9}$$

where $\mathcal{F}_t^Y$ is the σ-algebra generated by $\{Y_{t-u}, u \geq 0\}$. From (9) it is clear that the RMINAR(p) model allows both conditional overdispersion and underdispersion. E.g. when the inputs Φ_{it} and ε_t are Poisson distributed, then (9) becomes

$$\mathbb{E}\left(Y_t | \mathcal{F}_{t-1}^Y\right) = \sum_{i=1}^p \phi_i Y_{t-i} + \mu_\varepsilon \ \text{ and } \ \mathbb{V}\left(Y_t | \mathcal{F}_{t-1}^Y\right) = \sum_{i=1}^p \phi_i Y_{t-i}^2 + \mu_\varepsilon,$$

and $\mathbb{V}\left(Y_t | \mathcal{F}_{t-1}^Y\right) > \mathbb{E}\left(Y_t | \mathcal{F}_{t-1}^Y\right)$ as $Y_t \leq Y_t^2$ for all t. Also, (9) can lead to $\mathbb{V}\left(Y_t | \mathcal{F}_{t-1}^Y\right) < \mathbb{E}\left(Y_t | \mathcal{F}_{t-1}^Y\right)$ e.g. when the inputs are binomial distributed with appropriate parameters.

Note that for $p = 1$ the model (8) reduces to

$$Y_t = \Phi_t Y_{t-1} + \varepsilon_t, \ \ t \in \mathbb{Z},$$

which is a homogeneous Markov chain with transition probabilities given by

$$
\begin{aligned}
\mathbb{P}\left(Y_t = j | Y_{t-1} = i\right) &= \mathbb{P}\left(\Phi_t Y_{t-1} + \varepsilon_t = j | Y_{t-1} = i\right) \\
&= \mathbb{P}\left(\Phi_t i + \varepsilon_t = j\right) \\
&= \begin{cases} \sum_{k \leq j, \frac{j-k}{i} \in \mathbb{N}_0} \mathbb{P}\left(\varepsilon_t = k\right) \mathbb{P}\left(\Phi_t = \frac{j-k}{i}\right) &, i > 0 \\ \mathbb{P}\left(\varepsilon_t = j\right) &, i = 0 \end{cases}.
\end{aligned}
$$

The novel RMINAR(p) model (8) can be compared to the random coefficient INAR model of Zheng *et al.* (2006, 2007), denoted as RCINAR(p) and defined as

$$X_t = \sum_{i=1}^{p} \alpha_{it} \circ X_{t-i} + \zeta_t, \quad t \in \mathbb{Z}, \tag{10}$$

where " $\circ$ " is the binomial thinning operator whereas $\{\alpha_{it}, \ t \in \mathbb{Z}\}$ $(i = 1, \ldots, p)$ and $\{\zeta_t, \ t \in \mathbb{Z}\}$ are mutually independent iid sequences valued in $(0, 1)$ and $\mathbb{N}_0$, respectively. Models in (8) and (10) share the same conditional mean structure and similar quadratic conditional variances in terms of past observations. Note that (10) reduces to the INAR(p) model of Du and Li (1991) (see also McKenzie, 1985; Al-Osh and Alzaid, 1987 for the case $p = 1$) and thus the conditional mean and variance of (10) are

$$\mathbb{E}\left(X_t|\mathcal{F}_{t-1}^X\right) = \sum_{i=1}^{p} \alpha_i \circ X_{t-i} + \zeta \quad \text{and} \quad \mathbb{V}\left(X_t|\mathcal{F}_{t-1}^X\right) = \sum_{i=1}^{p} \sigma_{\alpha_i}^2 X_{t-i}^2 + \left(\alpha_i\left(\alpha_i - 1\right) - \sigma_{\alpha_i}^2\right) X_{t-i} + \sigma_\zeta^2,$$

where $\alpha_i := \mathbb{E}(\alpha_{it})$, $\sigma_{\alpha_i}^2 := \mathbb{V}(\alpha_{it})$, $\zeta := \mathbb{E}(\zeta_t)$ and $\mathbb{V}(\zeta_t) = \sigma_\zeta^2 > 0$ (Zheng *et al.*, 2006).

The probabilistic structure of (8) is already well known when the inputs are real-valued (e.g.. Nicholls and Quinn, 1982; Feigin and Tweedie, 1985; Tsay, 1987; Diaconis and Freedman, 1999). In the integer-valued case, however, there are some surprising properties. Let $\mathbf{Y}_t = (Y_t, \ldots, Y_{t-p+1})'$ and $\mathbf{\Xi}_t = (\varepsilon_t, 0, \ldots, 0)'$ be p-dimensional column vectors, and define the $p \times p$ companion matrix

$$A_t = \begin{pmatrix} \left(\Phi_{1t}, \ldots, \Phi_{(p-1)t}\right) & \Phi_{pt} \\ \mathbf{I}_{p-1} & \mathbf{0}_{(p-1)\times 1} \end{pmatrix}, \tag{11}$$

where $\mathbf{I}_p$ and $\mathbf{0}_{p\times 1}$ are the p-dimensional identity matrix and zero vector, respectively. Then model (8) can be written in the following vector form

$$\mathbf{Y}_t = A_t \mathbf{Y}_{t-1} + \mathbf{\Xi}_t, \quad t \in \mathbb{Z}. \tag{12}$$

For almost all common integer-valued time series models (observation-driven, parameter-driven, random sum based), the conditions for strict stationary and mean stationary on the auto-regressive parameter coincide. The formulation in (8), however, allows the output

process $\{Y_t, t \in \mathbb{Z}\}$ to be everywhere (i.e. for all parameter values) strictly stationary even with infinite mean. In fact, the following result shows that any solution of (8) is everywhere (i.e. universally) strictly stationary provided that $\mathbb{P}(\Phi_{it} = 0) > 0$ for all $i = 1, \ldots, p$. For simplicity in notation and readability, the latter condition is denoted as **A0**.

Theorem 1 *Under* **A0**, *the series* $\mathbf{Y}_t := \sum\limits_{j=0}^{\infty} \prod\limits_{i=0}^{j-1} A_{t-i} \mathbf{\Xi}_{t-j}$ *converges absolutely a.s. for all* $t \in \mathbb{Z}$ *and the process* $\{Y_t, t \in \mathbb{Z}\}$ *given by* $Y_t = (1, 0, \ldots, 0)' \mathbf{Y}_t$ *is the unique (causal) strictly stationary and ergodic solution to the RMINAR equation* (8).

Proof. See Appendix A. ■

Most of the usual count distributions (binomial, Poisson, negative binomial, etc.) satisfy **A0** and thus guarantee that model (8) has a (causal) strictly stationary and ergodic solution whatever the value of the inputs Φ_{it} and ε_t. Then, the RMINAR (8) is universally (or everywhere) stable with respect to stationarity and ergodicity.

Remark 1 *i) When* **A0** *is not satisfied (e.g. for the truncated geometric distribution, modeling the number of Bernoulli trials to get the first success, and taking* $\mathbb{P}(\Phi_{1t} \in \{1, 2, \ldots\}) = 1$), *assuming that* $\mathbb{E}(\log^+(\varepsilon_t)) < \infty$ *and* $\mathbb{E}(\log^+(\Phi_{it})) < \infty$ ($i = 1, \ldots, p$), *a sufficient condition for the existence of a strictly stationary and ergodic solution to* (8) *is that the largest Lyapunov exponent* γ *(Furstenberg and Kesten,* 1960; *Kesten,* 1973)*

$$\gamma = \lim_{t \to \infty} \tfrac{1}{t} \log \|A_1 A_2 \cdots A_t\| < 0 \ a.s.,$$

where $\log^+(x) := \max(\log(x), 0)$, $\|.\|$ *is an operator norm on the space of square real matrices of dimension* p *and* A_i ($i = 1, \ldots, t$) *are* $p \times p$ *random matrices.*

ii) When **A0** *is satisfied and* $p = 1$, *the Lyapunov exponent* $\mathbb{E}(\log(\Phi_{1t})) = -\infty$ *provided that* $\mathbb{E}(\log^+(\Phi_{1t})) < \infty$ *with the convention* $\max(-\infty, 0) = 0$.

Using stochastic recurrence equations theory (Kesten, 1973; Vervaat, 1979; Goldie, 1991; Grey, 1994), the tail behavior of model (8) can be easily revealed. Note that due to **A0**, the distribution of $\log(\Phi_{it})$ given $\Phi_{it} \neq 0$ is non-arithmetic for all $i = 1, \ldots, p$. Moreover,

the equation (10) admits a strictly stationary solution whatever the value of γ. Therefore, the following proposition states a result that is an obvious corollary of Theorems 3-5 of Kesten (1973) (see also Theorem 1 of Grey (1994) when $p = 1$). For this reason the proof of proposition 1 is omitted.

Proposition 1 *Consider the RMINAR model* (8) *under* ***A0*** *and* $\mathbb{E}\left(\log^{+}\left(\Phi_{it}\right)\right) < \infty$ *(i =* $1, \ldots, p$).

   i) For $p = 1$, assume that

$$\mathbb{E}\left(\Phi_{1t}^{\tau_0} \log^{+}\left(\Phi_{1t}\right)\right) < \infty \quad and \quad \mathbb{E}\left(\Phi_{1t}^{\tau_0}\right) \geq 1 \ , \ for \ some \ \tau_0 > 0. \tag{13}$$

Then there exists $\tau_1 \in [0, \tau_0]$ such that $\mathbb{E}\left(\Phi_{1t}^{\tau_1}\right) = 1$ has a unique solution. In addition, if $\mathbb{E}\left(\varepsilon_t^{\tau_1}\right) < \infty$ then

$$\mathbb{P}\left(Y_t \geq y\right) \to y^{-\tau_1}, \ as \ y \to \infty. \tag{14}$$

   ii) For $p > 1$, assume there exists $\tau_0 > 0$ such that

$$\mathbb{E}\left(\sum_{j=1}^{p} \Phi_{jt}^{\tau_0}\right) \geq p^{\frac{\tau_0}{2}}, \ \mathbb{E}\left(\|A_t\|^{\tau_0} \log^{+} \|A_t\|\right) < \infty \quad and \quad \mathbb{E}\left(\varepsilon_t^{\tau_1}\right) < \infty, \ \tau_1 \in [0, \tau_0]. \tag{15}$$

   Then for all non-negative p-vector ***x*** *such that $\|\mathbf{x}\| = 1$,*

$$\mathbb{P}\left(\mathbf{x}'\mathbf{Y}_t \geq y\right) \to y^{-\tau_1}, \ as \ y \to \infty. \tag{16}$$

Since the inputs of the RMINAR(p) model (8) have, in general, count distributions (binomial, Poisson, negative binomial, double Poisson, etc.) for which all moments exists, the conditions (13) and (15) are generally satisfied. Thus, the stationary solutions of (8) would have regularly-varying tails in the sense of (14) and (16).

While the conditions of strict stationarity, ergodicity and regular variation for the RMINAR model (8) are somewhat different from those of the real-valued RCAR case, the moment conditions for the two cases are the same. These conditions for model (8) are stated in proposition 2, the proof of which is very similar to that of Theorem 2.9 in Francq and Zakoian (2019) and hence is omitted. The reader is further referred to Feigin and Tweedie (1985) for more details (see also Tsay, 1987 and Ling, 1999 for similar models).

Proposition 2 *Suppose that* $\mathbb{E}\left(\Phi_{it}^m\right) < \infty$ *and* $\mathbb{E}\left(\varepsilon_t^m\right) < \infty$ *for some positive integer* m. *Then the RMINAR model* (8) *admits a strictly stationary and ergodic solution with* $\mathbb{E}\left(Y_t^m\right) < \infty$ *if*

$$\rho\left(\mathbb{E}\left(A_t^{\otimes m}\right)\right) < 1, \tag{17}$$

where $A^{\otimes m} = A \otimes \cdots \otimes A$ *is the Kronecker product with* m *factors. If* $\rho\left(\mathbb{E}\left(A_t^{\otimes m}\right)\right) \geq 1$ *then* $\mathbb{E}\left(Y_t^m\right) = \infty$.

From the result in (17) with $m = 1$, the RMINAR(p) model (8) admits a stationary and ergodic solution with $\mathbb{E}\left(Y_t\right) < \infty$ if

$$\rho\left(A_t\right) < 1,$$

which is equivalent to

$$\sum_{i=1}^p \phi_i < 1. \tag{18}$$

Under (18), the unconditional mean has the following expression

$$\mathbb{E}\left(Y_t\right) = \mu_\varepsilon \left(1 - \sum_{i=1}^p \phi_i\right)^{-1}, \tag{19}$$

which is the same as the mean of RCINAR(p) and AR(p) models. If $\sum_{i=1}^p \phi_i = 1$ or $\sum_{i=1}^p \phi_i > 1$ then the RMINAR(p) model (8) still admits a strictly stationary and ergodic solution $\{Y_t, t \in \mathbb{Z}\}$ (under **A0**), but with an infinite mean $\mathbb{E}\left(Y_t\right) = \infty$.

A sufficient condition for second-order stationary is given by (17) with $m = 2$. As an example, for the RMINAR(1) the condition is

$$0 \leq \mathbb{E}\left(\Phi_{1t}^2\right) = \sigma_{\phi_1}^2 + \phi_1^2 < 1,$$

and the unconditional variance of the process is

$$\mathbb{V}\left(Y_t\right) = \frac{\sigma_\varepsilon^2 + \sigma_{\phi_1}^2 \left(\mathbb{E}\left(Y_{t-1}\right)\right)^2}{1 - \left(\sigma_{\phi_1}^2 + \phi_1^2\right)}. \tag{20}$$

For any p, $\mathbb{V}\left(Y_t\right)$ can be given from (12) using the vector stacking operator *vec*. Let $A := \mathbb{E}\left(A_t\right)$, $\boldsymbol{\Xi} := \mathbb{E}\left(\boldsymbol{\Xi}_t\right)$, $\boldsymbol{\mu} := \mathbb{E}\left(\mathbf{Y}_t\right)$, $\Gamma_Y := \mathbb{V}(\mathbf{Y}_t) = \mathbb{E}\left(\mathbf{Y}_t - \boldsymbol{\mu}\right)\left(\mathbf{Y}_t - \boldsymbol{\mu}\right)'$ and $\Gamma_{\boldsymbol{\Xi}} := \mathbb{V}\left(\boldsymbol{\Xi}_t\right)$.

Then

$$vec(\Gamma_Y) = (I - E(A_t \otimes A_t))^{-1}((E(A_t \otimes A_t) - (A \otimes A))\,vec(\boldsymbol{\mu}\boldsymbol{\mu}') + vec(\Gamma_{\boldsymbol{\Xi}})).$$

In particular, for $p = 2$ it follows that

$$\mathbb{V}(Y_t) = \frac{(1 - \phi_2)\left(\left(\sigma_{\phi_1}^2 + \sigma_{\phi_2}^2\right)(\mathbb{E}(Y_t))^2 + \sigma_\varepsilon^2\right)}{1 - \left(\phi_1^2 + \phi_1^2\phi_2 - \phi_2^3 + \phi_2^2 + \phi_2 + (1 - \phi_2)\left(\sigma_{\phi_1}^2 + \sigma_{\phi_2}^2\right)\right)}.$$

On a final note, the RMINAR model is able to generate both (unconditional) over-dispersion (e.g. with Poisson distributed inputs) and under dispersion (e.g. with binomial inputs).

3.2 A $\mathbb{Z}$-valued RMINAR extension

The RMINAR(p) model (8) can be extended to address $\mathbb{Z}$-valued time series simply by considering $\mathbb{Z}$-valued random inputs. Thus, let $\{\Phi_{it},\ t \in \mathbb{Z}\}$ $(i = 1, \ldots, p)$ and $\{\varepsilon_t,\ t \in \mathbb{Z}\}$ be mutually independent $\mathbb{Z}$-valued iid sequences with $\phi_i \in \mathbb{R}$ $(i = 1, \ldots, p)$ and $\mu_\varepsilon \in \mathbb{R}$ instead of being constrained to be positive values. A $\mathbb{Z}$-valued RMINAR(p) writes as in (8) as follows

$$Y_t = \sum_{i=1}^p \Phi_{it} Y_{t-i} + \varepsilon_t,\ t \in \mathbb{Z}, \tag{21}$$

where the random multiplication $\Phi_{it}Y_{t-i} \equiv \phi_i \odot_m Y_{t-i}$ now acts on $\mathbb{Z}$ with $\phi_i \in \mathbb{R}$ and $Y_{t-i} \in \mathbb{Z}$. Many distributions supported on $\mathbb{Z}$ can be chosen for the inputs $\{\Phi_{it},\ t \in \mathbb{Z}\}$ $(i = 1, \ldots, p)$ and $\{\varepsilon_t,\ t \in \mathbb{Z}\}$. For example, a useful candidate is the Skellam distribution (Irwin, 1937; Skellam, 1946) for the difference of two independent Poisson variables (or simply Poisson difference, PD). Another useful choice with heavier tails is the negative binomial difference (NBD) distribution (cf. Barndorff-Nielsen *et al.*, 2012; Barra and Koopman, 2018).

The conditional mean and variance of model (21) are given exactly as (9). Many efforts have been made to define $\mathbb{Z}$-valued (or signed integer-valued) models following different approaches, namely thinning-based models (e.g. Kim and Park, 2008; Zhang *et al.*, 2009; Alzaid and Omair, 2014), parameter driven models (Koopman *et al.*, 2017; Barra and Koopman, 2018), and observation driven models (Alomani *et al.*, 2018; Cui *et al.*, 2021). The resulting specifications, however, appear to be less straightforward than those of the $\mathbb{Z}$-valued

16

RMINAR(p) proposed in this work. In particular, the INGARCH-type models (Alomani *et al.*, 2018; Cui *et al.*, 2021) only concern the conditional variance and not the conditional mean. On the contrary, the RMINAR(p) approach (21) simultaneously addresses the conditional mean and variance and is analytically much simpler. Furthermore, given the connections between RMINAR(p) and RCAR(p), the existing real-valued RCAR(p) tools (Nicholls and Quinn, 1982) can be adapted e.g. for the estimation of RMINAR(p) models.

The probability structure of (21) is similar to that of (8) as (21) can be written in the vector form (12) with obvious notations. Furthermore, the universal ergodicity property for the $\mathbb{N}_0$-valued RMINAR model (8) still holds for the $\mathbb{Z}$-valued RMINAR model (21). Under the assumption $\mathbb{P}\left(\Phi_{it} = 0\right) > 0 \ (i = 1, \ldots, p)$, which is condition **A0** where Φ_{it} is $\mathbb{Z}$-valued instead of $\mathbb{N}_0$-valued (hereafter referred as condition **A0***), Theorem 1 still holds true for the $\mathbb{Z}$-valued RMINAR model (21).

