entropy

Article

A Modified Multiplicative Thinning-Based INARCH Model: Properties, Saddlepoint Maximum Likelihood Estimation, and Application

Yue Xu [1], Qi Li [2] and Fukang Zhu [1,*]

[1] School of Mathematics, Jilin University, Changchun 130012, China
[2] College of Mathematics, Changchun Normal University, Changchun 130032, China
* Correspondence: fzhu@jlu.edu.cn

Abstract: In this article, we propose a modified multiplicative thinning-based integer-valued autoregressive conditional heteroscedasticity model and use the saddlepoint maximum likelihood estimation (SPMLE) method to estimate parameters. A simulation study is given to show a better performance of the SPMLE. The application of the real data, which is concerned with the number of tick changes by the minute of the euro to the British pound exchange rate, shows the superiority of our modified model and the SPMLE.

Keywords: INARCH model; saddlepoint approximation; thinning-based model; time series of counts

check for updates

Citation: Xu, Y.; Li, Q.; Zhu, F. A Modified Multiplicative Thinning-Based INARCH Model: Properties, Saddlepoint Maximum Likelihood Estimation, and Application. *Entropy* **2023**, *25*, 207. https://doi.org/10.3390/e25020207

Academic Editor: Christian H. Weiss

Received: 19 December 2022
Revised: 15 January 2023
Accepted: 18 January 2023
Published: 21 January 2023

1. Introduction

In practice, we can often observe a series of integer-valued data that have their own distinguishing characteristics, and many models were proposed for modeling integer-valued time series, such as the integer-valued autoregressive (INAR) process introduced by McKenzie (1985) [1], and Al-Osh and Alzaid (1987) [2]; the integer-valued moving average process proposed by Al-Osh and Alzaid (1988) [3]; the integer-valued autoregressive moving-average model defined by McKenize (1988) [4]; and the integer-valued generalized autoregressive conditional heteroscedasticity (INGARCH) model proposed by Ferland et al. (2006) [5], among others. Here we focus on two kinds of the models above: one is the INAR process, which was introduced as a convenient way to transfer the usual autoregressive structure to a discrete-valued time series, and a p-order model, which is defined as follows:

$$X_t = \sum_{i=1}^{p} \alpha_i \circ X_{t-i} + \varepsilon_t,$$

where $\alpha_i \in [0,1)$ for $i = 1,\ldots,p$, and $\{\varepsilon_t\}$ is a sequence of independent and identically distributed (i.i.d.) non-negative integer-valued random variables with $E(\varepsilon_t) = \mu$ and $\mathrm{Var}(\varepsilon_t) = \sigma_\varepsilon^2$. The binomial thinning operator $\circ$ is defined by Steutel and Van Harn (1979) [6] as:

$$\alpha \circ X = \sum_{i=1}^{X} Y_i, \text{ if } X > 0 \text{ and } 0 \text{ otherwise},$$

where Y_i are i.i.d. Bernoulli random variables, independent of X, with a success probability are defined by α. This model has been generalized by Qian and Zhu (2022) [7], and Huang et al. (2023) [8], among others.

The other is the INGARCH model which was proposed by Ferland et al. (2006) [5] to model the observations of integer-valued time series which exist heteroscedasticity; this INGARCH(p, q) model with a Poisson deviate is defined as:

$$X_t|\mathscr{F}_{t-1} : P(\lambda_t), \quad \lambda_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i} + \sum_{j-1}^{q} \beta_j \lambda_{t-j},$$

where $\alpha_0 > 0, \alpha_i \geq 0, \beta_j \geq 0, i = 1, \dots, p, j = 1, \dots, q, p \geq 1, q \geq 0$, and $\mathscr{F}_{t-1}$ is the σ-field generated by $\{X_{t-1}, X_{t-2}, \dots\}$. This model has been generalized by Hu (2016) [9], Liu et al. (2022) [10], and Weiß et al. (2022) [11], among others. Weiß (2018) [12] and Davis et al. (2021) [13] gave recent reviews. According to definitions of INAR and INGARCH models, we noticed that the INAR model is thinning-based, while the INGARCH model is specified by a conditional distribution with a time-varying mean depending on past observations. Combining the thinning-based stochastic equations and the INGARCH model, Aknouche and Scotto (2022) [14] proposed a multiplicative thinning-based INGARCH (MthINGARCH) model to model the integer-valued time series with high overdispersion and persistence. Furthermore, it fits well with heavy-tailed data regardless of the choice of innovation distribution and does not require recourse to complex random coefficient equations. The MthINGARCH model is denoted by:

$$\begin{cases} X_t = \lambda_t \varepsilon_t, \\ \lambda_t = 1 + \omega \circ m + \sum\limits_{i=1}^{q} \alpha_i \circ X_{t-i} + \sum\limits_{j=1}^{p} \beta_j \circ \lambda_{t-j}, \end{cases} \tag{1}$$

where the symbol $\circ$ stands for the binomial thinning operator, and $0 \leq \omega \leq 1, 0 \leq \alpha_i < 1$ and $0 \leq \beta_j < 1$ $(i = 1, \dots, q, j = 1, \dots, p)$, m is a fixed positive integer number that was introduced for more flexibility. Since there is no explicit probability mass function for the series $\{X_t\}$, then the traditional maximum likelihood estimation (MLE) cannot be applied to estimate the parameters; therefore, Aknouche and Scotto (2022) [14] used a two-stage weighted least squares estimation instead.

Note that the probability mass function of the random variables cannot be given directly for the likelihood function in some cases; to solve this problem, saddlepoint approximation has been proposed. Daniel (1954) [15] introduced saddlepoint techniques into the statistical field, which have been extended by Field and Ronchetti (1990) [16], Jensen (1995) [17], and Butler (2007) [18]. Saddlepoint techniques have been used successfully in many applications because of the high accuracy with which they can approximate intractable densities and tail probabilities. Pedeli et al. (2015) [19] proposed an alternative approach based on the saddlepoint approximation to log-likelihood, and the saddlepoint maximum likelihood estimation (SPMLE) was used to estimate the parameters of the INAR model, which demonstrates the usefulness of this technique. Thus, through combining the MthINGARCH model of Aknouche and Scotto (2022) [14] and the saddlepoint approximation, we propose a modified multiplicative thinning-based INARCH model for modeling high overdispersion, before applying the saddlepoint method to the estimated parameters. Although the two-stage weighted least squares estimation could be used to estimate the parameters of our modified model, we still adopted the SPMLE as it was still expected to have a better performance than the two-stage weighted least squares estimation in practice. Here, we just consider the INARCH model instead of the INGARCH model because it is difficult and complex to give the conditional cumulant-generating function of random variables for the latter model when applying the saddlepoint approximation.

This article has the following structure. A modified multiplicative thinning-based INARCH model is given, alongside some related properties in Section 2. Moreover, we use the Poisson distribution and geometric distribution for innovations. Section 3 discusses the SPMLE and its asymptotic properties, then simulation studies for both models with SPMLE are also given. A real data example is analyzed with our modified models in Section 4, and comparisons with existing models are made. In-sample and out-of-sample forecasts are used to show the superiority of the SPMLE and our modified model. The conclusion is given in Section 5. Some details of SPMLE and proof of some theorems are presented in the Appendix A.