Theorem 2 *Under **A0*** the conclusions of Theorem 1 remain true for the $\mathbb{Z}$-valued RMINAR model (21).*

Remark 2 *i) If **A0*** fails (e.g. $\mathbb{P}\left(\Phi_{1t} \in \mathbb{Z} \setminus \{0\}\right) = 1$), then assuming that $\mathbb{E}\left(\log^+\left(|\varepsilon_t|\right)\right) < \infty$ and $\mathbb{E}\left(\log^+\left(|\Phi_{it}|\right)\right) < \infty \ (i = 1, \ldots, p)$, the $\mathbb{Z}$-valued RMINAR model (21) still admits a unique strictly stationary and ergodic solution if $\gamma < 0$.*

The tail behavior is much more difficult to analyze for the $\mathbb{Z}$-valued RMINAR model (21) than for the $\mathbb{N}_0$-valued RMINAR model (8). However, it can be easily unmasked for $p = 1$ by using the following Kesten-Goldie theorem (Kesten, 1973; Goldie, 1991). This result is presented in proposition 3 which constitutes a corollary of Kesten's theorem 5.

Proposition 3 *Consider the $\mathbb{Z}$-valued RMINAR(1) model (21) subject to condition **A0*** and $\mathbb{E}\left(\log^+\left(|\Phi_{1t}|\right)\right) < \infty$. Suppose further that*

$$\mathbb{E}\left(|\Phi_{1t}|^{\tau_1}\right) = 1, \ \mathbb{E}\left(|\Phi_{1t}|^{\tau_1} \log^+ |\Phi_{1t}|\right) < \infty \quad and \quad \mathbb{E}\left(|\varepsilon_t|^{\tau_1}\right) < \infty \ , \ for \ some \ \tau_1 > 0.$$

Then

$$\mathbb{P}\left(Y_t \geq y\right) \to c^+ y^{-\tau_1} \quad and \quad \mathbb{P}(Y_t < -y) \to c^- y^{-\alpha}, \ as \ y \to \infty,$$

if and only if $\mathbb{P}(\varepsilon_1 + \mu_\varepsilon = (1 - \phi_1)c) < 1$ *for any* $c \in \mathbb{R}$ *such that* $c^+ + c^- > 0$.

Concerning the existence of moments for the $\mathbb{Z}$-valued RMINAR(p) model (21) assume that $\mathbb{E}\left(|\varepsilon_t|^{2m}\right) < \infty$, for $m \geq 1$. Using again the results of Feigin and Tweedie (1985, Theorems 4-5), a sufficient condition for the existence of a stationary and ergodic solution with $\mathbb{E}(|Y_t|^{2m}) < \infty$ is that

$$\rho\left(\mathbb{E}\left(A_t^{\otimes 2m}\right)\right) < 1.$$

As for the $\mathbb{N}_0$-valued RMINAR(p) model (8), the marginal mean and variance of $\mathbb{Z}$-valued RMINAR(p) model (21) are given, respectively, by (19) and (20) for $p = 1$.

3.3 RMINAR model with multiplicative errors

The RMINAR(p) models can be combined together with a scheme of multiplicative errors instead of additive ones. An example is the multiplicative thinning-based INGARCH model (MthINGARCH) of Aknouche and Scotto (2024), which has been shown to be flexible in representing many important features commonly observed integer-valued time series, such as over-dispersion and heavy-tailedness. The MthINGARCH(p,q) process is defined as

$$Y_t = \lambda_t\, \varepsilon_t \ \text{ and } \ \lambda_t = 1 + \omega \circ m + \sum_{i=1}^{p} \alpha_i \circ Y_{t-i} + \sum_{j=1}^{q} \beta_j \circ \lambda_{t-j}, \tag{22}$$

where "$\circ$" stands for the BTO. The multiplicative error (ε_t) is an iid integer-valued sequence with $\mathbb{E}\left(\varepsilon_t\right) = 1$ and variance $\sigma_\varepsilon^2 := \mathbb{V}\left(\varepsilon_t\right)$ while the random variable λ_t is such that $\lambda_t \in \mathbb{N}$, as a consequence of (22). Given the restrictions on the BTO, the coefficients of the MthINGARCH satisfy $0 \leq \alpha_i < 1$, $0 \leq \beta_j < 1$ $(i = 1, \ldots, p,\ j = 1, \ldots, q)$ and $0 \leq \omega \leq 1$, being ω the expected value of the Bernoulli random variables (ξ_i) $(i = 1, \ldots, m)$ underlying the BTO. Finally, being m defined as a fixed positive integer then $\omega \circ m \sim B(m, \omega)$.

The MthINGARCH model (22) can be modified by replacing the BTO $\circ$ by the RMO $\odot_m$ in (5). This case considers i) the MthINGARCH framework (22) with $\beta_j = 0$ $(j = 1, \ldots, q)$ for simplicity and ii) the replacement of the random variable $\omega \circ m$ by an integer-valued

random coefficient ω_t with $\mathbb{E}\left(\omega_t\right) = \omega$ and $\sigma_\omega^2 := \mathbb{V}\left(\omega_t\right) > 0$. These specifications result in the following RMINAR(p) model with multiplicative errors

$$Y_t = \lambda_t\,\varepsilon_t = \left(1 + \omega_t + \sum_{i=1}^{p} \phi_i \odot_m Y_{t-i}\right)\varepsilon_t = \left(1 + \omega_t + \sum_{i=1}^{p} \Phi_{it} Y_{t-i}\right)\varepsilon_t, \qquad (23)$$

where the integer-valued iid sequences $\{\omega_t,\ t \in \mathbb{Z}\}$, $\{\Phi_{it},\ t \in \mathbb{Z}\}$ $(i = 1, \ldots, p)$ and $\{\varepsilon_t,\ t \in \mathbb{Z}\}$ are mutually independent. As before, $\phi_i := \mathbb{E}\left(\Phi_{it}\right)$ $(i = 1, \ldots, p)$ and $\sigma_{\phi_i}^2 := \mathbb{V}\left(\Phi_{it}\right) > 0$. It follows from the RMINAR(p) formulation (23) that

$$\mathbb{E}\left(Y_t|\mathcal{F}_{t-1}^Y\right) = 1 + \omega + \sum_{i=1}^{p} \phi_i Y_{t-i}, \qquad (24)$$

$$\mathbb{V}\left(Y_t|\mathcal{F}_{t-1}^Y\right) = \left(\sigma_\varepsilon^2 + 1\right)\mathbb{V}\left(\lambda_t|\mathcal{F}_{t-1}^Y\right) + \sigma_\varepsilon^2\left(\mathbb{E}\left(Y_t|\mathcal{F}_{t-1}^Y\right)\right)^2, \qquad (25)$$

where

$$\mathbb{V}\left(\lambda_t|\mathcal{F}_{t-1}^Y\right) = \sigma_\omega^2 + \sum_{i=1}^{p} \sigma_{\phi_i}^2 Y_{t-i}^2,$$

which highlights that the RMINAR(p) with multiplicative errors (23) presents the same conditional-variance-to-mean relationship as that of the BTO-based MthINGARCH(p,0) model (Aknouche and Scotto, 2024).

Under a similar assumption as **A0**, the ergodic properties of the RMINAR(p) model with multiplicative errors (23) are obtained everywhere similarly to the above cases. The model (23) can be written in the following vector form

$$\mathbf{Y}_t = A_t \mathbf{Y}_{t-1} + \mathbf{\Psi}_t, \quad t \in \mathbb{Z},$$

with the definition of the p-dimensional column vectors $\mathbf{Y}_t = \left(Y_t, \ldots, Y_{t-p+1}\right)'$ and $\mathbf{\Psi}_t = \left(\varepsilon_t\left(1 + \omega_t\right), 0, \ldots, 0\right)'$, and the $p \times p$ companion matrix

$$A_t = \begin{pmatrix} \left(\varepsilon_t\phi_{1t}, \ldots, \varepsilon_t\phi_{(p-1)t}\right) & \varepsilon_t\phi_{pt} \\ \mathbf{I}_{p-1} & \mathbf{0}_{(p-1)\times 1} \end{pmatrix}.$$

Similarly to Theorem 1, the following result shows that under $\mathbb{P}\left(\varepsilon_t = 0\right) > 0$ and $\mathbb{P}\left(\Phi_{it} = 0\right) > 0$ $(i = 1, \ldots, p)$ (which will be referred as condition **A0****), any solution of the RMINAR(p) model with multiplicative errors (23) is everywhere stationary and ergodic.

Theorem 3 *Under **A0****, the series*

$$\mathbf{Y}_t := \sum_{j=0}^{\infty} \prod_{i=0}^{j-1} A_{t-i} \boldsymbol{\Psi}_{t-j},$$

converges absolutely a.s. for all $t \in \mathbb{Z}$. In addition, the process $\{Y_t, t \in \mathbb{Z}\}$ given by $Y_t = (1, 0, \ldots, 0)' \mathbf{Y}_t$ $(t \in \mathbb{Z})$ is the unique strictly stationary and ergodic solution to (23).

The tail behavior of the RMINAR model with multiplicative errors (23) can be obtained similarly to that of the $\mathbb{N}_0$-valued RMINAR model (8) using Kesten's theorem. Proposition 4 states such result and its proof its omitted, given that it is an obvious corollary of Kesten's theorems 3-5 (Kesten, 1973).

Proposition 4 *Under **A0****, all conclusions of Proposition 1 hold for the RMINAR model with multiplicative errors (23) while replacing Φ_{it} by $\varepsilon_t \Phi_{it}$.*

4 Four-stage WLS estimation

This section presents an estimation strategy for the parameters mean θ_0 and variance Λ_0 of the RMINAR models. For the $\mathbb{N}_0$-valued (8) and $\mathbb{Z}$-valued (21) models, the parameters are set as $\theta_0 = \left(\mu_{0\varepsilon}, \phi_{01}, \ldots, \phi_{0p}\right)'$ and $\Lambda_0 = \left(\sigma_{0\varepsilon}^2, \sigma_{0\phi_1}^2, \ldots, \sigma_{0\phi_p}^2\right)'$ while for the model with multiplicative errors (23), $\theta_0 = \left(\omega_0, \phi_{01}, \ldots, \phi_{0p}\right)'$ and $\Lambda_0 = \left(\sigma_{0\varepsilon}^2, \sigma_{0\omega}^2, \sigma_{0\phi_1}^2, \ldots, \sigma_{0\phi_p}^2\right)'$. The subscript 0 in the notation of θ_0 and Λ_0 aims at distinguishing the true parameters from generic ones $\theta \in \Theta$ and $\Lambda \in \Pi$, where Θ and Π represent the parameter spaces.

As the RMINAR models (8), (21) and (23) have similar conditional means, being linear in the models' parameters, this work provides a unified estimation procedure based on a four-stage weighted least squares estimation (4SWLSE) approach (see e.g. Aknouche, 2015). Let $\mu_t \equiv \mu_t (\theta_0) := \mathbb{E}\left(Y_t | \mathcal{F}_{t-1}^Y\right)$ and $V_t \equiv V_t (\Lambda_0) := \mathbb{V}\left(Y_t | \mathcal{F}_{t-1}^Y\right)$ be, respectively, the conditional mean and variance of the RMINAR process Y_t in one of the three models (8), (21) or (23). In some cases, $V_t := V_t (\theta_0, \Lambda_0)$ also depends on θ_0 (as in the multiplicative model). The principle and the rationale of the method are as follows. The 4SWLSE develops in 4 stages;

20

firstly, θ_0 is first estimated by $\widehat{\theta}_{1n}$ from the regression $Y_t = \mu_t + e_t$, using a WLS estimator weighted by the conditional variance $V_t(\Lambda_*)$, evaluated at some arbitrarily and fixed known vector $\Lambda_* \in \Pi$. The sequence (e_t) is a martingale difference with respect to $(\mathcal{F}_t^Y)$. Secondly, Λ_0 is estimated by $\widehat{\Lambda}_{1n}$ from the regression $(Y_t - \mu_t)^2 = V_t + u_t$, where the conditional variance residual (u_t) is also a martingale difference with respect to $(\mathcal{F}_t^Y)$. Thus, this step makes use of a WLS estimator weighted by $V_t^2(\Lambda_*)$ arbitrarily evaluated. The third stage reestimates θ_0 by $\widehat{\theta}_{2n}$ from the WLS approach used in the first stage but considering the estimated conditional variance as weights. The same reasoning is used in the fourth stage to reestimate Λ_0, where the weight is $V_t^2(\widehat{\Lambda}_{1n})$. For any generic parameters $\theta \in \Theta$ and $\Lambda \in \Pi$, let $e_t(\theta) = Y_t - \mu_t(\theta)$ and $u_t(\theta, \Lambda) = e_t^2(\theta) - V_t(\theta, \Lambda)$, where $V_t(\theta, \Lambda)$ is the conditional variance function in which the true Λ_0 is replaced by a generic Λ. Naturally, $e_t(\theta_0) = e_t$ and $u_t(\theta_0, \Lambda_0) = u_t$. Formally, the proposed 4SWLSE method is given by the following cascade of four optimization problems

$$\text{i) } \widehat{\theta}_{1n} = \arg\min_{\theta \in \Theta} \sum_{t=1}^n \frac{e_t^2(\theta)}{V_t(\Lambda_*)}, \text{ ii) } \widehat{\Lambda}_{1n} = \arg\min_{\Lambda \in \Pi} \sum_{t=1}^n \frac{u_t^2(\widehat{\theta}_{1n}, \Lambda)}{V_t^2(\Lambda_*)}, \tag{26}$$

$$\text{iii) } \widehat{\theta}_{2n} = \arg\min_{\theta \in \Theta} \sum_{t=1}^n \frac{e_t^2(\theta)}{V_t(\widehat{\Lambda}_{1n})}, \text{ iv) } \widehat{\Lambda}_{2n} = \arg\min_{\Lambda \in \Pi} \sum_{t=1}^n \frac{u_t^2(\widehat{\theta}_{2n}, \Lambda)}{V_t^2(\Lambda_*)}, \tag{27}$$

where the estimators $\widehat{\theta}_{\cdot n}$ and $\widehat{\Lambda}_{\cdot n}$ have the same notation as the corresponding target parameters, with an extra *hat* symbol and the n to refer to the sample size. Furthermore, the subscript 1 or 2 is used to highlight if the step performs estimation or re-estimation, respectively, of the corresponding parameter.

There is relevant literature connected with the 4SWLSE approach and the derivation of the properties of the estimators in (26) and (27) with respect to consistency, efficiency and asymptotic distribution. As an instance of that, Nicholls and Quinn (1982) proposed the two-stage least squares estimation (2SLSE) approach based on

$$\text{i) } \widehat{\theta}_n = \arg\min_{\theta \in \Theta} \sum_{t=1}^n e_t^2(\theta) \quad \text{and} \quad \text{ii) } \widehat{\Lambda}_n = \arg\min_{\Lambda \in \Pi} \sum_{t=1}^n u_t^2\left(\widehat{\theta}_{1n}, \Lambda\right), \tag{28}$$

to estimate θ_0 and Λ_0 in a consecutive manner. Schick (1996) studied the properties of $\widehat{\theta}_{1n}$ in (26) when Λ_* is replaced by any consistent estimator $\widehat{\Lambda}$ of Λ_0. For a RCAR(1) model,

21

Aknouche (2015) proposed a four-stage WLSE similar to (26) considering the weights $V_t^2(\Lambda_*)$ and $V_t(\widehat{\Lambda}_{1n})$, respectively, for the second and fourth stages. The choice for the latter weights results in an estimator for Λ_0 with the same asymptotic variance as that of the Gaussian QMLE. In our case, the 4SWLSE approach makes use of the latter weights in the second and third stages leading to estimators that are everywhere consistent and asymptotically Gaussian with a certain optimality. Moreover, it will be proved that $\widehat{\theta}_{2n}$ and $\widehat{\Lambda}_{2n}$ (27) are asymptotically more efficient than $\widehat{\theta}_{1n}$ and $\widehat{\Lambda}_{1n}$ (26), respectively, and more efficient than the 2SLSE estimators $\widehat{\theta}_n$ and $\widehat{\Lambda}_n$ (28) and any of the exponential family QMLE.

Note that for most usual discrete distributions, Λ_0 is a function of θ_0 and (possibly) some additional parameters. E.g. $\Lambda_0 = \theta_0$ when all coefficients are Poisson-distributed, and $\Lambda_0 = \theta_0 \oplus (1_{1+p} + \theta_0)$ when are geometric distributed, where 1_p is a p-length vector of ones and $\oplus$ stands for the Hadamard matrix product. These cases only require a two-stage estimation procedure corresponding to the first and third stages of the 4SWLSE. If the random coefficients $\Phi_{it} \sim \mathcal{NB}\left(\nu_{0i}, \frac{\nu_{0i}}{\nu_i + \phi_{0i}}\right)$ $(i = 1, \ldots, p)$ and $\epsilon_t \sim \mathcal{NB}\left(\nu_0, \frac{\nu_0}{\nu_0 + \omega_0}\right)$ are negative binomial II distributed, then $\sigma_{\Phi_i}^2 = \phi_{0i}\left(1 + \frac{1}{\nu_i}\phi_{0i}\right)$ and $\sigma_\varepsilon^2 = \mu_\epsilon\left(1 + \frac{1}{\nu_0}\mu_\epsilon\right)$. Hence, ν_0 and $\nu_{0i} > 0$ are estimated by $\widehat{\nu}_n^{-1}$ and $\widehat{\nu}_{in}^{-1}$, respectively, from the stage (iv) in (27) using

$$\widehat{\nu}_n^{-1} = \frac{\widehat{\sigma}_{2n}^2 - \widehat{\mu}_{\epsilon n}}{\widehat{\mu}_{\epsilon n}^2} \quad \text{and} \quad \widehat{\nu}_{in}^{-1} = \frac{\widehat{\sigma}_{\Phi_i n}^2 - \widehat{\phi}_{in}}{\widehat{\phi}_{in}^2}, \, i = 1, \ldots, p.$$

4.1 The $\mathbb{N}_0$-valued RMINAR model

Due to the mutual independence of the random sequences $\{\Phi_{it}, \, t \in \mathbb{Z}\}$ $(i = 1, \ldots, p)$ and $\{\varepsilon_t, \, t \in \mathbb{Z}\}$, the conditional mean and variance of the RMINAR model (8) have the following linear-in-parameters expressions

$$\mu_t(\theta_0) = \theta_0' \mathcal{Y}_{t-1} \quad \text{and} \quad V_t(\theta_0, \Lambda_0) = \mathcal{Z}_{t-1}' \Lambda_0, \tag{29}$$

where $\mathcal{Y}_{t-1} = (1, Y_{t-1}, \ldots, Y_{t-p})'$ and $\mathcal{Z}_{t-1} = \left(1, Y_{t-1}^2, \ldots, Y_{t-p}^2\right)' = \mathcal{Y}_{t-1} \oplus \mathcal{Y}_{t-1}$. Then, the estimators $[\widehat{\theta}_{1n}, \widehat{\Lambda}_{1n}, \widehat{\theta}_{2n}, \widehat{\Lambda}_{2n}]$ in (26)-(27) have the more specific closed-form expressions

$$
\text{i) } \widehat{\theta}_{1n} = \left(\sum_{t=1}^{n} \frac{\mathcal{Y}_{t-1}\mathcal{Y}_{t-1}'}{\mathcal{Z}_{t-1}'\Lambda_*}\right)^{-1} \sum_{t=1}^{n} \frac{\mathcal{Y}_{t-1}Y_t}{\mathcal{Z}_{t-1}'\Lambda_*}, \tag{30}
$$

$$
\text{ii) } \widehat{\Lambda}_{1n} = \left(\sum_{t=1}^{n} \frac{\mathcal{Z}_{t-1}\mathcal{Z}_{t-1}'}{\left(\mathcal{Z}_{t-1}'\Lambda_*\right)^2}\right)^{-1} \sum_{t=1}^{n} \mathcal{Z}_{t-1}\frac{\left(Y_t - \mathcal{Y}_{t-1}'\widehat{\theta}_{1n}\right)^2}{\left(\mathcal{Z}_{t-1}'\Lambda_*\right)^2}, \tag{31}
$$

$$
\text{iii) } \widehat{\theta}_{2n} = \left(\sum_{t=1}^{n} \frac{\mathcal{Y}_{t-1}\mathcal{Y}_{t-1}'}{\mathcal{Z}_{t-1}'\widehat{\Lambda}_{1n}}\right)^{-1} \sum_{t=1}^{n} \frac{\mathcal{Y}_{t-1}Y_t}{\mathcal{Z}_{t-1}'\widehat{\Lambda}_{1n}}, \tag{32}
$$

$$
\text{iv) } \widehat{\Lambda}_{2n} = \left(\sum_{t=1}^{n} \frac{\mathcal{Z}_{t-1}\mathcal{Z}_{t-1}'}{\left(\mathcal{Z}_{t-1}'\widehat{\Lambda}_{1n}\right)^2}\right)^{-1} \sum_{t=1}^{n} \mathcal{Z}_{t-1}\frac{\left(Y_t - \mathcal{Y}_{t-1}'\widehat{\theta}_{2n}\right)^2}{\left(\mathcal{Z}_{t-1}'\widehat{\Lambda}_{1n}\right)^2}, \tag{33}
$$

where $\Lambda_* = \left(\sigma_{*\varepsilon}^2, \sigma_{*\phi_1}^2, \ldots, \sigma_{*\phi_p}^2\right)' > 0$ is arbitrarily fixed. When all random coefficients are Poisson or geometric distributed, then the estimation steps (31) and (33) are useless but can be used to test the Poisson hypothesis $H_0 : \theta_0 = \Lambda_0$ or the geometric one $H_0 : \Lambda_0 = \theta_0 \oplus (1_{1+p} + \theta_0)$.

The everywhere consistency and asymptotic normality of the estimators in (30)-(33) are established in Theorem (4), with complete proof in Appendix B. Consider

$$
\Sigma\left(\Lambda_0, \Lambda_*\right) := A\left(\Lambda_*\right)^{-1} B\left(\Lambda_0, \Lambda_*\right) A\left(\Lambda_*\right)^{-1} \text{ and } \Omega\left(\Lambda_*\right) := C\left(\Lambda_*\right)^{-1} D\left(\Lambda_*\right) C\left(\Lambda_*\right)^{-1},
$$

with auxiliary matrices A, B, C and D defined as

$$
A\left(\Lambda_*\right) := \mathbb{E}\left(\frac{\mathcal{Y}_{t-1}\mathcal{Y}_{t-1}'}{\mathcal{Z}_{t-1}'\Lambda_*}\right) \quad, \quad B\left(\Lambda_0, \Lambda_*\right) := \mathbb{E}\left(\frac{\left(\mathcal{Z}_{t-1}'\Lambda_0\right)}{\left(\mathcal{Z}_{t-1}'\Lambda_*\right)^2}\mathcal{Y}_{t-1}\mathcal{Y}_{t-1}'\right),
$$

$$
C\left(\Lambda_*\right) := \mathbb{E}\left(\frac{\mathcal{Z}_{t-1}\mathcal{Z}_{t-1}'}{\left(\mathcal{Z}_{t-1}'\Lambda_*\right)^2}\right) \quad, \quad D\left(\Lambda_*\right) := \mathbb{E}\left(\frac{u_t^2}{\left(\mathcal{Z}_{t-1}'\Lambda_*\right)^4}\mathcal{Z}_{t-1}\mathcal{Z}_{t-1}'\right).
$$

Note that $B\left(\Lambda_0, \Lambda_0\right) = A\left(\Lambda_0\right)$ implies $\Sigma\left(\Lambda_0, \Lambda_0\right) = A\left(\Lambda_0\right)^{-1}$ and that $D\left(\Lambda_0\right) = C\left(\Lambda_0\right)$ leads to $\Omega\left(\Lambda_0\right) = C\left(\Lambda_0\right)^{-1}$. Also, recall that the finiteness of $\mathbb{E}\left(\varepsilon_t^2\right)$ and $\mathbb{E}\left(\Phi_{it}^2\right)$ holds by the model's definition.

Theorem 4 *For the RMINAR model (8) under **A0**, for all $\Lambda_* > 0$ it holds that*

$$
i) \quad \widehat{\theta}_{1n} \underset{n\to\infty}{\overset{a.s.}{\to}} \theta_0, \quad ii) \ \widehat{\theta}_{2n} \underset{n\to\infty}{\overset{a.s.}{\to}} \theta_0, \quad iii) \ \widehat{\Lambda}_{1n} \underset{n\to\infty}{\overset{a.s.}{\to}} \Lambda_0, \quad iv) \ \widehat{\Lambda}_{2n} \underset{n\to\infty}{\overset{a.s.}{\to}} \Lambda_0. \tag{34}
$$

23

If, in addition, $\mathbb{E}\left(\varepsilon_t^4\right) < \infty$ and $\mathbb{E}\left(\Phi_{it}^4\right) < 1$ $(i = 1, \ldots, p)$ then

$$i) \quad \sqrt{n}\left(\widehat{\theta}_{1n} - \theta_0\right) \xrightarrow[n\to\infty]{D} \mathcal{N}\left(0, \Sigma\left(\Lambda_0, \Lambda_*\right)\right), ii)\sqrt{n}\left(\widehat{\theta}_{2n} - \theta_0\right) \xrightarrow[n\to\infty]{D} \mathcal{N}\left(0, \Sigma\left(\Lambda_0, \Lambda_0\right)\right)$$

$$iii) \quad \sqrt{n}\left(\widehat{\Lambda}_{1n} - \Lambda_0\right) \xrightarrow[n\to\infty]{D} \mathcal{N}\left(0, \Omega\left(\Lambda_*\right)\right), \ iv) \ \sqrt{n}\left(\widehat{\Lambda}_{2n} - \Lambda_0\right) \xrightarrow[n\to\infty]{D} \mathcal{N}\left(0, \Omega\left(\Lambda_0\right)\right). \quad (35)$$

Proof. See Appendix B. ∎

The most surprising fact given by the above result is that the proposed 4SWLSE estimators are consistent and asymptotically normal everywhere, even when the RMINAR model has an infinite mean. In addition, $\widehat{\theta}_{2n}$ and $\widehat{\Lambda}_{2n}$ are more efficient than $\widehat{\theta}_{1n}$ and $\widehat{\Lambda}_{1n}$, respectively, as $\Sigma\left(\Lambda_0, \Lambda_0\right) \leq \Sigma\left(\Lambda_0, \Lambda_*\right)$ and $\Omega\left(\Lambda_0\right) \leq \Omega\left(\Lambda_*\right)$ for all Λ_*. The finiteness of the moments $\mathbb{E}\left(\varepsilon_t^4\right) < \infty$ and $\mathbb{E}\left(\Phi_{it}^4\right) < 1$ $(i = 1, \ldots, p)$, required for the asymptotic normality to hold, does not constitute a restrictive condition as most usual discrete distributions (Poisson, binomial, negative binomial, etc.) have finite higher-order moments.

Finally, note that the main advantage of precisely including the weights $\left(\mathcal{Z}_{t-1}'\Lambda_*\right)^2$ and $\left(\mathcal{Z}_{t-1}'\widehat{\Lambda}_{1n}\right)^2$ in the second and fourth stages is to have the everywhere-consistency property of the corresponding estimates $\widehat{\Lambda}_{\cdot n}$. However, while $\widehat{\theta}_{2n}$ is optimal in the sense of Schick (1996), the resulting estimate $\widehat{\Lambda}_{2n}$ is not asymptotically optimal in the class of all WLSEs. In fact, the optimal weight in the sense of obtaining an estimate such that $C\left(\Lambda_0\right)$ and $D\left(\Lambda_0\right)$ are proportional is $\mathbb{E}\left(u_t^2|\mathcal{F}_{t-1}^Y\right) = \mathbb{V}\left(u_t|\mathcal{F}_{t-1}^Y\right)$ and the resulting estimate would be more efficient than $\widehat{\Lambda}_{2n}$. Nonetheless, the closed form expression of $\mathbb{V}\left(u_t|\mathcal{F}_{t-1}^Y\right)$ in terms of model parameters is not always available and depends on the model distribution which is not necessarily known in our framework. Moreover, $\mathbb{V}\left(u_t|\mathcal{F}_{t-1}^Y\right)$ may depend on some third- and fourth-moment parameters that need to be specified. E.g. the RMINAR(1) model has

$$\mathbb{V}\left(u_t|\mathcal{F}_{t-1}^Y\right) = \mathbb{E}\left(\Phi_{1t} - \phi_{01}\right)^4 Y_{t-1}^4 + 6\sigma_\phi^2\sigma_\varepsilon^2 Y_{t-1}^2 + \mathbb{E}\left(\left(\varepsilon_t - \mu_\varepsilon\right)^4\right) - \left(\mathcal{Z}_{t-1}'\Lambda_0\right)^2,$$

which requires specifying $\mathbb{E}\left(\Phi_{1t} - \phi_{01}\right)^4$ and $\mathbb{E}\left(\left(\varepsilon_t - \mu_\varepsilon\right)^4\right)$ in the model's definition. Therefore, the above weight consisting of the squared conditional variance seems a reasonably good choice, especially since for the ARCH model and the double auto-regression (DAR, Ling, 2007) model, $\left(\mathcal{Z}_{t-1}'\Lambda_0\right)^2$ and $\mathbb{V}\left(u_t|\mathcal{F}_{t-1}^Y\right)$ are proportional (see also Aknouche, 2012).

24

4.2 The $\mathbb{Z}$-valued RMINAR model

It is clear that the $\mathbb{Z}$-valued RMINAR model (21) has a similar structure as its $\mathbb{N}_0$-valued counterpart (8). In particular, the conditional mean and variance of (21) are the same as those for (8) and are given by (29). Thus the general 4SWLSE procedure (26)-(27) reduces for model (21) to (30)-(33). Keeping the same notations for the parameters, namely $\theta_0 = \left(\mu_{0\varepsilon}, \phi_{01}, \ldots, \phi_{0p} \right)'$ and $\Lambda_0 = \left(\sigma_{0\varepsilon}^2, \sigma_{0\phi_1}^2, \ldots, \sigma_{0\phi_p}^2 \right)'$, the estimators $[\widehat{\theta}_{1n}, \widehat{\Lambda}_{1n}, \widehat{\theta}_{2n}, \widehat{\Lambda}_{2n}]$, and the asymptotic matrices Σ and Ω in (35), the following result states that the same conclusions of Theorem 4 hold for the $\mathbb{Z}$-valued RMINAR model (21).

Theorem 5 *For the RMINAR model* (21) *under **A0***, *the same conclusions of Theorem 4 hold for all arbitrarily and fixed known* $\Lambda_* > 0$.

Proof. See Appendix B. ∎

4.3 The RMINAR model with multiplicative errors

For the RMINAR model with multiplicative errors (23), the conditional mean (24) is linear in the parameter $\theta_0 = \left(\omega_0, \phi_{01}, \ldots, \phi_{0p} \right)'$ as in models (8) and (21) and rewrites as

$$\mu_t \left(\theta_0 \right) := \mathbb{E} \left(Y_t | \mathcal{F}_{t-1} \right) = 1 + \mathcal{Y}_{t-1}' \theta_0.$$

In view of (25), however, it turns out that the conditional variance

$$V_t \left(\Lambda_0, \theta_0 \right) := \mathbb{V} \left(Y_t | \mathcal{F}_{t-1} \right) = \left(\sigma_\varepsilon^2 + 1 \right) \left(\mathcal{Z}_{t-1}' \Lambda_0 \right) + \sigma_\varepsilon^2 \left(1 + \mathcal{Y}_{t-1}' \theta_0 \right)^2,$$

is not linear in the variance parameter $\Lambda_0 = \left(\sigma_{0\varepsilon}^2, \sigma_{0\omega}^2, \sigma_{0\phi_1}^2, \ldots, \sigma_{0\phi_p}^2 \right)'$, which contains an extra parameter $\sigma_{0\varepsilon}^2$. Hence, the first and third stages to estimate θ_0 are given in closed-form as for model (8) and (21), but the second and fourth stages are now iterative. Thus, the general 4SWLSE procedure (26)-(27) reduces for model (23) to the following problems

$$\text{i)} \ \widehat{\theta}_{1n} = \left(\sum_{t=1}^n \frac{\mathcal{Y}_{t-1} \mathcal{Y}_{t-1}'}{V_t(\theta_*, \Lambda_*)} \right)^{-1} \sum_{t=1}^n \frac{\mathcal{Y}_{t-1} Y_t}{V_t(\theta_*, \Lambda_*)} \ \text{ and ii)} \ \widehat{\Lambda}_{1n} = \arg\min_{\Lambda \in \Pi} \frac{1}{n} \sum_{t=1}^n \frac{u_t^2\left(\widehat{\theta}_{1n}, \Lambda\right)}{V_t^2(\theta_*, \Lambda_*)}, \tag{36}$$

$$\text{iii)} \ \widehat{\theta}_{2n} = \left(\sum_{t=1}^n \frac{\mathcal{Y}_{t-1} \mathcal{Y}_{t-1}'}{V_t\left(\widehat{\theta}_{1n}, \widehat{\Lambda}_{1n}\right)} \right)^{-1} \sum_{t=1}^n \frac{\mathcal{Y}_{t-1} Y_t}{V_t\left(\widehat{\theta}_{1n}, \widehat{\Lambda}_{1n}\right)} \ \text{ and iv)} \ \widehat{\Lambda}_{2n} = \arg\min_{\Lambda \in \Pi} \frac{1}{n} \sum_{t=1}^n \frac{u_t^2\left(\widehat{\theta}_{2n}, \Lambda\right)}{V_t^2\left(\widehat{\theta}_{2n}, \widehat{\Lambda}_{1n}\right)}, \tag{37}$$

25

where $\Lambda = \left(\sigma_\varepsilon^2, \sigma_\omega^2, \sigma_{\phi_1}^2, \dots, \sigma_{\phi_p}^2 \right)' \in \Pi$ and $\Lambda_* = \left(\sigma_{*\varepsilon}^2, \dots, \sigma_{*\phi_p}^2 \right)' > 0$ are the generic and the fixed known parameter vectors, respectively, and θ_* is fixed and known.

Next, Theorem 6 establishes the consistency and asymptotic normality of the 4SWLSE given by (36)-(37), relying on the assumptions that $\Lambda \in \Pi$ is compact and Λ_0 belongs to the interior of Π. To state the result, further consider the matrices

$$\Sigma\left(\theta_*, \Lambda_* \right) := A\left(\theta_*, \Lambda_* \right)^{-1} B\left(\theta_*, \Lambda_* \right) A\left(\theta_*, \Lambda_* \right)^{-1} \text{ and } \Omega\left(\Lambda_* \right) := C\left(\Lambda_* \right)^{-1} D\left(\Lambda_* \right) C\left(\Lambda_* \right)^{-1},$$

where the auxiliary matrices A, B, C and D are defined as

$$A\left(\theta_*, \Lambda_* \right) := \mathbb{E}\left(\tfrac{\mathcal{Y}_{t-1}\mathcal{Y}_{t-1}'}{V_t(\theta_*, \Lambda_*)} \right) \quad , \quad B\left(\theta_*, \Lambda_* \right) := \mathbb{E}\left(\tfrac{V_t(\theta_*, \Lambda_0)}{V_t(\theta_*, \Lambda_*)^2} \mathcal{Y}_{t-1}\mathcal{Y}_{t-1}' \right),$$

$$C\left(\Lambda_* \right) := \mathbb{E}\left(\tfrac{1}{V_t^2(\theta_0, \Lambda_*)} \tfrac{\partial V_t(\theta_0, \Lambda_0)}{\partial \Lambda} \tfrac{\partial V_t(\theta_0, \Lambda_0)}{\partial \Lambda'} \right) \quad , \quad D\left(\Lambda_* \right) := \mathbb{E}\left(\tfrac{u_t^2(\theta_0, \Lambda_0)}{V_t^4(\theta_0, \Lambda_*)} \tfrac{\partial V_t(\theta_0, \Lambda_0)}{\partial \Lambda} \tfrac{\partial V_t(\theta_0, \Lambda_0)}{\partial \Lambda'} \right).$$

Theorem 6 *Under model* (23) *subject to* **$A0^{**}$** *and* **$A1$:** $\Lambda \in \Pi$ *is compact,*

$$i) \ \widehat{\theta}_{1n} \overset{a.s.}{\underset{n \to \infty}{\to}} \theta_0, \quad ii) \ \widehat{\theta}_{2n} \overset{a.s.}{\underset{n \to \infty}{\to}} \theta_0, \quad iii) \ \widehat{\Lambda}_{1n} \overset{a.s.}{\underset{n \to \infty}{\to}} \Lambda_0, \quad iv) \ \widehat{\Lambda}_{2n} \overset{a.s.}{\underset{n \to \infty}{\to}} \Lambda_0. \tag{38}$$