2. A Multiplicative Thinning-Based INARCH Model

Note that $\mathbb{N} = \{0, 1, 2, \dots\}$ and $\mathbb{Z} = \{\dots, -1, 0, 1, \dots\}$ are the set of non-negative integers and integers, respectively. It can be supposed that $\{\varepsilon_t, t \in \mathbb{Z}\}$ is a sequence of i.i.d. random variables with a mean of one and finite variance of σ^2. The modified multiplicative thinning-based INARCH (denoted by the MthINARCH(q)) model, which we deal with in this paper, is defined by

$$X_t = \lambda_t \varepsilon_t, \quad \lambda_t = \omega \circ m + \sum_{i=1}^{q} \alpha_i \circ X_{t-i}, \tag{2}$$

where $0 < \omega \leq 1$, $0 \leq \alpha_i < 1, i = 1, \dots, q$, m is a fixed positive integer number. In real applications, we can set m as the upper integer part of the sample mean. It is assumed that the Bernoulli terms corresponding to the binomial variables $\omega \circ m$ and $\alpha_i \circ X_{t-i}$ are mutually independent and independent of the sequence $\{\varepsilon_t, t \in \mathbb{Z}\}$. The reason that we defined the new model in this way can be explained as follows. The additive term 1 in λ_t and in (1) is unnatural, and is posed to ensure $\lambda_t > 0$, but we can achieve this by adjusting the range of ω; therefore, we adopted a simple version of λ_t in (2).

Now that we discuss the conditional mean and conditional variance of X_t. Note that $\mathscr{F}_{t-1}$ is the σ-field generated by $X_{t-1}, X_{t-2}, \dots$. For $E(\varepsilon_t) = 1$, let $\mu_t := E(X_t|\mathscr{F}_{t-1}) = E(\lambda_t \varepsilon_t|\mathscr{F}_{t-1}) = E(\varepsilon_t)E(\lambda_t|\mathscr{F}_{t-1}) = E(\lambda_t|\mathscr{F}_{t-1}) = \omega m + \sum_{i=1}^{q} \alpha_i X_{t-i}$. Then we can obtain the conditional variance; first, let $\nu_t := \mathrm{Var}(\lambda_t|\mathscr{F}_{t-1})$ and $\sigma_t^2 := \mathrm{Var}(X_t|\mathscr{F}_{t-1})$. For $E(\varepsilon_t) = 1, \mathrm{Var}(\varepsilon_t) = \sigma^2$, so $E(\varepsilon_t^2) = \sigma^2 + 1$. Therefore,

$$\begin{aligned}
\nu_t :&= \mathrm{Var}(\lambda_t|\mathscr{F}_{t-1}) = \omega(1-\omega)m + \sum_{i=1}^{q} \alpha_i(1-\alpha_i)X_{t-i}, \\
\sigma_t^2 :&= \mathrm{Var}(X_t|\mathscr{F}_{t-1}) = E(X_t^2|\mathscr{F}_{t-1}) - [E(X_t|\mathscr{F}_{t-1})]^2 = E(\lambda_t^2|\mathscr{F}_{t-1})E(\varepsilon_t^2) - \mu_t^2 \\
&= [\mathrm{Var}(\lambda_t|\mathscr{F}_{t-1}) + (E(\lambda_t|\mathscr{F}_{t-1}))^2]E(\varepsilon_t^2) - \mu_t^2 \\
&= (\sigma^2+1)(\nu_t + \mu_t^2) - \mu_t^2 = (\sigma^2+1)\nu_t + \sigma^2\mu_t^2.
\end{aligned}$$

Proposition 1. *The necessary and sufficient condition for the first-order stationarity of X_t defined in (2) is that all roots of $1 - \sum_{i=1}^{q} \alpha_i z^i = 0$ should lie outside the unit circle.*

Proposition 2. *The necessary and sufficient condition for the second-order stationarity of X_t defined in (2) is that $(\sigma^2+1)\sum_{i=1}^{q} \alpha_i^2 < 1$.*

Proofs of Propositions 1 and 2 are similar to the proofs of Theorems 2.1 and 2.2 in Aknouche and Scotto (2022) [14], so we omit the details.

For convenience, we need to specify the distribution of $\{\varepsilon_t\}$ in (2). First, we let $\varepsilon_t \sim P(1)$, then $E(\varepsilon_t) = \mathrm{Var}(\varepsilon_t) = 1$, and this model is denoted by PMthINARCH(q). It is easy to obtain

$$\mu_t = \omega m + \sum_{i=1}^{q} \alpha_i X_{t-i}, \quad \sigma_t^2 = 2\nu_t + \mu_t^2.$$

Second, let $\varepsilon_t \sim Ge(p^*)$. The mean of ε_t is $(1-p^*)/p^* = 1$, so we have $p^* = 0.5$ and the variance is $\mathrm{Var}(\varepsilon_t) = 2$. This model is denoted by GMthINARCH(q), then we have

$$\mu_t = \omega m + \sum_{i=1}^{q} \alpha_i X_{t-i}, \quad \sigma_t^2 = 3\nu_t + 2\mu_t^2.$$

3. Parameter Estimation

In this section, we will consider the SPMLE and its asymptotic properties, and a simulation study will be conducted to assess the performance of this estimator.

3.1. Saddlepoint Maximum Likelihood Estimation

Let $\theta = (\omega, \alpha_1, \ldots, \alpha_q)^{\mathrm{T}}$ be the unknown parameter vector. Note that according to the condition on ε_t, σ^2 is no longer an unknown parameter. The maximum likelihood estimator of θ was obtained by maximizing the conditional log-likelihood function

$$l(\theta) = \sum_{t=1}^{n} \log P(X_t = x_t | X_{t-1} = x_{t-1}, \ldots, X_{t-q} = x_{t-q}), \tag{3}$$

giving $\hat{\theta} = \arg\max_\theta l(\theta)$. But the above procedure is challenging to implement because it is difficult to give the likelihood function due to the thinning operations.

Now we discuss the SPMLE. The conditional moment generating function of X_t is

$$\begin{aligned} E(\mathrm{e}^{uX_t} | X_{t-1} = x_{t-1}, \ldots, X_{t-q} = x_{t-q}) &= E(\mathrm{e}^{u\lambda_t \varepsilon_t} | X_{t-1} = x_{t-1}, \ldots, X_{t-q} = x_{t-q}) \\ &= E(\mathrm{e}^{u(\omega \circ m + \sum_{i=1}^{q} \alpha_i \circ X_{t-i})\varepsilon_t} | X_{t-1} = x_{t-1}, \ldots, X_{t-q} = x_{t-q}) \\ &= E(\mathrm{e}^{u(\omega \circ m)\varepsilon_t}) \prod_{i=1}^{q} E(\mathrm{e}^{u(\alpha_i \circ x_{t-i})\varepsilon_t}). \end{aligned}$$

Remark 1. *Here we just consider the INARCH model instead of the INGARCH model because for the INGARCH model, the conditional cumulant-generating function of X_t should be given by $E(\mathrm{e}^{uX_t} | X_{t-1} = x_{t-1}, \ldots, X_{t-q} = x_{t-q}) = E(\mathrm{e}^{u(\omega \circ m + \sum_{i=1}^{q} \alpha_i \circ X_{t-i} + \sum_{j=1}^{p} \beta_j \circ \lambda_{t-i})\varepsilon_t} | X_{t-1} = x_{t-1}, \ldots, X_{t-q} = x_{t-q})$. Notice that X_t and λ_t are correlated, it is difficult and complex to show the conditional cumulant-generating function.*