If, in addition, $\mathbb{E}\left(\varepsilon_t^4 \right) < \infty$, $\mathbb{E}\left(\Phi_{it}^4 \right) < 1$ $(i = 1, \dots, p)$ *and* **$A2$:** Λ_0 *belongs to the interior of* Π *holds, then*

$$i) \quad \sqrt{n}\left(\widehat{\theta}_{1n} - \theta_0 \right) \overset{D}{\underset{n \to \infty}{\to}} \mathcal{N}\left(0, \Sigma\left(\Lambda_0, \Lambda_* \right) \right), ii) \sqrt{n}\left(\widehat{\theta}_{2n} - \theta_0 \right) \overset{D}{\underset{n \to \infty}{\to}} \mathcal{N}\left(0, \Sigma\left(\Lambda_0, \Lambda_0 \right) \right)$$

$$iii) \quad \sqrt{n}\left(\widehat{\Lambda}_{1n} - \Lambda_0 \right) \overset{D}{\underset{n \to \infty}{\to}} \mathcal{N}\left(0, \Omega\left(\Lambda_* \right) \right), \ iv) \ \sqrt{n}\left(\widehat{\Lambda}_{2n} - \Lambda_0 \right) \overset{D}{\underset{n \to \infty}{\to}} \mathcal{N}\left(0, \Omega\left(\Lambda_0 \right) \right). \tag{39}$$

Proof. See Appendix B. ∎

For some particular cases, the estimators in (36)-(37) can be considerably simplified. This is the case of the Poisson and the geometric distributions, for which the random coefficient variances $\Delta_0 := \left(\sigma_{0\omega}^2, \sigma_{0\phi_1}^2, \dots, \sigma_{0\phi_p}^2 \right)'$ are functions of the random coefficient means θ_0 with no extra parameters. Indeed, only the variance $\sigma_{0\varepsilon}^2$ of the innovation sequence has to be estimated in $\Lambda_0 = (\sigma_{0\varepsilon}^2, \Delta_0')$, as the components in Δ_0' are functions of the mean parameters in θ_0. Thus, the 4SWLSE procedure (36)-(37) simplifies to the triplet $[\widehat{\theta}_{1n}, \widehat{\sigma}_{\varepsilon n}^2, \widehat{\theta}_{2n}]$ where $\widehat{\sigma}_{\varepsilon n}^2$ is an estimator for $\sigma_{0\varepsilon}^2$. Since the RMINAR model with multiplicative errors (23) shares

the same variance-to-mean structure as the MthINGARCH model of Aknouche and Scotto (2023), the same estimator is used for $\sigma_{0\varepsilon}^2$. Rewrite (25) as follows

$$V_t := \mathbb{V}\left(Y_t|\mathcal{F}_{t-1}^Y\right) = V_t\left(\theta_0, \Lambda_0\right) = \left(\sigma_\varepsilon^2 + 1\right)\delta_t^2 + \sigma_\varepsilon^2\mu_t^2,$$

where $\delta_t^2 \equiv \delta_t^2\left(\Delta_0\right) = \mathbb{V}\left(\lambda_t|\mathcal{F}_{t-1}^Y\right) = \Delta_0\mathcal{Z}_{t-1}'$. Let $\widehat{\mu}_t := \mu_t\left(\widehat{\theta}_{1n}\right) = 1 + \mathcal{Y}_{t-1}'\widehat{\theta}_{1n}$, and $\widehat{\delta}_t^2 := \delta_t^2\left(\widehat{\Delta}_n\right)$ where $\widehat{\Delta}_n$ is a consistent estimator for Δ_0 obtained as the corresponding function of $\widehat{\theta}_{1n}$. Hence, (36)-(37) becomes

$$\text{i) } \widehat{\theta}_{1n} = \left(\sum_{t=1}^n \frac{\mathcal{Y}_{t-1}\mathcal{Y}_{t-1}'}{V_t(\theta_*, \Lambda_*)}\right)^{-1}\sum_{t=1}^n \frac{\mathcal{Y}_{t-1}Y_t}{V_t(\theta_*, \Lambda_*)}, \tag{40}$$

$$\text{ii) } \widehat{\sigma}_{\varepsilon n}^2 = \frac{1}{n}\sum_{t=1}^n \frac{(Y_t - \widehat{\mu}_t)^2 - \widehat{\delta}_t^2}{\widehat{\delta}_t^2 + \widehat{\mu}_t^2}, \tag{41}$$

$$\text{iii) } \widehat{\theta}_{2n} = \left(\sum_{t=1}^n \frac{\mathcal{Y}_{t-1}\mathcal{Y}_{t-1}'}{V_t(\widehat{\theta}_{1n}, \widehat{\Lambda}_{1n})}\right)^{-1}\sum_{t=1}^n \frac{\mathcal{Y}_{t-1}Y_t}{V_t(\widehat{\theta}_{1n}, \Lambda_0)}, \tag{42}$$

where $\widehat{\Lambda}_{1n}$ is the corresponding function of $\widehat{\theta}_{1n}$ and (40) is directly obtained from Aknouche and Scotto (2024, eq. 3.13). For this particular case, the consistency and distribution of the resulting estimators are organized in the following corollary to Theorem 6.

Proposition 5 *Under model* (23) *subject to* **A0****

$$i) \ \widehat{\theta}_{1n} \overset{a.s.}{\underset{n\to\infty}{\to}} \theta_0, \ ii) \ \widehat{\sigma}_{\varepsilon n}^2 \overset{a.s.}{\underset{n\to\infty}{\to}} \sigma_{0\varepsilon}^2, \ iii) \ \widehat{\theta}_{2n} \overset{a.s.}{\underset{n\to\infty}{\to}} \theta_0.$$

In addition,

$$i) \quad \sqrt{n}\left(\widehat{\theta}_{1n} - \theta_0\right) \overset{D}{\underset{n\to\infty}{\to}} \mathcal{N}\left(0, \Sigma\left(\theta_*, \Lambda_*\right)\right),$$

$$ii) \quad \sqrt{n}\left(\widehat{\sigma}_{\varepsilon n}^2 - \sigma_{0\varepsilon}^2\right) \overset{\mathcal{D}}{\underset{n\to\infty}{\to}} \mathcal{N}\left(0, \Gamma\right),$$

$$iii) \quad \sqrt{n}\left(\widehat{\theta}_{2n} - \theta_0\right) \overset{D}{\underset{n\to\infty}{\to}} \mathcal{N}\left(0, \Sigma\left(\theta_0, \Lambda_0\right)\right),$$

where

$$\Gamma := \mathbb{E}\left(\frac{(Y_t - \mu_t(\theta_0))^2 - \left(\delta_t^2(\Delta_0) + \left(\delta_t^2(\Delta_0) + \mu_t^2(\theta_0)\right)\sigma_{0\varepsilon}^2\right)}{\delta_t^2(\Delta_0) + \mu_t^2(\theta_0)}\right)^2.$$

5 Numerical illustrations

The behavior of the 4SWLSE in finite samples is assessed for the three above models via simulation experiments. Furthermore, two real data applications are analyzed.

5.1 Simulation study

For the $\mathbb{N}_0$-valued RMINAR model (8) two instances are considered. Firstly, a RMINAR(4) driven by Poisson inputs with three sets of parameters, namely: i) stationary RMINAR(4) model with finite mean, $\sum_{i=1}^{4} \phi_{0i} = 0.7 < 1$, ii) stationary RMINAR(4) model with infinite mean and $\sum_{i=1}^{4} \phi_{0i} = 1$, and iii) stationary RMINAR(4) model with infinite mean and $\sum_{i=1}^{4} \phi_{0i} = 1.2 > 1$ (cf. Table 1). The second instance (cf. Table 2), still concerns a $\mathbb{N}_0$-valued RMINAR(3) model, but considering several distributions for the inputs, namely $\varepsilon_t \sim \mathcal{B}in\left(5, \frac{\mu_\varepsilon}{5}\right)$ (binomial) $\Phi_{1t} \sim \mathcal{P}\left(\phi_{01}\right)$ (Poisson), $\Phi_{2t} \sim \mathcal{NB}\left(3, \frac{3}{3+\phi_{02}}\right)$ (NB2), and $\Phi_{3t} \sim \mathcal{NB}\left(2\phi_{03}, \frac{2}{3}\right)$ (NB1). As previously, two instances of parameters are considered: the stationary model with finite mean $\sum_{i=1}^{3} \phi_{0i} < 1$ and the stationary model with infinite mean, $\sum_{i=1}^{3} \phi_{0i} > 1$. Regarding the $\mathbb{Z}$-valued RMINAR model, $p = 3$ is considered where all inputs are Skellam distributed with two cases of stationary, namely stationary with finite mean, corresponding to $\rho\left(\Phi\right) = 0.5761 < 1$, and stationary with infinite mean and $\rho\left(\Phi\right) = 1$, where

$$\Phi := E\left(\Phi_t\right) = \begin{pmatrix} \phi_{01} & \phi_{02} & \phi_{03} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

and $\rho\left(\Phi\right)$ denoting the spectral radius of Φ (cf. Table 3). Finally, for the multiplicative RMINAR(3) model, all inputs are Poisson distributed with three cases for the parameters just like for the $\mathbb{N}_0$-valued RMINAR.

The 4SWLSE in (30)-(33) and (36)-(37) are run on 1000 simulated series generated from the mentioned RMINAR models with sample-sizes 1000. For all instances, the means, the standard deviations (StD), and asymptotic standard errors (ASE) of $\widehat{\theta}_{2n}$ and $\widehat{\Lambda}_{2n}$ are obtained over the 1000 replications (cf. Tables 1-3). For the third model, only $\widehat{\theta}_{2n}$ and $\widehat{\sigma}_{\varepsilon n}^2$ are

obtained, since all inputs are Poisson distributed. The ASEs of all estimates are obtained from the sample estimates of the asymptotic variances in (34)-(34) and (4.3), where the expectations are replaced by their sample mean estimates.

(Table 1 here)

(Table 2 here)

(Table 3 here)

(Table 4 here)

From Tables 1-4 some general conclusions can be drawn. Firstly, the 4SWLE generates fairly good estimates with small bias and small standard errors for the three models. In particular, StDs and ASEs are quite close to each other. Secondly, the 4SWLE is convergent regardless of the placement of the parameters in the strict stationary domain, which is the entire parameter space. In particular, the estimates seem insensitive to the existence or not of the model mean. Other unreported simulations show that the results are consistent with the asymptotic theory in that the larger the sample size, the more accurate the results.

5.2   Application to the Euro-Pound sterling exchange rate data

The RMINAR model with either additive or multiplicative forms (8 and 22) is applied to the number of ticks per minute changes in the Euro to British Pound exchange rate (ExRate for short) on December 12, from 9:00 $a.m.$ to 9:00 $p.m.$. The dataset is taken from Gorgi (2020) who applied (heavy-tailed) observation-driven count models to it. The ExRate series was also considered by Aknouche and Scotto (2024) using the MthINGARCH$(1, 1)$ model. The dataset contains 720 observations and is highly over-dispersed with a sample mean and variance of 13.2153 and 224.2498, respectively (see Figure 1).
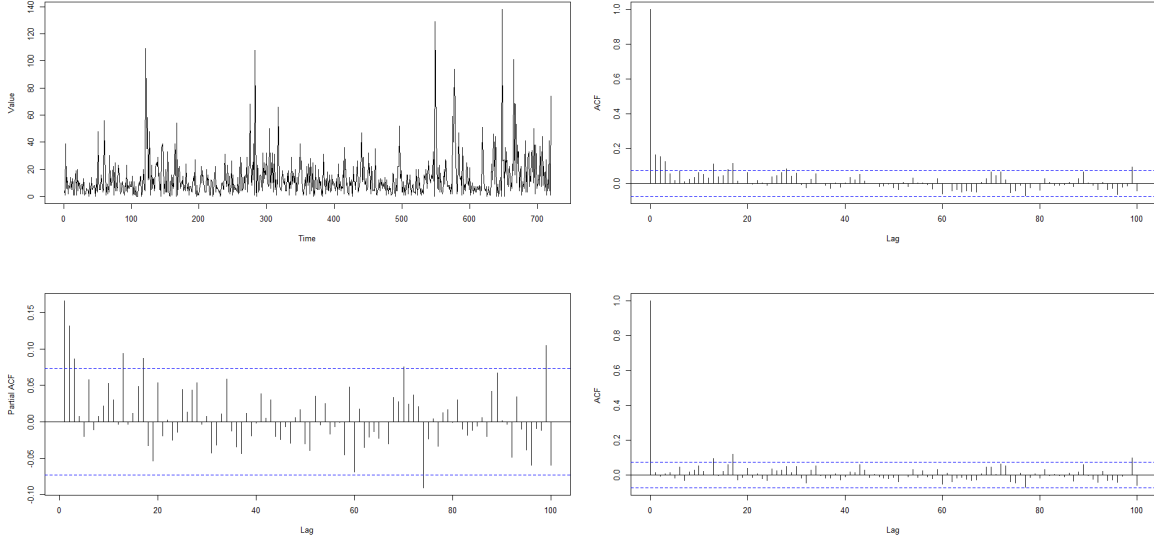
Figure 1: (a) The ExRate series, (b) sample auto-correlation, (c) sample partial auto-correlation, (d) sample auto-correlation of the Pearson residuals of the additive RMINAR(3) model.

(Figure 1 here)

The first step in modeling the ExRate series consists of identifying the order p of the RMINAR(p) models. Since the likelihood of the unspecified RMINAR(p) models is not simple to obtain due to the model random coefficients, standard information-based criteria (AIC, BIC, etc.) are out of reach. However, the additive and multiplicative RMINAR(p) models (8) and (22), having an RCAR form, have the same auto-covariance structure as an AR model since they are both a weak AR(p). Thus, the identification of p can be carried out by inspecting the shape of the sample partial auto-correlation of the series (cf. Figure 1) and reinforced by other evocative criteria. For example, the generated unconditional mean and variance (cf. (19) and (20), respectively) by the additive RMINAR(p) model could be compared to the sample mean and sample variance of the series. Finally, the mean absolute residual (MAR) defined as MAR$:= \frac{1}{n} \sum_{t=1}^{n} |Y_t - \widehat{\mu}_t|$, the mean square residual (MSR) MSR$:= \frac{1}{n} \sum_{t=1}^{n} (Y_t - \widehat{\mu}_t)^2$, and the mean square Pearson residual, MSPR$:= \frac{1}{n} \sum_{t=1}^{n} \frac{(Y_t - \widehat{\mu}_t)^2}{\widehat{V}_t}$ are

30

also used. Note that the sample partial auto-correlation of the series (cf. Figure 5.1) shows a pronounced cut-off after the lag $p = 3$, which suggests that $p = 3$ is an adequate choice, but the values $p = 1$ and $p = 2$ are also possible. Table 5 displays the mean, variance, MAR, and MSPR of the estimated additive and multiplicative RMINAR(p) models for $p = \in \{1, 2\}$. For the additive RMINAR(p) model, $p = 3$ leads to the smallest MAR and MSR while the MSPR is comparable to those given by $p \in \{1, 2\}$. Additionally, $p = 3$ provides the generated mean closest to the sample mean which is not the case for the generated variance. The multiplicative RMINAR(3) model provides the smallest MAR and MSR while the best MSPR is recorded for $p = 1$, which generates the closest mean. So the analysis proceeds with $p = 3$, but following the parsimony principle, $p \in \{1, 2\}$ are also considered for comparison. The choice for optimal p will be based on the out-of-sample forecast ability of the RMINAR(p) model in the set $p \in \{1, 2, 3\}$, as shown below (see Table 11 in the Supplementary material).

(Table 5 here)

The estimation of the RMINAR(p) parameters is performed with the WLS procedure described in Section 4, where the scheme in (28) is directly applied for the additive RMINAR(p) model. Regarding the multiplicative RMINAR(p) model, the procedure (26) is used for simplicity and to inspect the impact of assuming a given variance-to-mean relationship for the random parameters. In particular, the parameter variances are assumed proportional to the parameter means, i.e. $\left(\sigma_\omega^2, \sigma_{\alpha_1}^2, \ldots, \sigma_{\alpha_p}^2\right) = c\left(\omega, \alpha_1, \ldots, \alpha_p\right)$ where the constant of proportionality $c = 0.135$ was estimated as to minimize the |MSPR-1|. For instance, unreported results showed that both the Poisson assumption $\left(\sigma_\omega^2, \sigma_{\alpha_1}^2, \ldots, \sigma_{\alpha_p}^2\right) = \left(\omega, \alpha_1, \ldots, \alpha_p\right)$ and the geometric assumption $\left(\sigma_\omega^2, \sigma_{\alpha_1}^2, \ldots, \sigma_{\alpha_p}^2\right) = \left(\omega, \alpha_1, \ldots, \alpha_p\right) \oplus \left(\omega + 1, \alpha_1 + 1, \ldots, \alpha_p + 1\right)$ give bad MSPR values. Another reason for not applying directly the general nonlinear procedures (36 ii) and (37 ii) for estimating the variances of the multiplicative RMINAR(p) model is that the estimates were sensitive to the choice of the initial parameter values in the optimization routine.

The initial values for the weights used for the estimation procedure (28) follow the general principle of starting the 4SWLSE with arbitrary weights. Therefore, the 4SWLE is applied

31

as many times as possible to arrive at very close successive estimates. At each iteration, the estimated variance parameters are injected as initial weights for the conditional variance in the next iteration. It was concluded that all procedures stabilized at most at the sixth iteration.

Table 6 displays the parameter estimates and estimated asymptotic standard errors for the RMINAR(p) models with $p \in \{1, 2, 3\}$. All parameters are significant and the corresponding models are stationary in mean. In addition, all estimated additive RMINAR(p) models are second-order stationary (SSC) as the SSC condition (17) is verified. Note that the selected additive RMINAR(3) model has, in general, better in-sample properties than the MthINGARCH model (Aknouche and Scotto, 2024) since in the former case, the MAR and MSR are smaller while the MSPR is closer to 1. In addition, the RMINAR(3) generates an unconditional mean that is closer to the sample mean of the series. However, the unconditional variance of the RMINAR(3) is not the best out of the set of considered models, and the RMINAR(1) is that providing a better variance than that of the MthINGARCH model.