Using the binomial theorem $(a+b)^n = \sum_{k=0}^{n} C_n^k a^{n-k} b^k$, we have

$$\begin{aligned} E(\mathrm{e}^{u(\omega \circ m)\varepsilon_t}) &= E\left[E(\mathrm{e}^{u(\omega \circ m)\varepsilon_t} | \varepsilon_t)\right] = E(\omega \mathrm{e}^{u\varepsilon_t} + (1-\omega))^m \\ &= E\left[\sum_{r=0}^{m} C_m^r (1-\omega)^r \omega^{m-r} \mathrm{e}^{u(m-r)\varepsilon_t}\right] = \sum_{r=0}^{m} C_m^r (1-\omega)^r \omega^{m-r} E(\mathrm{e}^{u(m-r)\varepsilon_t}). \end{aligned}$$

Similarly, we also have

$$E(\mathrm{e}^{u(\alpha_i \circ x_{t-i})\varepsilon_t}) = \sum_{r=0}^{x_{t-i}} C_{x_{t-i}}^r (1-\alpha_i)^r \alpha_i^{x_{t-i}-r} E(\mathrm{e}^{u(x_{t-i}-r)\varepsilon_t}).$$

Therefore, for the PMthINARCH(q) model, we have

$$\begin{aligned} E(\mathrm{e}^{u(\omega \circ m)\varepsilon_t}) &= \sum_{r=0}^{m} C_m^r (1-\omega)^r \omega^{m-r} \mathrm{e}^{(\mathrm{e}^{u(m-r)}-1)}, \\ E(\mathrm{e}^{u(\alpha_i \circ x_{t-i})\varepsilon_t}) &= \sum_{r=0}^{x_{t-i}} C_{x_{t-i}}^r (1-\alpha_i)^r \alpha_i^{x_{t-i}-r} \mathrm{e}^{(\mathrm{e}^{u(x_{t-i}-r)}-1)}, \end{aligned}$$

while for the GMthINARCH(q) model, we have

$$\begin{aligned} E(\mathrm{e}^{u(\omega \circ m)\varepsilon_t}) &= \sum_{r=0}^{m} C_m^r (1-\omega)^r \omega^{m-r} \frac{1}{2-\mathrm{e}^{u(m-r)}}, \\ E(\mathrm{e}^{u(\alpha_i \circ x_{t-i})\varepsilon_t}) &= \sum_{r=0}^{x_{t-i}} C_{x_{t-i}}^r (1-\alpha_i)^r \alpha_i^{x_{t-i}-r} \frac{1}{2-\mathrm{e}^{u(x_{t-i}-r)}}. \end{aligned}$$

Thus the conditional cumulant-generating function of X_t is:

$$K_t(u) = \log[E(\mathrm{e}^{uX_t}|X_{t-1} = x_{t-1}, \ldots, X_{t-q} = x_{t-q})] = \log E(\mathrm{e}^{u(\omega \circ m)\varepsilon_t}) + \sum_{i=1}^{q} \log E(\mathrm{e}^{u(\alpha_i \circ x_{t-i})\varepsilon_t}).$$

A highly accurate approximation to the conditional mass function of X_t at x_t is provided by the saddlepoint approximation:

$$\tilde{f}_{X_t|X_{t-1}=x_{t-1},\ldots,X_{t-q}=x_{t-q}}(x_t) = \left(2\pi K_t''(\tilde{u}_t)\right)^{-\frac{1}{2}} \exp\{K_t(\tilde{u}_t) - \tilde{u}_t x_t\}, \tag{4}$$

where $\tilde{u}_t$ is the unique value of u which satisfies the saddlepoint equation $K_t'(u) = x_t$, with K_t' and K_t'' represent the first and second order derivatives of K_t with respect to u. Notice that it is difficult to solve the saddlepoint equation $K_t'(u) = x_t$ analytically; similar to that mentioned in Pedeli et al. (2015) [19], we can use the Newton–Raphson method to solve this equation.

The log-likelihood function (3) can be approximated by summing the logarithms of the corresponding density approximations (4), yielding:

$$\tilde{L}_n(\theta) = \sum_{t=1}^{n} \tilde{l}_t(\theta) := \sum_{t=1}^{n} \log \tilde{f}_{X_t|X_{t-1}=x_{t-1},\ldots,X_{t-q}=x_{t-q}}(x_t). \tag{5}$$

The value θ maximizing this expression is called the saddlepoint maximum likelihood estimator (SPMLE).

3.2. Asymptotic Properties of the SPMLE

Now we discuss the asymptotic properties of the SPMLE. First we give the first-order Taylor expansion of $K_t'(u)$ at $u = 0$ yields,

$$K_t'(u) = K_t'(0) + uK_t''(0) + o(u) = \mu_t(\theta) + u\sigma_t^2(\theta) + o(u), \tag{6}$$

where $\mu_t(\theta)$ and $\sigma_t^2(\theta)$ are the conditional mean and conditional variance of X_t. Notice that $\tilde{u}_t$ can be given by $K_t'(\tilde{u}_t) = x_t$, so with the Taylor series expansion of $K_t'(u)$ in (6), we have:

$$\tilde{u}_t = \frac{x_t - \mu_t(\theta)}{\sigma_t^2(\theta)} + o(1), \quad t = q+1, \ldots, n. \tag{7}$$

Then, we can obtain the second-order Taylor expansion of $K_t(u)$ at $u = 0$, which is:

$$K_t(u) \approx uK_t'(0) + \frac{u^2}{2}K_t''(0) = u\mu_t(\theta) + \frac{u^2}{2}\sigma_t^2(\theta). \tag{8}$$

Focusing on the exponent of the saddlepoint approximation (4), Equation (8) gives

$$K_t(u) - ux_t \approx u(\mu_t(\theta) - x_t) + \frac{u^2}{2}\sigma_t^2(\theta).$$

Then using Equation (7), we have

$$K_t(\tilde{u}_t) - \tilde{u}_t x_t \approx -\frac{[x_t - \mu_t(\theta)]^2}{2\sigma_t^2(\theta)}. \tag{9}$$

Hence, we can derive from (8) and (9) that the first-order saddlepoint approximation to the conditional probability mass function is approximately:

$$\tilde{f}_{X_t|X_{t-1}=x_{t-1},\ldots,X_{t-q}=x_{t-q}}(x_t) = \big(2\pi K_t''(\tilde{u}_t)\big)^{-\frac{1}{2}}$$
$$\times \exp\left[-\frac{(x_t - \omega m - \sum_{i=1}^{q}\alpha_i x_{t-i})^2}{2\Big[(\sigma^2+1)(\omega(1-\omega)m + \sum_{i=1}^{q}\alpha_i(1-\alpha_i)x_{t-i}) + \sigma^2(\omega m + \sum_{i=1}^{q}\alpha_i x_{t-i})^2\Big]}\right].$$

Therefore, $\tilde{L}_n(\theta) = \sum_{t=1}^{n}\tilde{l}_t(\theta) = \sum_{t=1}^{n}\log\tilde{f}_{X_t|X_{t-1}=x_{t-1},\ldots,X_{t-q}=x_{t-q}}(x_t)$ is the quasi-likelihood function for the estimation of θ. To establish the large-sample properties, we have

$$L_n(\theta) = \sum_{t=1}^{n} l_t(\theta) = \sum_{t=1}^{n}\log f_{X_t|X_{t-1}=x_{t-1},\ldots,X_{t-q}=x_{t-q}}(x_t),$$

which is the ergodic approximation of $\tilde{L}_n(\theta)$. The first and second derivatives of the quasi-likelihood function are given in the Appendix A. The strong convergence and asymptotic normality for the SPMLE $\hat{\theta}_n$ are established in the following theorems.

First of all, the assumptions for Theorems 1 and 2 are listed as follows.