(Table 6 here)

Finally, the out-of-sample forecasting ability of the RMINAR(3) models is examined and compared to that of the MthINGARCH(1,1) (Aknouche and Scotto, 2024). The setting $p = 3$ is selected beause it yields the best out-of-sample forecast performance among the RMINAR models (see the Supplementary material). The RMINAR(3) models were estimated using the first n_c observations of the series ($1 < n_c < n = 720$). Then, the one-step ahead forecasts, $\widehat{\mu}_t$, over the period ($n_c + 1, \ldots, n$) were computed. The evaluation was based on i) the mean square forecast error given by MSFE $= \frac{1}{n - n_c} \sum_{t=n_c+1}^{n} (Y_t - \widehat{\mu}_t)^2$, ii) the mean absolute forecast error, MAFE $= \frac{1}{n - n_c} \sum_{t=n_c+1}^{n} |Y_t - \widehat{\mu}_t|$, and iii) the mean squared Pearson forecast error, MSPFE $= \frac{1}{n - n_c} \sum_{t=n_c+1}^{n} \frac{(Y_t - \widehat{\mu}_t)^2}{\widehat{V}_t}$. The estimated conditional variance is given by $\widehat{V}_t := \mathcal{Z}'_{t-1} \widehat{\Lambda}_n$ for the additive model and by $\widehat{V}_t := (\widehat{\sigma}_n^2 + 1) \mathcal{Z}'_{t-1} \widehat{\Delta}_n + \widehat{\sigma}_n^2 \widehat{\mu}_t^2$ for the multiplicative model. Table 7 displays MSFE, MAFE, and MSPFE for the RMINAR models estimated from the series with sample size $n_c \in \{300, 420, 520, 680, 700\}$. The corresponding results

for the MthINGARCH(1,1) are directly taken from Aknouche and Scotto (2024). Note that the additive and multiplicative RMINAR(p) models provide, in general, a comparable out-of-sample ability to the MthINGARCH model. Specifically, for some values of n_c, especially those closer to the sample size $n = 720$, the RMINAR models provide better MSFE and MAFE. In contrast, the MthINGARCH slightly outperforms the RMINARs for smaller n_c values. This could be explained by the fact that the MthINGACH model, having a moving average component, is more persistent (and then has higher memory) than the RMINAR models. Note also that the multiplicative RMINAR provides the worst MSPFE compared to the remaining models due to the assumed variance-to-mean relationship.

(Table 7 here)

5.3 Application to the daily (integer-valued) stock return series

The second application concerns fitting the daily stock returns (Return) of Bank of America from July 1, 2016 to September 28, 2018. The dataset with a total of 566 observations is taken from Xu and Zhu (2022) who divided the returns by tick price to get a signed integer-valued series. The sample mean and variance of the series are 0.1184 and 978.8231, respectively, showing a strong overdispersion (see Figure 2). The authors proposed a $\mathbb{Z}$-valued (GARCH-like) conditional volatility model for the series. Such an approach could ignore the conditional mean effect and, in contrast, the RMINAR model can represent both the conditional mean and conditional volatility observed in the data.

(Figure 2 here)

The sample partial auto-correlation of the series shows a small and isolated peak at lag 3 while the remaining auto-correlations are fairly small. Thus, as in the above ExRate example, the orders $p \in \{1, 2, 3\}$ were considered. Table 8 shows some measures to support the choice of the model order p. It can be seen that $p = 1$ gives the best mean and variance approximations, while $p = 3$ provides the smallest MAR and MSR and simultaneously the worst MSPR.
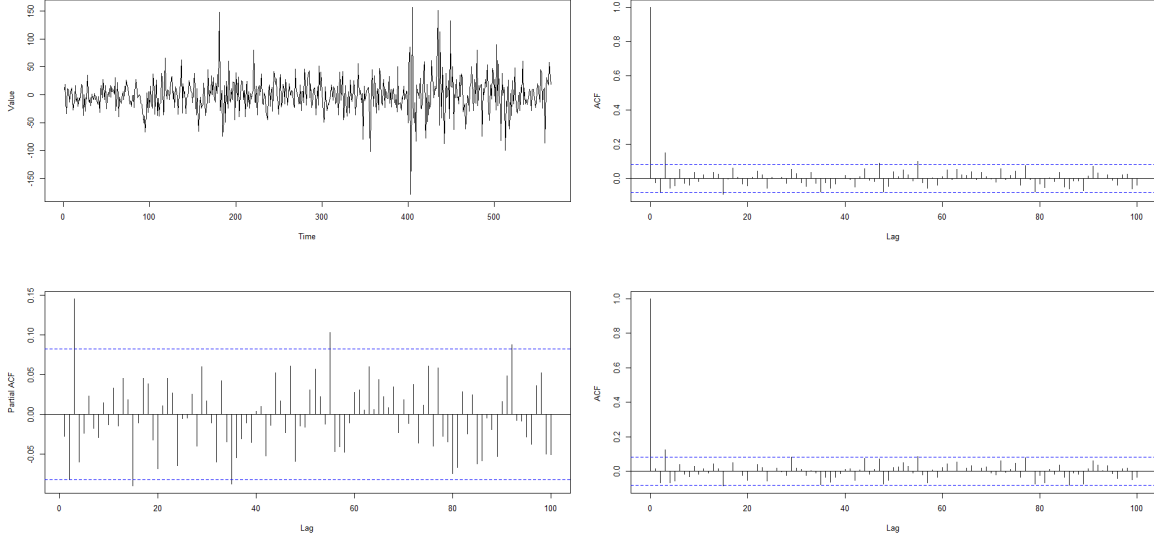
Figure 2: (a) The Return series, (b) sample auto-correlation, (c) sample partial auto-correlation, (d) sample auto-correlation of the Pearson residuals of the $\mathbb{Z}$-valued RMINAR(1) model.

(Table 8 here)

Parameter estimation is carried out using the algorithm (28) with results reported in Table 9. All estimated models are second-order stationary and all parameters are significant except $\sigma^2_{\phi_3}$ which is zero due to the artificial non-negative least squares constraint used. Thus the random coefficient Φ_{t3} is degenerated at $\widehat{\phi}_3 = 0.1021$ and since Φ_{t3} should be integer valued this implies it should be zero, which excludes the order $p = 3$. On the other hand, the retained orders $p = 1$ and $p = 2$ provide comparable in-sample performance (cf. Table 8).

(Table 9 here)

The truncated series with $n_c \in \{300, 350, 400, 450, 500\}$ are used to analyze the out-of-sample performance of the RMINAR(p) model. Table 10 shows that the order $p = 2$ provides the smallest MSFE and MAFE, for the sizes $n_c \in \{300, 350, 400, 550\}$. However, the order $p = 1$ gives better measures when $n_c \in \{450, 500\}$. Overall, there is no dominant order for all sample sizes n_c and $p = 1$ is be preferred following the parsimony principle.

(Table 10 here)

6 Conclusion

This paper proposed a random multiplication operator (RMO) to construct integer-valued time series models in both $\mathbb{N}$-valued and $\mathbb{Z}$-valued cases. Compared to the random sum operator (RSO) which generally requires specifying the full distribution (e.g. binomial, geometric, Poisson, etc.), the RMO is semi-parametric in the sense that it only specifies the mean and variance. Additionally, the RMO allows larger over-dispersion and can be more heavy tailed but, more importantly, the variance does not necessarily have to depend on the mean as is the case for most RSOs. This allows more flexibility in modeling since the operator variance is estimated in a separate procedure independently of the operator mean. Finally, the RMO is simpler than the RSO because it consists of the usual multiplication of the constant and variable operands.

This paper also shows how to build up simple and tractable RMO-based integer-valued time series models that consist of auto-regressive type models with integer-valued random coefficients. Although random coefficient time series models have been well-known for a long time, their use was entirely focused on real-valued random coefficients. In contrast, integer-valued random coefficients make the underlying model universally (or everywhere) stationary and ergodic when generated from stationary and ergodic inputs. An important consequence is that most important M estimation methods such as quasi-maximum likelihood and weighted least squares (WLS) are consistent and asymptotically Normal everywhere for all parameter components. This makes the models more flexible than other existing classes of integer-valued time series models mentioned in the introduction.

The RMINAR framework can be extended in many perspectives. Firstly, the persistence ability of the RMINAR models could be improved by proposing INARMA models based on random multiplications with some moving average dynamics. In particular, INGARCH-type models based on thinning (Aknouche and Scotto, 2024) could be adapted while replacing the

RSO with the RMO. Second, multivariate RMINAR forms could be easily introduced and estimated using the methods in Nicholls and Quinn (1982) and Praskova and Vanecek (2011). These models for both $\mathbb{N}$-valued and $\mathbb{Z}$-valued series could constitute very straightforward alternatives to multivariate INGARCH models and multivariate INAR-type models.

Funding

Appendices

Appendix A   Proof of Theorem 1

The proof of Theorem 1 relies on Lemma 1 and the proofs of 2 and 3 follow similarly.

Lemma 1 *Consider the stochastic recurrence equation*

$$Z_t = A_t Z_{t-1} + B_t, \quad t \in \mathbb{Z},$$

driven by the iid sequence of pairs $\{(A_t, B_t), t \in \mathbb{Z}\}$ *such that* A_t *is the companion matrix defined in* (11) *and* $\{a_{it}, \ t \in \mathbb{Z}\}$ $(i = 1, \ldots, p)$ *are iid and mutually independent, satisfying*

$$\mathbb{P}\left((a_{1t}, \ldots, a_{pt})' = \mathbf{0}_{p \times 1}\right) = \prod_{i=1}^{p} \mathbb{P}\left(a_{it} = 0\right) > 0. \tag{43}$$

Then, the random series

$$\sum_{j=0}^{\infty} \prod_{i=0}^{j-1} A_{t-i} B_{t-j}, \tag{44}$$

36

converges absolutely a.s. for all $t \in \mathbb{Z}$.

Proof. (Lemma 1)

For $p = 1$, the result in (44) directly follows from Vervaat (1979, Theorem 1.6 (c); Lemma 1.7). See also Brandt (1986, Theorem 1 and p. 215) for the case of stationary and ergodic non-necessarily independent sequences. For $p > 1$, by (43), the iid property of $\{(A_t, B_t), t \in \mathbb{Z}\}$, the mutual independence of $\{a_{it}, \ t \in \mathbb{Z}\}$ $(i = 1, \ldots, p)$, and the form (11) of the companion matrix A_t, it follows that

$$\mathbb{P}\left(A_t A_{t-1} \cdots A_{t-p+1} = \mathbf{0}_p\right) = \prod_{i=1}^{p} \mathbb{P}\left(a_{it} = 0\right) > 0.$$

Indeed, it can be easily seen that the matrix $A_t A_{t-1} \cdots A_{t-p+1}$ contains no element 1 and that all of its elements are formed by algebraic operations of $\{(a_{1k}, \ldots, a_{pk}), \ k = t - p + 1, \ldots, t\}$ vectors that satisfy (43). Therefore,

$$\#\left\{j \in \mathbb{N}_0 : \prod_{i=0}^{j-1} A_{t-i} B_{t-j} = 0\right\} = \infty \text{ a.s.},$$

where $\#A \in \mathbb{N}_0 \cup \{\infty\}$ is the number of elements of the set A. Thus, the series in (44) has only finitely many positive terms a.s., which implies its a.s. absolute convergence. $\blacksquare$

Proof. (Theorem 1)

The fact that

$$\sum_{j=0}^{\infty} \prod_{i=0}^{j-1} A_{t-i} B_{t-j} < \infty \text{ a.s.},$$

immediately follows from Lemma 1. By definition, $\mathbf{Y}_t := \sum_{j=0}^{\infty} \prod_{i=0}^{j-1} A_{t-i} B_{t-j}$, $t \in \mathbb{Z}$ and then $\{\mathbf{Y}_t, \ t \in \mathbb{Z}\}$ is a solution of the RMINAR model in the vector form (12). Such a solution is unique and causal (i.e. future independent). Moreover, it is strictly stationary and ergodic in view of the iid property of the random inputs of the RMINAR model (8). $\blacksquare$

Appendix B   Proof of Theorems 4, 5 and 6

The proof of Theorem 4 is first presented by introducing some lemmas (2-6) that support the results in the theorem. Then, the proof of Theorem 5 is omitted as it is quite similar to that of Theorem 4. The proofs of some results conveyed in Theorem 6 are also omitted as they can be derived similarly as in the proof of Theorem 4.

Lemma 2 *Consider the RMINAR model* (8) *under* **A0**. *For all positive real* (a_i) $(i = 0, \ldots, p)$ *and all integers* r, s, k *such that* $r + s \leq 2k$, *it holds that*

$$\frac{1}{n} \sum_{t=1}^{n} \frac{Y_{t-i}^r Y_{t-j}^s}{\left(a_0 + a_1 Y_{t-1}^2 + \cdots + a_p Y_{t-p}^2\right)^k} \xrightarrow[n \to \infty]{a.s.} \mathbb{E}\left(\frac{Y_{t-i}^r Y_{t-j}^s}{\left(a_0 + a_1 Y_{t-1}^2 + \cdots + a_p Y_{t-p}^2\right)^k}\right), \ i, j = 1, \ldots, p. \tag{45}$$

Proof. (Lemma 2)

The result in (45) follows from the stationar and ergodicity of $\{Y_t, t \in \mathbb{Z}\}$ under **A0**, the a.s. boundedness of $\frac{Y_{t-i}^r Y_{t-j}^s}{\left(a_0 + a_1 Y_{t-1}^2 + \cdots + a_p Y_{t-p}^2\right)^k}$ which entails $\mathbb{E}\left(\frac{Y_{t-i}^r Y_{t-j}^s}{\left(a_0 + a_1 Y_{t-1}^2 + \cdots + a_p Y_{t-p}^2\right)^k}\right) < \infty$ and the ergodic theorem. ∎

Lemma 3 *Consider the RMINAR model* (8) *under* **A0**. *Then, for all* $\Lambda_* > \mathbf{0}_{p+q+1}$,

$$\frac{1}{n} \sum_{t=1}^{n} \mathcal{Z}_{t-1} \left(\frac{\left(Y_t - \mathcal{Y}_{t-1}'\widehat{\theta}_{1n}\right)^2}{\left(\mathcal{Z}_{t-1}'\Lambda_*\right)^2} - \frac{\left(Y_t - \mathcal{Y}_{t-1}'\theta_0\right)^2}{\left(\mathcal{Z}_{t-1}'\Lambda_*\right)^2}\right) = o_{a.s.}(1), \tag{46}$$

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \mathcal{Z}_{t-1} \left(\frac{\left(Y_t - \mathcal{Y}_{t-1}'\widehat{\theta}_{1n}\right)^2}{\left(\mathcal{Z}_{t-1}'\Lambda_*\right)^2} - \frac{\left(Y_t - \mathcal{Y}_{t-1}'\theta_0\right)^2}{\left(\mathcal{Z}_{t-1}'\Lambda_*\right)^2}\right) = o_p(1), \tag{47}$$

where $o_{a.s.}(1)$ *and* $o_p(1)$ *converge to zero, respectively, as* $n \to \infty$ *a.s. and in probability.*

Proof. (Lemma 3)

i) A Taylor expansion of the function $g\left(\widehat{\theta}_{1n}\right) = \left(Y_t - \mathcal{Y}_{t-1}'\widehat{\theta}_{1n}\right)^2$ around θ_0 gives

$$\frac{1}{n} \sum_{t=1}^{n} \mathcal{Z}_{t-1} \frac{\left(Y_t - \mathcal{Y}_{t-1}'\widehat{\theta}_{1n}\right)^2}{\left(\mathcal{Z}_{t-1}'\Lambda_*\right)^2} = \frac{1}{n} \sum_{t=1}^{n} \frac{\left(Y_t - \mathcal{Y}_{t-1}'\theta_0\right)^2}{\left(\mathcal{Z}_{t-1}'\Lambda_*\right)^2} + \frac{1}{n} \sum_{t=1}^{n} \frac{2\left(Y_t - \mathcal{Y}_{t-1}'\theta_u\right)}{\left(\mathcal{Z}_{t-1}'\Lambda_*\right)^2} \mathcal{Y}_{t-1}' \left(\widehat{\theta}_{1n} - \theta_0\right). \tag{48}$$

Therefore, the result (46) follows from the a.s. boundedness of

$$\frac{2\left(Y_t - \mathcal{Y}_{t-1}'\theta_u\right)}{\left(\mathcal{Z}_{t-1}'\Lambda_*\right)^2} \mathcal{Y}_{t-1}',$$

which entails the finiteness of $\mathbb{E}\left(\frac{2\left(Y_t - \mathcal{Y}'_{t-1}\theta_u\right)}{\left(\mathcal{Z}'_{t-1}\Lambda_*\right)^2}\mathcal{Y}'_{t-1}\right)$, the ergodic theorem and the strong consistency of $\widehat{\theta}_{1n}$.

ii) One can write $\widehat{\theta}_{1n} - \theta_0 = \frac{1}{\sqrt{n}}O_p\left(1\right)$, where $O_p\left(1\right)$ denotes a term bounded in probability. Hence (48) becomes

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\mathcal{Z}_{t-1}\frac{\left(Y_t - \mathcal{Y}'_{t-1}\widehat{\theta}_{1n}\right)^2}{\left(\mathcal{Z}'_{t-1}\Lambda_*\right)^2} = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\left(Y_t - \mathcal{Y}'_{t-1}\theta_0\right)^2}{\left(\mathcal{Z}'_{t-1}\Lambda_*\right)^2} + \frac{1}{n}\sum_{t=1}^{n}\frac{2\left(Y_t - \mathcal{Y}'_{t-1}\theta_u\right)}{\left(\mathcal{Z}'_{t-1}\Lambda_*\right)^2}\mathcal{Y}'_{t-1}O_p\left(1\right),$$

so the result (47) follows from the ergodic theorem, the consistency of $\widehat{\theta}_{1n}$ and the fact that θ_u is between $\widehat{\theta}_{1n}$ and θ_0. ∎

Lemma 4 *Consider the RMINAR model* (8) *under* **A0**. *Then,*

$$\frac{1}{n}\sum_{t=1}^{n}\mathcal{Y}_{t-1}Y_t\left(\frac{1}{\left(\mathcal{Z}'_{t-1}\widehat{\Lambda}_{1n}\right)^2} - \frac{1}{\left(\mathcal{Z}'_{t-1}\Lambda_0\right)^2}\right) = o_{a.s.}\left(1\right), \tag{49}$$

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\mathcal{Y}_{t-1}Y_t\left(\frac{1}{\left(\mathcal{Z}'_{t-1}\widehat{\Lambda}_{1n}\right)^2} - \frac{1}{\left(\mathcal{Z}'_{t-1}\Lambda_0\right)^2}\right) = o_p\left(1\right), \tag{50}$$

$$\frac{1}{n}\sum_{t=1}^{n}\left(\frac{1}{\left(\mathcal{Z}'_{t-1}\widehat{\Lambda}_{1n}\right)^2} - \frac{1}{\left(\mathcal{Z}'_{t-1}\Lambda_0\right)^2}\right)\mathcal{Y}_{t-1}\mathcal{Y}'_{t-1} = o_{a.s.}\left(1\right). \tag{51}$$

Proof. (Lemma 4)
Similarly to the proof of condition (46), a Taylor expansion of the function $g\left(\widehat{\theta}_{1n}\right) = \frac{1}{\left(\mathcal{Z}'_{t-1}\widehat{\Lambda}_{1n}\right)^2}$ around θ_0 shows that $I = o_{a.s.}\left(1\right)$. The proofs of (50) and (51) follow in the same way as for the result (49). ∎