Assumption 1. *The solution of the MthINARCH process is strictly stationary and ergodic.*

Assumption 2. Θ *is compact and* $\theta_0 \in \mathring{\Theta}$*, where* $\mathring{\Theta}$ *denotes the interior of* Θ*. For technical reasons, we assumed the lower and upper values of each component of parameters as* $0 < \omega_L \le \omega \le \omega_U \le 1$ *and* $0 \le \alpha_L \le \alpha_i \le \alpha_U < 1$, $i = 1, \ldots, q$.

Theorem 1. *Let* $\hat{\theta}_n$ *be a sequence of SPMLEs satisfying* $\hat{\theta}_n = \arg\max_{\theta\in\Theta}\tilde{L}_n(\theta)$*, then under Assumptions 1 and 2,* $\hat{\theta}_n$ *converges to* θ_0 *almost as surely, as* $n \to \infty$.

Theorem 2. *Under Assumptions 1 and 2, there exists a sequence of maximizers* $\hat{\theta}_n$ *of* $\tilde{L}_n(\theta)$ *such as that of* $n \to \infty$,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, \Sigma^{-1}),$$

where

$$\Sigma = -E_{\theta_0}\left(\frac{\partial^2 l_t(\theta_0)}{\partial\theta\partial\theta^{\mathrm{T}}}\right),$$

and Σ *is positively definite.*

3.3. Simulation Study

In this section, simulation studies of PMthINARCH(q) and GMthINARCH(q) models for finite sample size are given, where $q = 2$. Here, we used several combinations to show the performance of SPMLE, and the mean absolute deviation error (MADE) $\frac{1}{s}\sum_{j=1}^{s}|\hat{\theta}_j - \theta_j|$ was used as the evaluation criterion; here, s is the number of replications. The sample size is $n = 100, 200, 500$, and the number of replications is $s = 200$. We used the following combinations of $(\omega, \alpha_1, \alpha_2)^{\mathrm{T}}$ as the true values to generate the random sample: A1 $= (0.65, 0.4, 0.4)^{\mathrm{T}}$, A2 $= (0.9, 0.5, 0.3)^{\mathrm{T}}$ for the PMthINARCH(2) model, and B1 $= (0.8, 0.4, 0.4)^{\mathrm{T}}$, B2 $= (0.65, 0.3, 0.5)^{\mathrm{T}}$ for the GMthINARCH(2) model. Tables 1 and 2 show the results of these simulations. Notice that as the sample sizes become larger, the MADEs become smaller, and the estimates seem to be close to the true values. Therefore, the SPMLE performs well.

Table 1. Mean and MADE of estimates for PMthINARCH(2) model with SPMLE.

Model				ω	α_1	α_2
A1	$m = 3$	$n = 100$	Mean	0.6069	0.5356	0.3569
			MADE	0.3681	0.2866	0.2510
		$n = 200$	Mean	0.5722	0.5026	0.3952
			MADE	0.3557	0.2434	0.2243
		$n = 500$	Mean	0.6436	0.4888	0.4140
			MADE	0.2724	0.1287	0.1005
A2	$m = 8$	$n = 100$	Mean	0.7782	0.5076	0.4750
			MADE	0.2533	0.2752	0.3007
		$n = 200$	Mean	0.7935	0.5161	0.4701
			MADE	0.2318	0.2527	0.2778
		$n = 500$	Mean	0.8703	0.5170	0.4677
			MADE	0.1752	0.2155	0.2390

Table 2. Mean and MADE of estimates for GMthINARCH(2) model with SPMLE.

Model				ω	α_1	α_2
B1	$m = 4$	$n = 100$	Mean	0.7821	0.2930	0.2870
			MADE	0.1195	0.1499	0.1766
		$n = 200$	Mean	0.8190	0.3611	0.3185
			MADE	0.1121	0.1425	0.1640
		$n = 500$	Mean	0.8456	0.3610	0.3298
			MADE	0.0601	0.1331	0.1414
B2	$m = 6$	$n = 100$	Mean	0.4718	0.2086	0.3811
			MADE	0.1965	0.1466	0.1463
		$n = 200$	Mean	0.5186	0.2632	0.5080
			MADE	0.1607	0.1198	0.1412
		$n = 500$	Mean	0.5468	0.2874	0.4896
			MADE	0.1415	0.1050	0.0770

4. A Real Example

Here, we considered the number of tick changes by the minute of the euro to the British pound exchange rate (ExRate for short) on December 12th from 9.00 a.m. to 9.00 p.m. The dataset is available at the website http://www.histdata.com/ (accessed on 17 January 2023). The series comprises of 720 observations with a sample mean of 13.2153 and a sample variance of 224.2498. Obviously, the sample variance is much larger than the sample mean, which shows high overdispersion, and this high overdispersion can also be seen in Figure 1a. Figure 1b,c are the plots of the autocorrelation function (ACF), and the partial autocorrelation function (PACF) means that we know the tick changes are correlated.

We analyzed the data using the PMthINARCH(3) model, GMthINARCH(3) model, Poisson INAR(3) (here denoted by PINAR(3) for short) model, and the INARCH(3) model. The Poisson INAR model is mentioned in Pedeli et al. (2015) [19], and the SPMLE was used to estimate the parameters. Here, the innovations in the PINAR model were assumed to be Poisson with a mean of one. The INARCH model with a Poisson deviate was proposed by Ferland et al. (2006) [5], and the MLE was used to estimate the parameters. According to Aknouche and Scotto (2022) [14], in real applications, we can set m as the upper integer part of the sample mean. Here the sample mean is 13.2153, so m is set to the value of 14. Table 3 gives the estimates of SPMLE and the values of the Akaike information criterion (AIC) and Bayesian information criterion (BIC). According to Table 3, it is clear to see that the values of AIC and BIC of PMthINARCH(3) and GMthINARCH(3) are smaller than those of the PINAR(3) and INARCH(3) models, the values of AIC and BIC of INARCH(3) are smaller than those of the PINAR(3) model. Moreover, the values of AIC and BIC of PMthINARCH(3) are smaller than those of GMthINARCH(3). In summary, the INARCH model performed better than the PINAR model; meanwhile, the PMthINARCH model and GMthINARCH model performed better than the PINAR model and INARCH model.

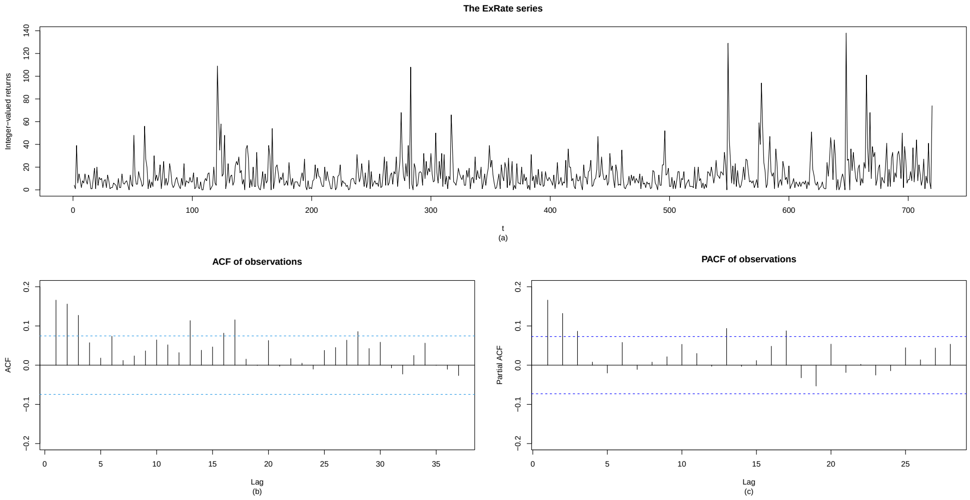

Figure 1. (**a**) The plot of integer-valued series of ExRate. (**b**) The plot of ACF of observations. (**c**) The plot of PACF of observations.