Lemma 5 *Consider the RMINAR model* (8) *under* **A0**. *Then,*

$$\frac{1}{n}\sum_{t=1}^{n}\mathcal{Z}_{t-1}\left(\frac{\left(Y_t - \mathcal{Y}'_{t-1}\widehat{\theta}_{2n}\right)^2}{\left(\mathcal{Z}'_{t-1}\widehat{\Lambda}_{1n}\right)^2} - \frac{\left(Y_t - \mathcal{Y}'_{t-1}\theta_0\right)^2}{\left(\mathcal{Z}'_{t-1}\Lambda_0\right)^2}\right) = o_{a.s.}\left(1\right), \tag{52}$$

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\mathcal{Z}_{t-1}\left(\frac{\left(Y_t - \mathcal{Y}'_{t-1}\widehat{\theta}_{2n}\right)^2}{\left(\mathcal{Z}'_{t-1}\widehat{\Lambda}_{1n}\right)^2} - \frac{\left(Y_t - \mathcal{Y}'_{t-1}\theta_0\right)^2}{\left(\mathcal{Z}'_{t-1}\Lambda_0\right)^2}\right) = o_p\left(1\right), \tag{53}$$

$$\frac{1}{n}\sum_{t=1}^{n}\left(\frac{1}{\left(\mathcal{Z}'_{t-1}\widehat{\Lambda}_{1n}\right)^2} - \frac{1}{\left(\mathcal{Z}'_{t-1}\Lambda_0\right)^2}\right)\mathcal{Z}_{t-1}\mathcal{Z}'_{t-1} = o_{a.s.}\left(1\right). \tag{54}$$

Proof. (Lemma 5)

The proof of condition (52) uses the same device as that used in the proof of (49), giving

$$\frac{1}{n} \sum_{t=1}^{n} \mathcal{Z}_{t-1} \left(\frac{\left(Y_t - \mathcal{Y}'_{t-1}\widehat{\theta}_{2n}\right)^2}{\left(\mathcal{Z}'_{t-1}\widehat{\Lambda}_{1n}\right)^2} - \frac{\left(Y_t - \mathcal{Y}'_{t-1}\theta_0\right)^2}{\left(\mathcal{Z}'_{t-1}\Lambda_0\right)^2} \right) = o_{a.s.} (1) .$$

Results (53) and (54) follow using similar arguments. ■

Lemma 6 *Consider the RMINAR model (8) under **A0**, $\mathbb{E}\left(\varepsilon_t^4\right) < \infty$ and $\mathbb{E}\left(\Phi_{it}^4\right) < 1$ (i = 1,\ldots, p) then, for all $\Lambda_* > \mathbf{0}_{p+q+1}$,*

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\mathcal{Y}_{t-1}e_t}{\mathcal{Z}'_{t-1}\Lambda_*} \xrightarrow[n\to\infty]{D} \mathcal{N}\left(0, B\left(\Lambda_0, \Lambda_*\right)\right), \tag{55}$$

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\mathcal{Z}_{t-1}u_t}{\left(\mathcal{Z}'_{t-1}\Lambda_*\right)^2} \xrightarrow[n\to\infty]{D} \mathcal{N}\left(0, D\left(\Lambda_*\right)\right). \tag{56}$$

Proof. (Lemma 6)

i) By **A0** and the ergodic theorem, it follows that

$$\sum_{t=1}^{n} \left(\frac{1}{\sqrt{n}} \frac{\mathcal{Y}_{t-1}e_t}{\mathcal{Z}'_{t-1}\Lambda_*} \right) \left(\frac{1}{\sqrt{n}} \frac{\mathcal{Y}_{t-1}e_t}{\mathcal{Z}'_{t-1}\Lambda_*} \right)' = \frac{1}{n} \sum_{t=1}^{n} \frac{e_t^2 \mathcal{Y}_{t-1}\mathcal{Y}'_{t-1}}{\left(\mathcal{Z}'_{t-1}\Lambda_*\right)^2} \xrightarrow[n\to\infty]{a.s.} B\left(\Lambda_0, \Lambda_*\right),$$

where $B\left(\Lambda_0, \Lambda_*\right)$ is finite by the a.s. boundedness of $\frac{\mathcal{Z}'_{t-1}\Lambda_0\mathcal{Y}_{t-1}\mathcal{Y}'_{t-1}}{\left(\mathcal{Z}'_{t-1}\Lambda_*\right)^2}$ and the finiteness of the second moments of the random coefficients. Since $\{e_t, t \in \mathbb{Z}\}$ is square integrable martingale difference with respect to $\{\mathcal{F}_t^Y, t \in \mathbb{Z}\}$, then (55) follows from the central limit theorem (CLT) for square-integrable martingales (e.g. Billingsley, 2008; Hall and Heyde, 1980).

ii) Note that

$$\sum_{t=1}^{n} \left(\frac{1}{\sqrt{n}} \frac{\mathcal{Z}_{t-1}u_t}{\left(\mathcal{Z}'_{t-1}\Lambda_*\right)^2} \right) \left(\frac{1}{\sqrt{n}} \frac{\mathcal{Z}_{t-1}u_t}{\left(\mathcal{Z}'_{t-1}\Lambda_*\right)^2} \right)' = \frac{1}{n} \sum_{t=1}^{n} \frac{u_t^2 \mathcal{Z}_{t-1}\mathcal{Z}'_{t-1}}{\left(\mathcal{Z}'_{t-1}\Lambda_*\right)^4} \xrightarrow[n\to\infty]{a.s.} D\left(\Lambda_*\right).$$

Hence (56) is a consequence of the CLT for the square-integrable $\{\mathcal{F}_t^Y, t \in \mathbb{Z}\}$-martingale difference $\{u_t, t \in \mathbb{Z}\}$. ■

Next, follows the proof of Theorem 4 based on the established lemmas 2 to 6.

Proof. (Theorem 4)

a) For the proof of (34)-i, combining (8) and (30) leads to

$$\widehat{\theta}_{1n} - \theta_0 = \left(\frac{1}{n} \sum_{t=1}^{n} \frac{\mathcal{Y}_{t-1} \mathcal{Y}'_{t-1}}{\mathcal{Z}'_{t-1} \Lambda_*} \right)^{-1} \frac{1}{n} \sum_{t=1}^{n} \mathcal{Y}_{t-1} \frac{e_t}{\mathcal{Z}'_{t-1} \Lambda_*}, \tag{57}$$

so the result (34)-i follows from Lemma 2 while using the fact that

$$\mathbb{E}\left(\mathcal{Y}_{t-1} \frac{e_t}{\mathcal{Z}'_{t-1} \Lambda_*} \right) = \mathbb{E}\left(\mathcal{Y}_{t-1} \frac{\mathbb{E}\left(e_t | \mathcal{F}^Y_{t-1} \right)}{\mathcal{Z}'_{t-1} \Lambda_*} \right) = 0.$$

b) For the proof of (34)-ii, in view of (31), the regression

$$\left(Y_t - \mathcal{Y}'_{t-1} \theta_0 \right)^2 = \mathcal{Z}'_{t-1} \Lambda_0 + u_t, \tag{58}$$

and Lemma 3, it follows that

$$\widehat{\Lambda}_{1n} - \Lambda_0 = \left(\frac{1}{n} \sum_{t=1}^{n} \frac{\mathcal{Z}_{t-1} \mathcal{Z}'_{t-1}}{\left(\mathcal{Z}'_{t-1} \Lambda_* \right)^2} \right)^{-1} \frac{1}{n} \sum_{t=1}^{n} \mathcal{Z}_{t-1} \frac{u_t}{\left(\mathcal{Z}'_{t-1} \Lambda_* \right)^2} + o_{a.s.}(1).$$

Therefore, the result (34)-ii follows from Lemma 2 and the fact that

$$\mathbb{E}\left(\frac{u_t}{\left(\mathcal{Z}'_{t-1} \Lambda_* \right)^2} \right) = \mathbb{E}\left(\frac{\mathbb{E}\left(u_t | \mathcal{F}^Y_{t-1} \right)}{\left(\mathcal{Z}'_{t-1} \Lambda_* \right)^2} \right) = 0.$$

c) The proof of (34)-iii is based on (32), (8) and Lemma 4, leading to

$$\widehat{\theta}_{2n} - \theta_0 = \left(\frac{1}{n} \sum_{t=1}^{n} \frac{\mathcal{Y}_{t-1} \mathcal{Y}'_{t-1}}{\mathcal{Z}'_{t-1} \Lambda_0} \right)^{-1} \frac{1}{n} \sum_{t=1}^{n} \mathcal{Y}_{t-1} \frac{e_t}{\mathcal{Z}'_{t-1} \Lambda_0} + o_{a.s.}(1),$$

so the result follows from Lemma 2 in the same way as (34)-i.

d) The proof of (34)-iv follows by combining (32), (58) and Lemma 5 to obtain

$$\widehat{\Lambda}_{2n} - \Lambda_0 = \left(\frac{1}{n} \sum_{t=1}^{n} \frac{\mathcal{Z}_{t-1} \mathcal{Z}'_{t-1}}{\left(\mathcal{Z}'_{t-1} \Lambda_0 \right)^2} \right)^{-1} \frac{1}{n} \sum_{t=1}^{n} \frac{\mathcal{Z}_{t-1} u_t}{\left(\mathcal{Z}'_{t-1} \Lambda_0 \right)^2} + o_{a.s.}(1)$$

and the result is a consequence of Lemma 2 using the same argument in proving (34)-ii.

e) For the proof of (35)-i, condition (57) is rewritten as

$$\sqrt{n}\left(\widehat{\theta}_{1n} - \theta_0 \right) = \left(\frac{1}{n} \sum_{t=1}^{n} \frac{\mathcal{Y}_{t-1} \mathcal{Y}'_{t-1}}{\mathcal{Z}'_{t-1} \Lambda_*} \right)^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\mathcal{Y}_{t-1} e_t}{\mathcal{Z}'_{t-1} \Lambda_*},$$

41

and the result follows from Lemma 2 and Lemma 6 while using Slutsky's Lemma.

f) For the proof of (35)-ii, from (32) and Lemma 4, one obtains

$$\sqrt{n}\left(\widehat{\theta}_{2n} - \theta_0\right) = \left(\frac{1}{n}\sum_{t=1}^{n}\frac{\mathcal{Y}_{t-1}\mathcal{Y}'_{t-1}}{\mathcal{Z}'_{t-1}\Lambda_0}\right)^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\mathcal{Y}_{t-1}e_t}{\mathcal{Z}'_{t-1}\Lambda_0} + o_p(1),$$

and the result follows from Lemma 6 and Lemma 2.

g) For the proof of (35)-iii, consider the Lemma 3. One can write

$$\sqrt{n}\left(\widehat{\Lambda}_{1n} - \Lambda_0\right) = \left(\frac{1}{n}\sum_{t=1}^{n}\frac{\mathcal{Z}_{t-1}\mathcal{Z}'_{t-1}}{\left(\mathcal{Z}'_{t-1}\Lambda_*\right)^2}\right)^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\mathcal{Z}_{t-1}u_t}{\left(\mathcal{Z}'_{t-1}\Lambda_*\right)^2} + o_p(1),$$

with obvious notation. Hence, the result (35)-ii follows from Lemma 6 and Lemma 2.

h) The proof of (35)-iv relies on (33) and Lemma 5. Then,

$$\sqrt{n}\left(\widehat{\Lambda}_{2n} - \Lambda_0\right) = \left(\frac{1}{n}\sum_{t=1}^{n}\frac{\mathcal{Z}_{t-1}\mathcal{Z}'_{t-1}}{\left(\mathcal{Z}'_{t-1}\Lambda_0\right)^2}\right)^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\mathcal{Z}_{t-1}u_t}{\left(\mathcal{Z}'_{t-1}\Lambda_0\right)^2} + o_p(1)$$

and the result follows from Lemma 6 (or 56 with $\Lambda_* = \Lambda_0$) and Lemma 2. This concludes the proof of the theorem. ■

The proof of the results conveyed in Theorem 6 is based on the intermediary results stated in the following lemmas. Consider the following notation

$$L_n\left(\Lambda, \Lambda_*, \widehat{\theta}_{1n}\right) = \frac{1}{n}\sum_{t=1}^{n}l_t\left(\Lambda, \Lambda_*, \widehat{\theta}_{1n}\right)$$

$$l_t\left(\Lambda, \Lambda_*, \widehat{\theta}_{1n}\right) = \frac{u_t^2\left(\widehat{\theta}_{1n}, \Lambda\right)}{V_t^2\left(\widehat{\theta}_{1n}, \Lambda_*\right)} = \frac{\left(\left(Y_t - \mathcal{Y}'_{t-1}\widehat{\theta}_{1n}\right)^2 - V_t\left(\widehat{\theta}_{1n}, \Lambda\right)\right)^2}{V_t^2\left(\widehat{\theta}_{1n}, \Lambda_*\right)}$$

where $V_t(\theta_0, \Lambda_0) := \mathbb{V}\left(Y_t | \mathcal{F}_{t-1}^Y\right)$.

Lemma 7 *Under **A0****, it holds that*

$$\sup_{\Lambda \in \Pi}\left|L_n\left(\Lambda, \Lambda_*, \widehat{\theta}_{1n}\right) - L_n\left(\Lambda, \Lambda_*, \theta_0\right)\right| \overset{a.s.}{\underset{n\to\infty}{\to}} 0.$$

Proof. (Lemma 7)

It holds that

$$\left| L_n\left(\Lambda, \Lambda_*, \widehat{\theta}_{1n}\right) - L_n\left(\Lambda, \Lambda_*, \theta_0\right) \right|$$

$$\leq \frac{1}{n}\sum_{t=1}^{n}\left| \frac{\left(\left(Y_t - \mathcal{Y}'_{t-1}\widehat{\theta}_{1n}\right)^2 - V_t\left(\widehat{\theta}_{1n}, \Lambda\right)\right)^2}{\left(V_t\left(\widehat{\theta}_{1n}, \Lambda_*\right)\right)^2} - \frac{\left(\left(Y_t - \mathcal{Y}'_{t-1}\theta_0\right)^2 - V_t\left(\theta_0, \Lambda\right)\right)^2}{\left(V_t\left(\theta_0, \Lambda_*\right)\right)^2} \right|.$$

Using the Taylor expansion of the function $g_1\left(\widehat{\theta}_{1n}\right) = \dfrac{\left(\left(Y_t - \mathcal{Y}'_{t-1}\widehat{\theta}_{1n}\right)^2 - V_t\left(\widehat{\theta}_{1n}, \Lambda\right)\right)^2}{\left(V_t\left(\widehat{\theta}_{1n}, \Lambda_*\right)\right)^2}$ around θ_0
as in Lemma 3, the result follows from the a.s. boundedness of $\frac{\partial}{\partial\theta'}g_1\left(\theta_u\right)$ and the strong
consistency of $\widehat{\theta}_{1n}$. $\blacksquare$

Lemma 8 *Under **A0****, it holds that*

$$i)\ \mathbb{E}\left(l_t\left(\Lambda, \Lambda_*, \theta_0\right)\right) < \infty,$$

$$ii)\ \mathbb{E}\left(l_t\left(\Lambda, \Lambda_*, \theta_0\right)\right) \leq \mathbb{E}\left(l_t\left(\Lambda_0, \Lambda_*, \theta_0\right)\right)\ \textit{for all } \Lambda \in \Pi,$$

$$iii)\ \mathbb{E}\left(l_t\left(\Lambda, \Lambda_*, \theta_0\right)\right) = \mathbb{E}\left(l_t\left(\Lambda_0, \Lambda_*, \theta_0\right)\right) \Longrightarrow \Lambda = \Lambda_0.$$

Proof. (Lemma 8)

i) The result obviously follows from the boundedness of $\dfrac{u_t^2\left(\widehat{\theta}_{1n}, \Lambda\right)}{V_t^2\left(\widehat{\theta}_{1n}, \Lambda_*\right)}$.

ii) Standard arguments show that for all $\Lambda \in \Pi$

$$
\begin{aligned}
\mathbb{E}\left(l_t\left(\Lambda, \Lambda_*, \theta_0\right)\right) &= \mathbb{E}\left(\frac{\left(e_t^2 - V_t(\theta_0, \Lambda_0) + V_t(\theta_0, \Lambda_0) - V_t(\theta_0, \Lambda)\right)^2}{V_t^2\left(\widehat{\theta}_{1n}, \Lambda_*\right)}\right)\\
&= \mathbb{E}\left(\frac{\left(e_t^2 - V_t(\theta_0, \Lambda_0)\right)^2}{V_t^2\left(\widehat{\theta}_{1n}, \Lambda_*\right)}\right) + \mathbb{E}\left(\frac{\left(V_t(\theta_0, \Lambda_0) - V_t(\theta_0, \Lambda)\right)^2}{V_t^2\left(\widehat{\theta}_{1n}, \Lambda_*\right)}\right)\\
&\geq \mathbb{E}\left(\frac{\left(e_t^2 - V_t(\theta_0, \Lambda_0)\right)^2}{V_t^2\left(\widehat{\theta}_{1n}, \Lambda_*\right)}\right) = \mathbb{E}\left(l_t\left(\Lambda_0, \Lambda_*, \theta_0\right)\right).
\end{aligned}
\tag{59}
$$

iii) Inequality (59) becomes an equality if and only if

$$\mathbb{E}\left(\frac{\left(V_t(\theta_0, \Lambda_0) - V_t(\theta_0, \Lambda)\right)^2}{V_t^2\left(\widehat{\theta}_{1n}, \Lambda_*\right)}\right) = 0,$$

which holds if and only if

$\dfrac{\left(V_t(\theta_0, \Lambda_0) - V_t(\theta_0, \Lambda)\right)^2}{u_t^2\left(\widehat{\theta}_{1n}, \Lambda_*\right)} = 0 \Leftrightarrow V_t\left(\theta_0, \Lambda\right) = V_t\left(\theta_0, \Lambda_0\right)$

$$\Leftrightarrow \sigma_\varepsilon^2 \left(\left(\mathcal{Y}_{t-1}' \theta_0 \right)^2 + \delta_t^2 (\Delta) \right) + \delta_t^2 (\Delta) = \sigma_{0\varepsilon}^2 \left(\left(\mathcal{Y}_{t-1}' \theta_0 \right)^2 + \delta_t^2 (\Delta_0) \right) + \delta_t^2 (\Delta_0)$$

$$\Leftrightarrow \sigma_\varepsilon^2 \left(\left(\mathcal{Y}_{t-1}' \theta_0 \right)^2 + \mathcal{Z}_{t-1}' \Delta \right) + \mathcal{Z}_{t-1}' \Delta = \sigma_{0\varepsilon}^2 \left(\left(\mathcal{Y}_{t-1}' \theta_0 \right)^2 + \mathcal{Z}_{t-1}' \Delta_0 \right) + \mathcal{Z}_{t-1}' \Delta_0$$

$$\Leftrightarrow \sigma_\varepsilon^2 \left(\left(\mathcal{Y}_{t-1}' \theta_0 \right)^2 + \mathcal{Z}_{t-1}' \Delta \right) = \sigma_{0\varepsilon}^2 \left(\left(\mathcal{Y}_{t-1}' \theta_0 \right)^2 + \mathcal{Z}_{t-1}' \Delta_0 \right) \text{ and } \mathcal{Z}_{t-1}' \Delta = \mathcal{Z}_{t-1}' \Delta_0$$

$$\Leftrightarrow \sigma_\varepsilon^2 = \sigma_{0\varepsilon}^2 \text{ and } \Delta = \Delta_0 \Leftrightarrow \Lambda = \Lambda_0. \quad \blacksquare$$