Table 3. Estimation results: AIC and BIC values for PMthINARCH(3), GMthINARCH(3), PINAR(3) and INARCH(3) models.

PMthINARCH(3)	ω 0.3242	α_1 0.5214	α_2 0.1945	α_3 0.0842	AIC 1395.296	BIC 1413.613
GMthINARCH(3)	ω 0.4904	α_1 0.2532	α_2 0.2155	α_3 0.2392	AIC 1402.472	BIC 1420.789
PINAR(3)	α_1 0.1335	α_2 0.4116	α_3 0.3901		AIC 1572.806	BIC 1586.544
INARCH(3)	ω 8.5670	α_1 0.1140	α_2 0.1379	α_3 0.1009	AIC 1524.638	BIC 1542.955

According to Aknouche and Scotto (2022) [14], the two-stage weighted least squares estimation (2SWLSE) was used to estimate the parameters of the MthINGARCH model. Therefore, to compare the performance of 2SWLSE and SPMLE, and the performance of PMthINARCH, GMthINARCH, and PINAR models, to consider the in-sample and out-of-sample forecasts of these two estimation methods and the three models above, respectively. First, we considered the in-sample forecast. We used all of the observations to estimate the model, and then we could forecast the last 10 observations 711–720, the last 15 observations 706–720, and the last 20 observations 701–720; these three-time horizons of in-sample forecast are denoted by C1, C2, and C3, respectively. Similar to the in-sample forecast process, we also considered the out-of-sample forecast and divided all the observations into three-time horizons: the first one was 1–710 and 711–720, the second one was 1–705 and 706–720, and the third one was 1–700 and 701–720, which are denoted by D1, D2, and D3, respectively.

Here we illustrate the performance of the considered models by comparing the MADEs of each forecast. The MADEs of in-sample forecasts and out-of-sample forecasts for three models with SPMLE are shown in Table 4. The MADEs of the in-sample forecasts and out-of-sample forecasts for the PMthINARCH model with 2SWLSE and SPMLE are shown in Table 5, and the in-sample forecasts and out-of-sample forecasts for the GMthINARCH model with 2SWLSE and SPMLE are shown in Table 6. According to Table 4, the MADEs of PMthINARCH(3) and GMthINARCH(3) are smaller than those of PINAR(3), Tables 5 and 6 show that the MADEs of PMthINARCH(3) and GMthINARCH(3) of SPMLE are smaller than those of 2SWLSE; meanwhile, in these three Tables, the MADEs of in-sample forecasts were smaller than those of out-of-sample forecasts. In summary, the PMthINARCH model and GMthINARCH model were superior to the PINAR model in modeling this real data set, and the PMthINARCH model performed better than the GMthINARCH model. Meanwhile, the performance of SPMLE was better than 2SWLSE for MthINARCH models.

Table 4. MADEs of in-sample forecasts and out-of-sample forecasts for PMthINARCH(3), GMthINARCH(3), and PINAR(3) models with SPMLE.

Methods of Forecast		PMthINARCH	GMthINARCH	PINAR
In-sample	C1	15.30	16.80	17.40
	C2	15.87	17.67	18.40
	C3	16.65	20.70	21.90
Out-of-sample	D1	17.50	17.70	22.50
	D2	19.47	19.80	23.80
	D3	20.50	25.25	27.50

Table 5. MADEs of in-sample forecasts and out-of-sample forecasts for PMthINARCH(3) model with SPMLE and 2SWLSE.

Methods of Forecast		SPMLE	2SWLSE
In-sample	C1	15.30	16.20
	C2	15.87	17.20
	C3	16.65	18.55
Out-of-sample	D1	17.50	18.60
	D2	19.47	21.67
	D3	20.50	22.70

Table 6. MADEs of in-sample forecasts and out-of-sample forecasts for GMthINARCH(3) model with SPMLE and 2SWLSE.

Methods of Forecast		SPMLE	2SWLSE
In-sample	C1	16.80	17.20
	C2	17.67	18.07
	C3	20.70	21.05
Out-of-sample	D1	17.70	19.90
	D2	19.80	22.87
	D3	25.25	26.50

5. Conclusions

In this paper, we modified a multiplicative thinning-based INARCH model. The probability mass function of random variables is provided by saddlepoint approximation. We used the SPMLE to estimate the parameters and obtain the asymptotic distribution of the SPMLE. Moreover, to show the superiority of the MthINARCH models and the

SPMLE, we used the PMthINARCH(q) process and GMthINARCH(q) process for discussion and comparison. The SPMLE performs well in the simulation studies. A real dataset indicates that the PMthINARCH model and the GMthINARCH model are able to describe the overdispersed integer-valued data, and the real data example leads to a superior performance of the MthINARCH models compared with the PINAR and INARCH models. In addition, the results also show a superior performance of SPMLE compared with 2SWLSE.

For further discussion, more research is needed for some aspects. Here we used the Poisson distribution and geometric distribution for ε_t; however, we could use the negative binomial distribution or some zero-inflated distributions as well. Moreover, we just considered the INARCH model, so the corresponding INGARCH model should be considered as well.

Author Contributions: Conceptualization, F.Z.; methodology, Y.X.; software, Y.X. and Q.L.; validation, Y.X. and Q.L.; formal analysis, Y.X. and Q.L.; investigation, Y.X. and F.Z.; resources, Q.L.; data curation, Y.X. and Q.L.; writing—original draft preparation, Y.X., Q.L. and F.Z.; writing—review and editing, Y.X., Q.L. and F.Z.; visualization, Y.X.; supervision, F.Z.; project administration, F.Z.; funding acquisition, Q.L. and F.Z. All authors have read and agreed to the published version of the manuscript.

Funding: Li's work is supported by the National Natural Science Foundation of China (No. 12201069), the Natural Science Foundation of Jilin Province (No. 20210101160JC), the Science and Technology Research Project of Education Bureau of Jilin Province (No. JJKH20220820KJ), and Natural Science Foundation Projects of CCNU (CSJJ2022006ZK). Zhu's work is supported by the National Natural Science Foundation of China (No. 12271206) and the Natural Science Foundation of Jilin Province (No. 20210101143JC).

Data Availability Statement: The dataset is available at the website http://www.histdata.com/ (accessed on 17 January 2023).

Acknowledgments: The authors are very grateful to three reviewers for their constructive suggestions and comments, leading to a substantial improvement in the presentation and contents.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Appendix A.1. Details of SPMLE

Here, we give the derivatives of $K_t(u)$ mentioned in Section 3.1 of PMthINARCH(q) and GMthINARCH(q). Now we give $K_t'(u)$ and $K_t''(u)$ of PMthINARCH(q). In Section 3.1, we have

$$K_t(u) = \log E(\mathrm{e}^{u(\omega \circ m)\varepsilon_t}) + \sum_{i=1}^{q} \log E(\mathrm{e}^{u(\alpha_i \circ x_{t-i})\varepsilon_t}) = \log a_1 + \sum_{i=1}^{q} \log b_1,$$

so the derivatives of $K_t(u)$ are given by

$$K_t'(u) = \frac{c_1}{a_1} + \sum_{i=1}^{q} \frac{d_1}{b_1}, \quad K_t''(u) = \frac{e_1 a_1 - c_1^2}{a_1^2} + \sum_{i=1}^{q} \frac{f_1 b_1 - d_1^2}{b_1^2},$$

where

$$a_1 = \sum_{r=0}^{m} C_m^r (1-\omega)^r \omega^{m-r} \mathrm{e}^{\mathrm{e}^{u(m-r)}-1},$$

$$b_1 = \sum_{r=0}^{x_{t-i}} C_{x_{t-i}}^r (1-\alpha_i)^r \alpha_i^{x_{t-i}-r} \mathrm{e}^{\mathrm{e}^{u(x_{t-i}-r)}-1},$$

$$c_1 = \sum_{r=0}^{m} C_m^r (1-\omega)^r \omega^{m-r} \mathrm{e}^{u(m-r)} \mathrm{e}^{\mathrm{e}^{u(m-r)}-1},$$

$$d_1 = \sum_{r=0}^{x_{t-i}} C_{x_{t-i}}^r (1-\alpha_i)^r \alpha_i^{x_{t-i}-r} \mathrm{e}^{u(x_{t-i}-r)} \mathrm{e}^{\mathrm{e}^{u(x_{t-i}-r)}-1},$$
$$e_1 = \sum_{r=0}^{m} C_m^r (1-\omega)^r \omega^{m-r} \mathrm{e}^{u(m-r)} (m-r)^2 \mathrm{e}^{\mathrm{e}^{u(m-r)}-1}[1+\mathrm{e}^{u(m-r)}],$$
$$f_1 = \sum_{r=0}^{x_{t-i}} C_{x_{t-i}}^r (1-\alpha_i)^r \alpha_i^{x_{t-i}-r} (x_{t-i}-r)^2 \mathrm{e}^{u(x_{t-i}-r)} \mathrm{e}^{\mathrm{e}^{u(x_{t-i}-r)}-1}[1+\mathrm{e}^{u(x_{t-i}-r)}].$$

Then we give $K_t'(u)$ and $K_t''(u)$ of GMthINARCH(q). In Section 3.1, we have

$$K_t(u) = \log E(\mathrm{e}^{u(\omega \circ m)\varepsilon_t}) + \sum_{i=1}^{q} \log E(\mathrm{e}^{u(\alpha_i \circ x_{t-i})\varepsilon_t}) = \log a_2 + \sum_{i=1}^{q} \log b_2,$$

so the derivatives of $K_t(u)$ are given by

$$K_t'(u) = \frac{c_2}{a_2} + \sum_{t=1}^{q} \frac{d_2}{b_2}, \quad K_t''(u) = \frac{e_2 a_2 - c_2^2}{a_2^2} + \sum_{t=1}^{q} \frac{f_2 b_2 - d_2^2}{b_2^2},$$

where

$$a_2 = \sum_{r=0}^{m} C_m^r (1-\omega)^r \omega^{m-r} \frac{1}{2-(2-\mathrm{e}^{u(m-r)})},$$
$$b_2 = \sum_{r=0}^{x_{t-i}} C_{x_{t-i}}^r (1-\alpha_i)^r \alpha_i^{x_{t-i}-r} \frac{1}{2-(2-\mathrm{e}^{u(x_{t-i}-r)})},$$
$$c_2 = \frac{1}{4} \sum_{r=0}^{m} C_m^r (1-\omega)^r \omega^{m-r} (m-r) \frac{\mathrm{e}^{u(m-r)}}{[1-(1-\frac{1}{2}\mathrm{e}^{u(m-r)})]^2},$$
$$d_2 = \frac{1}{4} \sum_{r=0}^{x_{t-i}} C_{x_{t-i}}^r (1-\alpha_i)^r \alpha_i^{x_{t-i}-r} (x_{t-i}-r) \frac{\mathrm{e}^{u(x_{t-i}-r)}}{[1-(1-\frac{1}{2}\mathrm{e}^{u(x_{t-i}-r)})]^2},$$
$$e_2 = \frac{1}{4} \sum_{r=0}^{m} C_m^r (1-\omega)^r \omega^{m-r} (m-r)^2 \mathrm{e}^{u(m-r)} \frac{1+\frac{1}{2}\mathrm{e}^{u(m-r)}}{[1-(1-\frac{1}{2}\mathrm{e}^{u(m-r)})]^3},$$
$$f_2 = \frac{1}{4} \sum_{r=0}^{x_{t-i}} C_{x_{t-i}}^r (1-\alpha_i)^r \alpha_i^{x_{t-i}-r} (x_{t-i}-r)^2 \mathrm{e}^{u(x_{t-i}-r)} \frac{1+\frac{1}{2}\mathrm{e}^{u(x_{t-i}-r)}}{[1-(1-\frac{1}{2}\mathrm{e}^{u(x_{t-i}-r)})]^3}.$$

Appendix A.2. Derivatives of the Quasi-Likelihood Function

The conditional log-quasi-likelihood function $l_t(\theta)$ is continuous on Θ: for $1 \le t \le n$,

$$\frac{\partial l_t(\theta)}{\partial \theta} = m_1 \frac{\partial \mu_t(\theta)}{\partial \theta} + m_2 \frac{\partial \sigma_t^2(\theta)}{\partial \theta},$$
$$\frac{\partial^2 l_t(\theta)}{\partial \theta \partial \theta^{\mathrm{T}}} = (m_1 - m_3) \frac{\partial^2 \mu_t(\theta)}{\partial \theta \partial \theta^{\mathrm{T}}} - 2 m_1 m_3 \frac{\partial \mu_t(\theta)}{\partial \theta} \frac{\partial \sigma_t^2(\theta)}{\partial \theta^{\mathrm{T}}} + (m_2 + \frac{m_3^2}{2} - m_1^2 m_3) \frac{\partial^2 \sigma_t^2(\theta)}{\partial \theta \partial \theta^{\mathrm{T}}},$$

where

$$m_1 = \frac{X_t - \mu_t(\theta)}{\sigma_t^2(\theta)}, \quad m_2 = \frac{(X_t - \mu_t(\theta))^2 - \sigma_t^2(\theta)}{2\sigma_t^4(\theta)}, \quad m_3 = \frac{1}{\sigma_t^2(\theta)}.$$

Then the first and second derivatives of $\mu_t(\theta)$ and $\sigma_t^2(\theta)$ can be easily expressed by

$$\frac{\partial \mu_t(\theta)}{\partial \omega} = m, \quad \frac{\partial \mu_t(\theta)}{\partial \alpha_i} = X_{t-i},$$
$$\frac{\partial \sigma_t^2(\theta)}{\partial \omega} = (\sigma^2 + 1)(m - 2\omega m) + 2\sigma^2 (m^2 \omega + m \sum_{i=1}^{q} \alpha_i X_{t-i}),$$

$$\frac{\partial \sigma_t^2(\theta)}{\partial \alpha_i} = (\sigma^2+1)(X_{t-i} - 2\alpha_i X_{t-i}) + 2\sigma^2(m\omega X_{t-i} + \alpha_i X_{t-i}^2),$$
$$\frac{\partial^2 \mu_t(\theta)}{\partial \omega^2} = 0, \quad \frac{\partial^2 \mu_t(\theta)}{\partial \alpha_i^2} = 1, \quad \frac{\partial^2 \mu_t(\theta)}{\partial \omega \alpha_i} = 0,$$
$$\frac{\partial^2 \sigma_t^2(\theta)}{\partial \omega^2} = -2m(\sigma^2+1) + 2m^2\sigma^2, \quad \frac{\partial^2 \sigma_t^2(\theta)}{\partial \alpha_i^2} = -2X_{t-i}(\sigma^2+1) + 2X_{t-i}^2\sigma^2,$$
$$\frac{\partial^2 \sigma_t^2(\theta)}{\partial \omega \alpha_i} = 2m\sigma^2 X_{t-i}.$$