Lemma 9 *Under **A0**** for all $\Lambda \neq \Lambda_0$, there is a neighborhood $\mathcal{V}(\Lambda)$ such that*

$$\liminf_{n\to\infty} \inf_{\Lambda_1 \in \mathcal{V}(\Lambda)} L_n \left(\Lambda_1, \Lambda_*, \widehat{\theta}_{1n} \right) > \liminf_{n\to\infty} L_n \left(\Lambda_0, \Lambda_*, \widehat{\theta}_{1n} \right) a.s.$$

Proof. (Lemma 9)

For all $\overline{\Lambda} \in \Pi$ and $k \in \mathbb{N}^*$, let $\mathcal{V}_k(\overline{\Lambda})$ be the open ball with center $\overline{\Lambda}$ and radius $1/k$. Since $\inf_{\Lambda \in \mathcal{V}_k(\overline{\Lambda}) \cap \Pi} l_t (\Lambda, \Lambda_*, \theta_0)$ is a measurable function of the terms of $\{Y_t, t \in \mathbb{Z}\}$, which is strictly stationary and ergodic under **A0**, then $\left\{ \inf_{\Lambda \in \mathcal{V}_k(\overline{\Lambda}) \cap \Pi} l_t (\Lambda, \Lambda_*, \theta_0), \ t \in \mathbb{Z} \right\}$ is also strictly stationary and ergodic where, by Lemma 8,

$$\mathbb{E} \left(\inf_{\Lambda \in \mathcal{V}_k(\overline{\Lambda}) \cap \Pi} l_t (\Lambda, \Lambda_*, \theta_0) \right) \in [-\infty, +\infty[.$$

Therefore, in view of Lemma 7 and the ergodic theorem (Billingsley, 2008), it follows that

$$\liminf_{n\to\infty} \inf_{\Lambda \in \mathcal{V}_k(\overline{\Lambda}) \cap \Pi} L_n \left(\Lambda, \Lambda_*, \widehat{\theta}_{1n} \right) = \liminf_{n\to\infty} \inf_{\Lambda \in \mathcal{V}_k(\overline{\Lambda}) \cap \Pi} L_n (\Lambda_0, \Lambda_*, \theta_0)$$

$$\geq \mathbb{E} \left(\inf_{\Lambda \in \mathcal{V}_k(\overline{\Lambda}) \cap \Pi} l_t (\Lambda, \Lambda_*, \theta_0) \right).$$

By the Beppo-Levi theorem, $\mathbb{E} \left(\inf_{\Lambda \in \mathcal{V}_k(\overline{\Lambda}) \cap \Pi} l_t (\Lambda, \Lambda_*, \theta_0) \right)$ converges while increasing to $\mathbb{E} \left(l_t \left(\overline{\Lambda}, \Lambda_*, \theta_0 \right) \right)$ as $k \to \infty$. Hence, Lemma 9 follows from Lemma 8 ii). $\quad \blacksquare$

Next, it follows the proof of the Theorem 6, based on the previous lemmas.

Proof. (Theorem 6)

a) The proof of (38)-i is based on the results in Lemma (7) and (9). It is shown that for all $\overline{\Lambda} \neq \Lambda_0$, there exists a neighborhood $\mathcal{V}(\overline{\Lambda})$ such that

$$\liminf_{n\to\infty} \inf_{\Lambda \in \mathcal{V}_k(\overline{\Lambda}) \cap \Pi} L_n \left(\Lambda, \Lambda_*, \widehat{\theta}_{1n} \right) \geq \liminf_{n\to\infty} L_n \left(\Lambda_0, \Lambda_*, \widehat{\theta}_{1n} \right)$$

$$= \liminf_{n\to\infty} L_n (\Lambda_0, \Lambda_*, \theta_0) = \mathbb{E} \left(l_t (\Lambda_0, \Lambda_*, \theta_0) \right),$$

44

which concludes the proof.

b) The proof of (38)-ii is similar to that of (38)-i by replacing $\widehat{\theta}_{1n}$ by $\widehat{\theta}_{2n}$. Similarly, the proof of (38)-iii and (38)-iv is similar to that of (38)-i by replacing Λ_* by $\widehat{\Lambda}_{1n}$ and Λ_* by $\widehat{\Lambda}_{2n}$, respectively.

c) The proof of the results in (39) is similar to all estimators and thus only the consistency for $\widehat{\Lambda}_{1n}$ is shown.

d) From standard arguments, the proof of the consistency of $\widehat{\Lambda}_{1n}$ is completed while using the compactness assumption **A1** of Π. Using again a Taylor expansion of

$$\frac{\left(\left(Y_t - \mathcal{Y}_{t-1}'\widehat{\theta}_{1n}\right)^2 - V_t\left(\widehat{\theta}_{1n}, \Lambda\right)\right)^2}{\left(V_t\left(\widehat{\theta}_{1n}, \Lambda_*\right)\right)^2},$$

around θ_0, it holds that

$$\sqrt{n}\sup_{\theta \in \Theta}\left\|L_n\left(\Lambda, \Lambda_*, \widehat{\theta}_{1n}\right) - L_n\left(\Lambda, \Lambda_*, \theta_0\right)\right\|$$

$$\leq \quad \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left|\frac{\left(\left(Y_t - \mathcal{Y}_{t-1}'\widehat{\theta}_{1n}\right)^2 - V_t\left(\widehat{\theta}_{1n}, \Lambda\right)\right)^2}{\left(V_t\left(\widehat{\theta}_{1n}, \Lambda_*\right)\right)^2} - \frac{\left(\left(Y_t - \mathcal{Y}_{t-1}'\theta_0\right)^2 - V_t(\theta_0, \Lambda)\right)^2}{(V_t(\theta_0, \Lambda_*))^2}\right| = o_{a.s.}(1).$$

Now from **A2** and (38)-i, the estimate $\widehat{\Lambda}_{1n}$ cannot be on the boundary of Π for a sufficiently large n. Therefore, a Taylor expansion of $\sqrt{n}\frac{\partial L_n\left(\widehat{\Lambda}_{1n}, \Lambda_*, \widehat{\theta}_{1n}\right)}{\partial \Lambda}$ around Λ_0 implies that for some $\Lambda_u \in \Pi$ between $\widehat{\Lambda}_{1n}$ and Λ_0,

$$\begin{aligned} 0 &= \sqrt{n}\frac{\partial L_n\left(\widehat{\Lambda}_{1n}, \Lambda_*, \widehat{\theta}_{1n}\right)}{\partial \Lambda} = \sqrt{n}\frac{\partial L_n\left(\widehat{\Lambda}_{1n}, \Lambda_*, \theta_0\right)}{\partial \Lambda} + o_{a.s.}(1) \\ &= \sqrt{n}\frac{L_n(\Lambda_0, \Lambda_*, \theta_0)}{\partial \Lambda} + \sqrt{n}\frac{L_n(\Lambda_u, \Lambda_*, \theta_0)}{\partial \Lambda}\left(\widehat{\Lambda}_{1n} - \Lambda_0\right) + o_{a.s.}(1). \end{aligned} \tag{60}$$

Finally, given (60) the result is established while the following two lemmas are shown. ∎

Lemma 10 *Under **A0****, **A1**, **A2** and the finiteness of the fourth moments of the random coefficients, then*

$$\sqrt{n}\frac{L_n(\Lambda_0, \Lambda_*, \theta_0)}{\partial \Lambda} \xrightarrow[n \to \infty]{\mathcal{L}} \mathcal{N}(0, D(\Lambda_*)).$$

Proof. (Lemma 10)

Clearly $\left\{ \frac{\partial l_t(\Lambda_0, \Lambda_*, \theta_0)}{\partial \theta}, t \in \mathbb{Z} \right\}$ is a martingale difference with respect to $\left\{ \mathcal{F}_t^Y, t \in \mathbb{Z} \right\}$ where

$$\frac{\partial l_t(\Lambda, \Lambda_*, \theta_0)}{\partial \theta} = -2 \frac{\partial V_t(\theta_0, \Lambda)}{\partial \Lambda} \frac{e_t^2 - V_t(\theta_0, \Lambda)}{u_t^2(\theta_0, \Lambda_*)} \text{ and } \sqrt{n} \frac{L_n(\Lambda_0, \Lambda_*, \theta_0)}{\partial \Lambda} = \sum_{t=1}^n \frac{1}{\sqrt{n}} \frac{\partial l_t(\Lambda_0, \Lambda_*, \theta_0)}{\partial \Lambda}.$$

By **A1** and **A2**, it holds that $\mathbb{E}\left(\frac{\partial l_t(\Lambda_0, \Lambda_*, \theta_0)}{\partial \Lambda} \frac{\partial l_t(\Lambda_0, \Lambda_*, \theta_0)}{\partial \Lambda'} \right) = D(\Lambda_*)$, and the lemma follows from the martingale CLT. ∎

Lemma 11 *Under* **A0****, **A1** *and* **A2** *it follows that* $\frac{\partial^2 L_n(\Lambda_u, \Lambda_*, \theta_0)}{\partial \Lambda \partial \Lambda'} \overset{a.s.}{\underset{n \to \infty}{\rightarrow}} C(\Lambda_*).$

Proof. (Lemma 11)

Let $\mathcal{V}_k(\Lambda_0)$ $(k \in \mathbb{N}^*)$ be the open ball of center Λ_0 and radius $1/k$. Assume that n is large enough so that Λ_u belongs to $\mathcal{V}_k(\Lambda_0)$. From the stationarity and ergodicity of

$$\left\{ \inf_{\Lambda \in \mathcal{V}_k(\Lambda_0)} \left| \frac{\partial^2 l_t(\Lambda_0, \Lambda_*, \theta_0)}{\partial \Lambda_i \partial \Lambda_j} - \mathbb{E}\left(\frac{\partial^2 l_t(\Lambda_0, \Lambda_*, \theta_0)}{\partial \Lambda_i \partial \Lambda_j} \right) \right|, t \in \mathbb{Z} \right\},$$

if follows that

$$\left| \frac{\partial^2 L_n(\Lambda_u, \Lambda_*, \theta_0)}{\partial \Lambda \partial \Lambda'} - C(\Lambda_*) \right| \leq \frac{1}{n} \sum_{t=1}^n \inf_{\Lambda \in \mathcal{V}(\Lambda_0)} \left| \frac{\partial^2 l_t(\Lambda, \Lambda_*, \theta_0)}{\partial \Lambda_i \partial \Lambda_j} - \mathbb{E}\left(\frac{\partial^2 l_t(\Lambda_0, \Lambda_*, \theta_0)}{\partial \Lambda_i \partial \Lambda_j} \right) \right|$$

$$\overset{a.s.}{\underset{n \to \infty}{\rightarrow}} \mathbb{E}\left(\inf_{\Lambda \in \mathcal{V}_k(\Lambda_0)} \left| \frac{\partial^2 l_t(\Lambda, \Lambda_*, \theta_0)}{\partial \Lambda_i \partial \Lambda_j} - \mathbb{E}\left(\frac{\partial^2 l_t(\Lambda_0, \Lambda_*, \theta_0)}{\partial \Lambda_i \partial \Lambda_j} \right) \right| \right).$$

Note that the Lebesgue-dominated convergence theorem entails

$$\lim_{k \to \infty} \mathbb{E}\left(\inf_{\Lambda \in \mathcal{V}_k(\Lambda_0)} \left| \frac{\partial^2 l_t(\Lambda, \Lambda_*, \theta_0)}{\partial \Lambda_i \partial \Lambda_j} - \mathbb{E}\left(\frac{\partial^2 l_t(\Lambda_0, \Lambda_*, \theta_0)}{\partial \Lambda_i \partial \Lambda_j} \right) \right| \right)$$

$$= \mathbb{E}\left(\lim_{k \to \infty} \sup_{\Lambda \in \mathcal{V}_k(\Lambda_0)} \left| \frac{\partial^2 l_t(\Lambda, \Lambda_*, \theta_0)}{\partial \Lambda_i \partial \Lambda_j} - \mathbb{E}\left(\frac{\partial^2 l_t(\Lambda_0, \Lambda_*, \theta_0)}{\partial \Lambda_i \partial \Lambda_j} \right) \right| \right) = 0,$$

which completes the proof of the result. ∎

Supplementary Material

(Table 11 here)

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. (2013). Two-stage weighted least squares estimation of nonstationary random coefficient autoregressions. *Journal of Time Series Econometrics* **5**, 25–46.

[4] Aknouche, A. (2015). Quadratic random coefficient autoregression with linear-in-parameters volatility. *Statistical Inference for Stochastic Processes* **18**, 99–125.

[5] 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.

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

[7] Aknouche, A. and Francq, C. (2023). Two-stage weighted least squares estimator of the conditional mean of observation-driven time series models. *Journal of Econometrics*, to appear.

[8] Aknouche, A. and Scotto, M.G. (2024). A multiplicative thinning-based integer-valued GARCH model. *Journal of Time Series Analysis* **45**, 4–26.

[9] Alomani, G., Alzaid, A.A. and Omair, M.A. (2018). A Skellam INGARCH model. *Brazilian Journal of Probability and Statistics* **32**, 200–214.

[10] 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.

[11] Alzaid, A.A. and Omair, M.A. (2010). On the Poisson difference distribution inference and applications. *Bulletin of the Malaysian Mathematical Sciences Society* **33**, 17–45.

[12] Alzaid, A.A. and Omair, M.A. (2014). Poisson difference integer valued autoregressive model of order one. *Bulletin of the Malaysian Mathematical Sciences Society* **37**, 465–485.

[13] Aue, A., Horváth, L. and Steinebach, J. (2006). Estimation in random coefficient autoregressive models. *Journal of Time Series Analysis* **27**, 61–76.

[14] Aue, A. and Horváth, L. (2011). Quasi-likelihood estimation in stationary and nonstationary autoregressive models with random coefficients. *Statistica Sinica* **21**, 973–99.

[15] Barndorff-Nielsen, O.E., Pollard, D.G. and Shephard, N. (2012). Integer-valued Levy procces and low latency financial econometrics. *Quantitative Finance* **12**, 587–605.

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

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

[18] Brandt, A. (1986). The stochastic equation $X_{n+1} = A_n X_n + B_n$ with stationary coefficients. *Advances in Applied Probability* **18**, 211–220.

[19] Cox, D.R. (1981). Statistical analysis of time series: Some recent developments. *Scandinavian Journal of Statistics* **8**, 93–115.

[20] Cui, Y., Li, Q. and Zhu, F. (2021). Modeling $\mathbb{Z}$-valued time series based on new versions of the Skellam INGARCH model *Brazilian Journal of Probability and Statistics* **35**, 93–314.

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

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

[23] 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.

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

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

[26] Diaconis, P. and Freedman, D. (1999). Iterated random functions. *SIAM Review* **41**, 45–76.

[27] Du, J.G. and Li, Y. (1991). The integer-valued autoregressive (INAR(p)) model. *Journal of Time Series Analysis* **12**, 129–142.

[28] Feigin, P.D. and Tweedie, R.L. (1985). Random coefficient autoregressive processes: a Markov chain analysis of stationarity and finiteness of moments. *Journal of Time Series Analysis* **6**, 1–14.

[29] Ferland, R., Latour, A. and Oraichi, D. (2006). Integer-valued GARCH process. *Journal of Time Series Analysis* **7**, 923–942.

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

[31] Francq, C. and Zakoian, J.M. (2019). *GARCH Models, Structure, Statistical Inference and Financial Applications.* John Wiley & Sons, Ltd.

[32] Furstenberg, H. and Kesten, H. (1960). Products of random matrices. *Annals of Mathematical Statistics* **1**, 457–469.

[33] Goldie, C.M. (1991). Implicit renewal theory and tails of solutions of random equations. *Annals of Applied Probability* **1**, 126–166.

[34] 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.

[35] Grey, D.R. (1994). Regular variation in the tail behaviour of solutions of random difference equations. *Annals of Applied Probability* **4**, 169–183.

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

[37] Irwin, J.O. (1937). The frequency distribution of the difference between two independent variates following the same Poisson distribution. *Journal of the Royal Statistical Society: Series A* **100**, 415–416.

[38] Kesten, H. (1973). Random difference equations and renewal theory for products of random. *Acta Mathematica* **131**, 207–248.

[39] Kim, H.-Y. and Park, Y. (2008). A non-stationary integer-valued autoregressive model. *Statistical Papers* **49**, 485–502.

[40] Koopman, S., Lit, R. and Lucas, A. (2017). Intraday stochastic volatility in discrete price changes: the dynamic Skellam model. *Journal of the American Statistical Association* **112**, 1490–1503.

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

[42] Ling, S. (1999). On the stationarity and the existence of moments of conditional heteroskedastic $ARMA$ models. *Statistica Sinica* **9**, 1119–1130.

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

[44] Nicholls, F. and Quinn, B.G. (1982). *Random Coefficient Autoregressive Models*: *An Introduction*. Springer-Verlag, New York.

[45] Praskova, Z. and Vanecek, P. (2011). On a class of estimators in multivariate RCA(1) model. *Kybernetika* **47**, 501–518.

[46] Regis, M., Serra, P. and van den Heuvel, E.R. (2022). Random autoregressive models: A structured overview. *Econometric Reviews* **41**, 207–230.

[47] Rydberg, T.H. and Shephard, N. (2000). A modelling framework for the prices and times of trades made on the New York stock exchange. *Nonlinear and Nonstationary Signal Processing*, 217–246.

[48] Schick, A. (1996). Consistent estimation in a random coefficient autoregressive model. *Australian Journal of Statistics* **38**, 155–160.

[49] 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.

[50] Skellam, J.G. (1946). The frequency distribution of the difference between two Poisson variates belonging to different populations. *Journal of the Royal Statistical Society, Series A* **109**, 296.

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

[52] Trapani, L. (2021). Testing for strict stationarity in a random coefficient autoregressive model. *Econometric Reviews* **40**, 220–256.

[53] Tsay, R.S. (1987). Conditional heteroskedastic time series models. *Journal of the American Statistical Association* **7**, 590–604.