Appendix A.3. Proof of Theorem 1

The techniques used here are mainly based on Francq and Zakoïan (2004) [20]. We will establish the following intermediate results:

(i) $\lim_{n\to\infty} \sup_{\theta\in\Theta} \left| \frac{1}{n}(L_n(\theta) - \tilde{L}_n(\theta)) \right| = 0 \quad a.s.$
(ii) $E(l_t(\theta))$ is continuous in θ.
(iii) It exists $t \in \mathbb{Z}$ such that $\sigma_t^2(\theta) = \sigma_t^2(\theta_0)$ a.s., then $\Rightarrow \theta = \theta_0$.
(iv) Any $\theta \neq \theta_0$ has a neighbourhood $V(\theta)$ such that

$$\limsup_{n\to\infty} \sup_{\theta^*\in V_k(\theta)\cap\Theta} \frac{1}{n}\tilde{L}_n(\theta^*) > E_{\theta_0} l_1(\theta_0) \quad a.s.$$

First we prove (i). Let $at := \sup_{\theta\in\Theta} |\tilde{\mu}_t(\theta) - \mu_t(\theta)|$, $b_t := \sup_{\theta\in\Theta} |\tilde{\sigma}_t^2(\theta) - \sigma_t^2(\theta)|$. Standard arguments from Corollary 2.2 in Aknouche and Francq (2023) [21] show that $a_t(1 + X_t + \sup_{\theta\in\Theta}\mu_t(\theta)) \to 0, a.s.$ and $b_t(1 + X_t^2 + \sup_{\theta\in\Theta}\mu_t^2(\theta)) \to 0, a.s., t\to\infty$, so we obtain the inequality

$$\begin{aligned}
\sup_{\theta\in\Theta}\left|\frac{1}{n}(L_n(\theta) - \tilde{L}_n(\theta))\right| &= \sup_{\theta\in\Theta}\left|\frac{1}{2n}\sum_{t=1}^n \log\frac{\tilde{\sigma}_t^2(\theta)}{\sigma_t^2(\theta)} + (\frac{(x_t-\tilde{\mu}_t)^2}{\tilde{\sigma}_t^2} - \frac{(x_t-\mu_t(\theta))^2}{\sigma_t^2})\right| \\
&\leq \sup_{\theta\in\Theta}\left|\frac{1}{2n}\sum_{t=1}^n \frac{\tilde{\sigma}_t^2(\theta) - \sigma_t^2(\theta)}{\sigma_t^2(\theta)} + (\frac{(x_t-\tilde{\mu}_t(\theta))^2}{\tilde{\sigma}_t^2(\theta)} - \frac{(x_t-\mu_t(\theta))^2}{\sigma_t^2})\right| \\
&\leq \sup_{\theta\in\Theta}\frac{1}{2n}\sum_{t=1}^n \frac{|\tilde{\sigma}_t^2(\theta) - \sigma_t^2(\theta)|}{\sigma_t^2(\theta)} + \frac{|\tilde{\mu}_t(\theta) - \mu_t(\theta)||\mu_t(\theta) + \tilde{\mu}_t(\theta) - 2X_t|}{\tilde{\sigma}_t^2(\theta)} \\
&+ \frac{|\tilde{\sigma}_t^2(\theta) - \sigma_t^2(\theta)||X_t - \mu_t(\theta)|^2}{\sigma_t^2(\theta)\tilde{\sigma}_t^2(\theta)} \\
&\leq \frac{1}{2n}\sum_{t=1}^n \frac{2}{\sigma_t^2(\theta)} a_t(1 + X_t + \sup_{\theta\in\Theta}\mu_t(\theta)) + \frac{1+\tilde{\sigma}_t^2(\theta)}{\sigma_t^2(\theta)\tilde{\sigma}_t^2(\theta)} c_t(1 + X_t^2 + \sup_{\theta\in\Theta}\mu_t^2(\theta)).
\end{aligned}$$

The a.s. limit holds because of the Cesàro lemma.

We prove (ii) now. For any $\theta\in\Theta$, let $V_\eta(\theta) = B(\theta,\eta)$ be an open ball centered at θ with radius η,

$$|l_t(\tilde{\theta}) - l_t(\theta)| \leq |\sigma_t^2(\tilde{\theta}) - \sigma_t^2(\theta)|\left|\frac{X_t^2 + \mu_t^2(\theta) + \sigma_t^2(\tilde{\theta})}{\sigma_t^2(\theta)\sigma_t^2(\tilde{\theta})}\right| + \frac{|\mu_t(\tilde{\theta}) - \mu_t(\theta)||\mu_t(\theta) + \mu_t(\tilde{\theta}) - 2X_t|}{\sigma_t^2(\tilde{\theta})}.$$

Then

$$\begin{aligned}
E\left(\sup_{\theta\in\tilde{V}_\eta(\theta)}|l_t(\tilde{\theta}) - l_t(\theta)|\right) &\leq \| \sigma_t^2(\tilde{\theta}) - \sigma_t^2(\theta) \|_2 \| \frac{X_t^2 + \mu_t^2(\theta) + \sigma_t^2(\tilde{\theta})}{\sigma_t^2(\theta)\sigma_t^2(\tilde{\theta})} \|_2 \\
&+ \frac{\| \mu_t(\tilde{\theta}) - \mu_t(\theta) \|_2 \| \mu_t(\theta) + \mu_t(\tilde{\theta}) - 2X_t \|_2}{\sigma_t^2(\tilde{\theta})} \to 0, \quad as \quad \eta\to 0.
\end{aligned}$$

Next, we check (iii). By Jensen's inequality, we have

$$\begin{aligned}
E[l_t(\theta) - l_t(\theta_0)] &= E\left[E\left(\frac{1}{2}\log\frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta)} + \frac{(x_t - \mu_t(\theta_0))^2}{2\sigma_t^2(\theta_0)} - \frac{(x_t - \mu_t(\theta))^2}{2\sigma_t^2(\theta)}\middle|\mathscr{F}_{t-1}\right)\right] \\
&\leq E\left[\log E\left(\frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta)}\middle|\mathscr{F}_{t-1}\right)\right] \\
&= E(\log(1)) = 0.
\end{aligned}$$

The equality holds if $\dfrac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta)} = 1$ a.s. $\mathscr{F}_{t-1}$, i.e. $\theta = \theta_0$.

Then the proof of (iv) is similar to that in the Supplementary Material A.4 in Xu and Zhu (2022) [22]. Here we omit the details.