[54] Vervaat, W. (1979). On a stochastic difference equation and a representation of non negative infinitely divisible random variables. *Advances in Applied Probability* **11**, 750–783.

[55] Weiss, C.H. (2018). *An Introduction to Discrete-valued Time Series.* Wiley, Oxford, UK.

[56] Xu, Y. and Zhu, F. (2022). A new asymmetric power ARCH model for Z-valued time series. *Journal of Time Series Analysis* **43**, 490–500.

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

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

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

[60] Zheng, H.T., Basawa, I.V. and Datta, S. (2007). The first order random coefficient integer-valued autoregressive processes. *Journal of Statistical Planning and Inference* **173**, 212–229.

[61] Zhang, H., Wang, D. and Zhu, F. (2009). Inference for inar(p) processes with signed generalized power series thinning operator. *Journal of Statistical Planning and Inference* **140**, 667–683.

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

$n = 1000$	(a) $\sum_{i=1}^{4} \phi_{0i} = 0.7 < 1$				
$\widehat{\theta}_{2n}$	$\mu_\varepsilon = 2$	$\phi_{01} = 0.3$	$\phi_{02} = 0.2$	$\phi_{03} = 0.1$	$\phi_{04} = 0.1$
Mean	2.1249	0.2913	0.1911	0.0946	0.0929
StD	0.3399	0.0562	0.0514	0.0475	0.0462
ASE	0.2540	0.0526	0.0504	0.0449	0.0436
$\widehat{\Lambda}_{2n}$	$\sigma_\varepsilon^2 = 2$	$\sigma_{\phi_1}^2 = 0.3$	$\sigma_{\phi_2}^2 = 0.2$	$\sigma_{\phi_3}^2 = 0.1$	$\sigma_{\phi_4}^2 = 0.1$
Mean	1.9826	0.2829	0.1863	0.0901	0.0891
StD	0.1321	0.0056	0.0078	0.0074	0.0076
ASE	0.0318	0.0019	0.0022	0.0020	0.0021
$n = 1000$	(b) $\sum_{i=1}^{4} \phi_{0i} = 1$				
$\widehat{\theta}_{2n}$	$\mu_\varepsilon = 1$	$\phi_{01} = 0.4$	$\phi_{02} = 0.3$	$\phi_{03} = 0.1$	$\phi_{04} = 0.2$
Mean	1.1151	0.3907	0.2932	0.0936	0.1963
StD	0.2513	0.0574	0.0605	0.0462	0.0574
ASE	0.1567	0.0537	0.0539	0.0447	0.0497
$\widehat{\Lambda}_{2n}$	$\sigma_\varepsilon^2 = 1$	$\sigma_{\phi_1}^2 = 0.4$	$\sigma_{\phi_2}^2 = 0.3$	$\sigma_{\phi_3}^2 = 0.1$	$\sigma_{\phi_4}^2 = 0.2$
Mean	0.9455	0.3769	0.2818	0.0900	0.1823
StD	0.0841	0.0083	0.0082	0.0076	0.0122
ASE	0.0130	0.0026	0.0028	0.0026	0.0032
$n = 1000$	(c) $\sum_{i=1}^{4} \phi_{0i} = 1.2 > 1$				
$\widehat{\theta}_{2n}$	$\mu_\varepsilon = 0.1$	$\phi_{01} = 0.5$	$\phi_{02} = 0.2$	$\phi_{03} = 0.3$	$\phi_{04} = 0.2$
Mean	0.1028	0.5087	0.2073	0.2892	0.2008
StD	0.0190	0.0662	0.0540	0.0671	0.0592
ASE	0.0188	0.0635	0.0573	0.0639	0.0583
$\widehat{\Lambda}_{2n}$	$\sigma_\varepsilon^2 = 0.1$	$\sigma_{\phi_1}^2 = 0.5$	$\sigma_{\phi_2}^2 = 0.2$	$\sigma_{\phi_3}^2 = 0.3$	$\sigma_{\phi_4}^2 = 0.2$
Mean	0.0551	0.5048	0.1956	0.3147	0.1703
StD	0.0141	0.0241	0.0295	0.0322	0.0430
ASE	0.0045	0.0043	0.0046	0.0046	0.0054

Table 1: Mean, StD and ASE of $\widehat{\theta}_{2n}$ and $\widehat{\Lambda}_{2n}$ for a Poisson RMINAR(4) under (a) strict stationarity with finite mean, (b)-(c) strict stationarity with infinite mean.

$n = 1000$	(a) $\sum_{i=1}^{3} \phi_{0i} = 0.6 < 1$			
$\widehat{\theta}_{2n}$	$\mu_{\varepsilon} = 2$	$\phi_{01} = 0.3$	$\phi_{02} = 0.2$	$\phi_{03} = 0.1$
Mean	2.0834	0.2952	0.1929	0.0866
StD	0.2228	0.0518	0.0467	0.0443
ASE	0.2120	0.0538	0.0519	0.0480
$\widehat{\Lambda}_{2n}$	$\sigma_{\varepsilon}^2 = 2$	$\sigma_{\phi_1}^2 = 0.3$	$\sigma_{\phi_2}^2 = 0.2$	$\sigma_{\phi_3}^2 = 0.1$
Mean	2.0834	0.2952	0.1929	0.0866
StD	0.2228	0.0518	0.0467	0.0443
ASE	0.0390	0.0024	0.0037	0.0038
$n = 1000$	(b) $\sum_{i=1}^{3} \phi_{0i} = 1.1 > 1$			
$\widehat{\theta}_{2n}$	$\mu_{\varepsilon} = 0.5$	$\phi_{01} = 0.3$	$\phi_{02} = 0.2$	$\phi_{03} = 0.6$
Mean	0.5261	0.2907	0.2019	0.6127
StD	0.0859	0.0501	0.0492	0.0757
ASE	0.0711	0.0491	0.0481	0.0682
$\widehat{\Lambda}_{2n}$	$\sigma_{\varepsilon}^2 = 0.45$	$\sigma_{\phi_1}^2 = 0.3$	$\sigma_{\phi_2}^2 = 0.2133$	$\sigma_{\phi_3}^2 = 0.9$
Mean	0.4216	0.2975	0.2075	0.8947
StD	0.0425	0.0113	0.0137	0.0227
ASE	0.0100	0.0026	0.0038	0.0063

Table 2: Mean, StD and ASE of $\widehat{\theta}_{2n}$ and $\widehat{\Lambda}_{2n}$ for a RMINAR(3) with $\varepsilon_t \sim \mathcal{B}in\left(5, \frac{\mu_{\varepsilon}}{5}\right), \Phi_{1t} \sim \mathcal{P}\left(\phi_{01}\right), \Phi_{2t} \sim \mathcal{NB}\left(3, \frac{3}{3+\phi_{02}}\right)$ and $\Phi_{3t} \sim \mathcal{NB}\left(2\phi_{03}, \frac{2}{3}\right)$ in both finite and infinite mean cases.

		$\mu_{1\varepsilon} = 0.7,\ \mu_{2\varepsilon} = 0.3,\ \phi_{11} = 0.1,\ \phi_{21} = 0.3$		
$n = 1000$	(a)	$\phi_{12} = 0.2,\ \phi_{22} = 0.1,\ \phi_{13} = 0.4,\ \phi_{23} = 0.2$		
		$\rho(\Phi) = 0.5761 < 1$		
$\widehat{\theta}_{2n}$	$\mu_\varepsilon = 0.4$	$\phi_{01} = -0.2$	$\phi_{02} = 0.1$	$\phi_{03} = 0.2$
Mean	0.4020	-0.1954	0.0958	0.1978
StD	0.0646	0.0394	0.0365	0.0449
ASE	0.0658	0.0408	0.0399	0.0457
$\widehat{\Lambda}_{2n}$	$\sigma^2_\varepsilon = 1$	$\sigma^2_{\phi_1} = 0.4$	$\sigma^2_{\phi_2} = 0.3$	$\sigma^2_{\phi_3} = 0.6$
Mean	0.9904	0.3969	0.2945	0.5980
StD	0.0417	0.0116	0.0153	0.0117
ASE	0.0079	0.0036	0.0041	0.0038
		$\mu_{1\varepsilon} = 0.3 = \mu_{2\varepsilon},\ \phi_{11} = 0.5,\ \phi_{21} = 0.1$		
	(b)	$\phi_{12} = 0.4,\ \phi_{22} = 0.1,\ \phi_{13} = 0.5,\ \phi_{23} = 0.2$		
		$\rho(\Phi) = 1$		
$\widehat{\theta}_{2n}$	$\mu_\varepsilon = 0$	$\phi_{01} = 0.4$	$\phi_{02} = 0.3$	$\phi_{03} = 0.3$
Mean	-0.0537	0.3917	0.2803	0.2767
StD	0.1403	0.0611	0.0709	0.0675
ASE	0.0681	0.0450	0.0466	0.0516
$\widehat{\Lambda}_{2n}$	$\sigma^2_\varepsilon = 0.6$	$\sigma^2_{\phi_1} = 0.6$	$\sigma^2_{\phi_2} = 0.5$	$\sigma^2_{\phi_3} = 0.7$
Mean	0.5809	0.5900	0.4939	0.6780
StD	0.0687	0.0183	0.0143	0.0297
ASE	0.0104	0.0055	0.0033	0.0050

Table 3: Mean, StD and ASE of $\widehat{\theta}_{2n}$ and $\widehat{\Lambda}_{2n}$ for a $\mathbb{Z}$-valued RMINAR(3) with $\varepsilon_t \sim$ $Skellam(\mu_{1\varepsilon}, \mu_{2\varepsilon})$, $\Phi_{it} \sim Skellam(\phi_{1i}, \phi_{2i})$, $\mu_{1\varepsilon} = \mu_{1\varepsilon} - \mu_{2\varepsilon}, \sigma^2_\varepsilon = \mu_{1\varepsilon} + \mu_{2\varepsilon}, \phi_{0i} = \phi_{1i} - \phi_{2i}$, and $\sigma^2_{\phi_i} = \phi_{1i} + \phi_{2i}$.

$n = 1000$	(a) $\sum_{i=1}^{4} \phi_{0i} = 0.7 < 1$			
$(\widehat{\theta}_{2n}, \widehat{\sigma}_{\varepsilon n}^2)$	$\omega_0 = 1$	$\phi_{01} = 0.4$	$\phi_{02} = 0.3$	$\sigma_\varepsilon^2 = 1$
Mean	0.9855	0.3981	0.3039	1.0042
StD	0.1521	0.0659	0.0620	0.1222
ASE	0.1553	0.0683	0.0638	0.0287
$n = 1000$	(b) $\sum_{i=1}^{4} \phi_{0i} = 1$			
$(\widehat{\theta}_{2n}, \widehat{\sigma}_{\varepsilon n}^2)$	$\omega_0 = 1$	$\phi_{01} = 0.5$	$\phi_{02} = 0.5$	$\sigma_\varepsilon^2 = 1$
Mean	1.0168	0.5008	0.5031	1.0064
StD	0.1609	0.0692	0.0814	0.1207
ASE	0.1669	0.0738	0.0768	0.0273
$n = 1000$	(c) $\sum_{i=1}^{4} \phi_{0i} = 1.1 > 1$			
$(\widehat{\theta}_{2n}, \widehat{\sigma}_{\varepsilon n}^2)$	$\omega_0 = 1$	$\phi_{01} = 0.5$	$\phi_{02} = 0.6$	$\sigma_\varepsilon^2 = 1$
Mean	1.0075	0.5013	0.6069	1.0472
StD	0.1579	0.0820	0.0817	0.1458
ASE	0.1684	0.0727	0.0826	0.0304

Table 4: Mean, StD and ASE of $\widehat{\theta}_{2n}$ and $\widehat{\sigma}_{\varepsilon n}^2$ for a multiplicative RMINAR(4) with Poisson inputs under (a) strict stationarity with finite mean, and (b)-(c) strict stationarity with infinite mean.

A	p	$\mathbb{E}(Y)$	$\mathbb{V}(Y)$	MAR	MSR	MSPR
	1	13.2026	223.1584	9.3449	217.5623	1.0000
	2	13.2793	254.0780	9.3048	213.7554	0.9996
	3	13.2255	303.9467	9.2035	211.0007	1.0005
M	p	$\mathbb{E}(Y)$	-	MAR	MSR	MSPR
	1	12.0650	-	9.3569	217.5965	0.9958
	2	11.8887	-	9.3381	214.0984	1.1203
	3	11.6881	-	9.2736	212.7709	1.2195

Table 5: Selecting the best order p, considering additive (A) or multiplicative (M) RMINAR(p) models on the ExRate time series.

A	$p = 1$	$p = 2$	$p = 3$	M	$p = 1$	$p = 2$	$p = 3$
$\widehat{\mu}_\varepsilon$	11.1727 (0.2540)	9.3941 (0.3672)	8.2255 (0.4674)	$\widehat{\mu}_\varepsilon$	1.0029	1.0035	1.0082
$\widehat{\phi}_1$	0.1537 (0.0298)	0.1277 (0.0343)	0.1346 (0.0382)	$\widehat{\omega}$	9.9797 (0.8275)	7.9889 (0.8404)	6.9170 (0.8148)
$\widehat{\phi}_2$	-	0.1649 (0.0356)	0.1520 (0.0393)	$\widehat{\alpha}_1$	0.1728 (0.0457)	0.1610 (0.0495)	0.1854 (0.0544)
$\widehat{\phi}_3$	-	-	0.0914 (0.0356)	$\widehat{\alpha}_2$	-	0.1670 (0.0524)	0.1441 (0.0516)
$\widehat{\sigma}^2_\varepsilon$	212.4713 (8.5459)	157.1334 (8.0192)	112.2556 (7.4415)	$\widehat{\alpha}_3$	-	-	0.0787 (0.0557)
$\widehat{\sigma}^2_{\phi_1}$	0.0136 (0.0099)	0.0103 (0.0084)	0.0183 (0.0108)	$\widehat{\sigma}^2_\varepsilon$	1.0143 (0.0733)	0.7118 (0.0508)	0.5585 (0.0390)
$\widehat{\sigma}^2_{\phi_2}$	-	0.1854 (0.0290)	0.1519 (0.0266)	-	-	-	-
$\widehat{\sigma}^2_{\phi_3}$	-	-	0.1882 (0.0293)	-	-	-	-
SSC	0.0372	0.3701	0.5238	-	-	-	-

Table 6: 4SWLSE estimates and asymptotic standard errors (in parenthesis) for the additive (A) or multiplicative (M) RMINAR(p) models on the ExRate time series. The SSC stands for second-order stationary condition.

	n_c	300	420	520	680	700
	MSFE	241.7308	295.1503	398.7492	309.7075	366.5074
A RMINAR(3)	MAFE	9.4727	10.2364	11.7017	13.3228	13.1845
	MSPFE	1.3535	1.7079	2.3473	1.2327	1.4226
	MSFE	241.1297	288.3198	396.5317	313.0352	369.2506
M RMINAR(3)	MAFE	9.4708	10.0013	11.6276	13.3125	13.1978
	MSPFE	1.6006	1.8979	2.3258	2.1731	2.1345
	MSFE	239.611	291.263	395.830	305.882	363.530
MthINGARCH	MAFE	9.4884	10.169	11.769	13.340	13.319
	MSPFE	1.1744	1.5721	2.2917	1.2246	1.6525

Table 7: Out-of-sample forecasting ability of the additive (A) RMINAR(3), the multiplicative (M) RMINAR(3), and the MthINGARCH(1, 1) on the ExRate time series. The n_c stands for the number of observations used for parameter estimation. The MthINGARCH(1, 1) results are obtained from Aknouche and Scotto (2024).

p	$\mathbb{E}(Y)$	$\mathbb{V}(Y)$	MAR	MSR	MSPR
1	0.1596	999.885	21.9768	976.2304	0.9982
2	−0.0584	962.0471	21.8832	971.7091	0.9961
3	0.0484	904.8307	21.6772	951.4233	1.0972

Table 8: Selecting the best order p for the $\mathbb{Z}$-valued RMINAR(p) model on the Return time series.

	$p = 1$	$p = 2$	$p = 3$
$\widehat{\mu}_{\varepsilon}$	-0.03067 (0.0733)	-0.0610 (0.0892)	0.0450 (0.1161)
$\widehat{\phi}_1$	0.1645 (0.0075)	-0.0120 (0.0074)	-0.0092 (0.0091)
$\widehat{\phi}_2$	-	-0.0322 (0.0097)	-0.0229 (0.0127)
$\widehat{\phi}_3$	-	-	0.1021 (0.0092)
$\widehat{\sigma}^2_{\varepsilon}$	783.9099 (41.9269)	671.1159 (41.9794)	535.3662 (38.0848)
$\widehat{\sigma}^2_{\phi_1}$	0.2151 (0.0468)	0.1781 (0.0422)	0.2484 (0.0461)
$\widehat{\sigma}^2_{\phi_2}$	-	0.1231 (0.0363)	0.1488 (0.0368)
$\widehat{\sigma}^2_{\phi_3}$	-	-	0.0000 (0.0158)
SSC	0.1741	0.2799	0.4004

Table 9: 4SWLSE results for the $\mathbb{Z}$-valued RMINAR(p) model on the Return time series. Same caption as in Table 6.

p	n_c	300	350	400	450	500	550
	MSFE	1404.3517	1581.2839	1854.6241	1042.7839	1106.3495	1203.23862
1	MAFE	26.0506	27.5554	30.1032	24.2280	24.5137	27.60783
	MSPFE	2.2259	2.5298	2.9709	1.2397	1.2872	1.37160
	MSFE	1413.7965	1582.6909	1843.1118	1060.6812	1122.3612	1196.0980
2	MAFE	26.1138	27.5715	30.0951	24.6735	24.8328	27.6674
	MSPFE	1.9150	2.5863	2.7693	1.3997	1.4153	1.5050

Table 10: Out-of-sample forecasting ability of the $\mathbb{Z}$-valued RMNINAR(p) model on the Return time series.

	n_c	300	420	520	680	700
	MSFE	246.9545	318.5958	413.4456	322.4681	377.6246
A RMINAR(1)	MAFE	9.4164	10.4993	11.8362	13.5432	13.4922
	MSPFE	1.360	1.9864	2.7857	1.5346	1.8009
	MSFE	243.8918	314.6527	406.6413	316.6054	381.6725
A RMINAR(2)	MAFE	9.4024	10.4726	11.7646	13.2168	13.1774
	MSPFE	1.4137	2.0281	2.7375	1.7056	2.2566
	MSFE	241.7308	295.1503	398.7492	309.7075	366.5074
A RMINAR(3)	MAFE	9.4727	10.2364	11.7017	13.3228	13.1845
	MSPFE	1.3535	1.7079	2.3473	1.2327	1.4226

Table 11: Out-of-sample forecasting ability of the additive (A) RMINAR(p) models on the ExRate time series.