Appendix A.4. Proof of the Positive Definiteness of Σ

Here, we prove the positive definiteness of Σ. By definition of positive definiteness, we need to prove for any $\xi = (\xi_0, \xi_1, \ldots, \xi_q)^{\mathrm{T}} \in \mathbb{R}^{q+1}$, if $\xi^{\mathrm{T}}\Sigma\xi = 0$, then $\xi = 0$.

$$\begin{aligned}
\xi^{\mathrm{T}}\Sigma\xi &= \xi^{\mathrm{T}} E\left[\frac{1}{2\sigma_t^4(\theta_0)}\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta}\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta^{\mathrm{T}}} + \frac{1}{\sigma_t^2(\theta_0)}\frac{\partial\mu_t(\theta_0)}{\partial\theta}\frac{\partial\mu_t(\theta_0)}{\partial\theta^{\mathrm{T}}}\right]\xi \\
&= E\left[\frac{1}{2\sigma_t^4(\theta_0)}(\xi^{\mathrm{T}}\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta})^2 + \frac{1}{\sigma_t^2(\theta_0)}(\xi^{\mathrm{T}}\frac{\partial\mu_t(\theta_0)}{\partial\theta})^2\right].
\end{aligned}$$

Suppose the left-hand side is 0, then under Assumption 1, the expectation in the right-hand side is 0 for any $t \in \mathbb{Z}$. Because $\sigma_t^2(\theta_0) > 0$, this expectation is always greater than or equal to 0. It equals 0 only when $\xi^{\mathrm{T}}\dfrac{\partial\sigma_t^2(\theta_0)}{\partial\theta} = 0$ and $\xi^{\mathrm{T}}\dfrac{\partial\mu_t(\theta_0)}{\partial\theta} = 0$ almost surely. Thus, $\xi^{\mathrm{T}}\Sigma\xi = 0$ yields $\xi^{\mathrm{T}}\dfrac{\partial\sigma_t^2(\theta_0)}{\partial\theta} = 0$ and $\xi^{\mathrm{T}}\dfrac{\partial\mu_t(\theta_0)}{\partial\theta} = 0$ a.s. for $t \in \mathbb{Z}$, and vice versa.

Using vector form of $\dfrac{\partial\sigma_t^2(\theta_0)}{\partial\theta}$, we have

$$\boldsymbol{\xi_a}^{\mathrm{T}}\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta} = \boldsymbol{\xi}^{\mathrm{T}}\begin{pmatrix}
(\sigma^2+1)(m - 2\omega m) + 2\sigma^2(\omega m^2 + m\sum_{i=1}^q \alpha_i X_{t-i}) \\
(\sigma^2+1)(X_{t-1} - 2\alpha_1 X_{t-1}) + 2\sigma^2(\omega m X_{t-1} + \alpha_1 X_{t-1}^2) \\
\vdots \\
(\sigma^2+1)(X_{t-q} - 2\alpha_q X_{t-q}) + 2\sigma^2(\omega m X_{t-q} + \alpha_q X_{t-q}^2)
\end{pmatrix}.$$

Suppose the left-hand side is 0 almost surely, then the right-hand side is also 0 almost surely, which can be written as

$$\begin{aligned}
&\xi_0(\sigma^2+1)(m - 2\omega m) + 2\sigma^2\xi_0(\omega m^2 + m\sum_{i=1}^q \alpha_i X_{t-i}) \\
&+ \xi_1(\sigma^2+1)(X_{t-1} - 2\alpha_1 X_{t-1}) + 2\sigma^2\xi_1(\omega m X_{t-1} + \alpha_1 X_{t-1}^2) + M_{t-2} = 0 \ \ a.s.,
\end{aligned}$$

where

$$M_{t-2} = \sum_{k=2}^p \xi_k\Big[(\sigma^2+1)(X_{t-k} - 2\alpha_k X_{t-k}) + 2\sigma^2(\omega m X_{t-k} + \alpha_k X_{t-k}^2)\Big].$$

So the coefficients of the above equation must satisfy

$$\xi_i(\sigma^2+1) = 0, \quad 2\sigma^2\xi_i = 0, \quad i = 0, \ldots, q.$$

For $\sigma^2 > 0$, we must have $\xi_i = 0, i = 0, \ldots, q$. Thus, $\xi = (\xi_0, \xi_1, \ldots, \xi_q)^{\mathrm{T}} = 0$, which completes the proof of the positive definiteness of Σ.

Appendix A.5. Lemmas for the Proof of Theorem 2

Similar to the proof of Theorem 1.2 in Hu (2016) [9], we give some related lemmas for the proof of Theorem 2. According to the derivatives of the quasi-likelihood function, we have

$$\begin{aligned}
\frac{\partial \mu_t(\theta)}{\partial \omega} &= m, \\
\frac{\partial \sigma_t^2(\theta)}{\partial \omega} &= (\sigma^2+1)(m-2\omega m) + 2\sigma^2\left(m^2\omega + m\sum_{i=1}^{q}\alpha_i X_{t-i}\right), \\
&\le (\sigma^2+1)m(1-2\omega_L) + 2\sigma^2\left(m^2\omega_U + m\sum_{i=1}^{q}\alpha_U X_{t-i}\right),
\end{aligned}$$

thus, $E(\frac{\partial \mu_t(\theta)}{\partial \omega})^2 < \infty$ and $E(\frac{\partial \sigma_t^2(\theta)}{\partial \omega})^2 < \infty$. Likewise for the other terms of parameters.

Lemma A1. *Under Assumptions 1 and 2, when $n \to \infty$,*

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial \tilde{l}_t(\theta_0)}{\partial \theta_i} \xrightarrow{d} N(0,\Sigma), \quad \frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 \tilde{l}_t(\theta_0)}{\partial \theta_i \partial \theta_j} \xrightarrow{P} -\Sigma.$$

Proof of Lemma A1. First, we show that

$$n^{-1/2}\sum_{t=1}^{n}\left|\frac{\partial l_t(\theta_0)}{\partial \theta_i} - \frac{\partial \tilde{l}_t(\theta_0)}{\partial \theta_i}\right| \xrightarrow{P} 0, \quad n^{-1}\sum_{t=1}^{n}\left|\frac{\partial^2 l_t(\theta_0)}{\partial \theta_i \partial \theta_j} - \frac{\partial^2 \tilde{l}_t(\theta_0)}{\partial \theta_i \partial \theta_j}\right| \xrightarrow{P} 0.$$

Notice that $\tilde{\mu}_t(\theta)$ and $\tilde{\sigma}_t^2(\theta)$ are stationary approximations of $\mu_t(\theta)$ and $\sigma_t^2(\theta)$, since X_t is stationary and ergodic, using arguments similar to Proposition 2.1.1 in Straumann (2005) [23], for fixed $\theta \in \Theta$, $\tilde{\mu}_t(\theta)$ and $\tilde{\sigma}_t^2(\theta)$, $\mu_t(\theta)$ and $\sigma_t^2(\theta)$ are also stationary and ergodic. Hence, similar to the proof of Lemma A2 in Hu and Andrews (2021) [24], it is easy to have

$$n^{-1/2}\sum_{t=1}^{n}\left|\frac{\partial l_t(\theta_0)}{\partial \theta_i} - \frac{\partial \tilde{l}_t(\theta_0)}{\partial \theta_i}\right| \xrightarrow{P} 0, \quad n^{-1}\sum_{t=1}^{n}\left|\frac{\partial^2 l_t(\theta_0)}{\partial \theta_i \partial \theta_j} - \frac{\partial^2 \tilde{l}_t(\theta_0)}{\partial \theta_i \partial \theta_j}\right| \xrightarrow{P} 0.$$

Therefore, it suffices to show that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial l_t(\theta_0)}{\partial \theta} \xrightarrow{d} N(0,\Sigma), \quad \frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t(\theta_0)}{\partial \theta \partial \theta^{\mathrm{T}}} \xrightarrow{P} -\Sigma.$$

First, we should guarantee that

$$E_{\theta_0}\left\|\frac{\partial l_t(\theta_0)}{\partial \theta}\frac{\partial l_t(\theta_0)}{\partial \theta^{\mathrm{T}}}\right\| < \infty, \quad E_{\theta_0}\left\|\frac{\partial^2 l_t(\theta_0)}{\partial \theta \partial \theta^{\mathrm{T}}}\right\| < \infty. \tag{A1}$$

Now we prove the first part of (A1).

$$E_{\theta_0}\left(\frac{\partial l_t(\theta_0)}{\partial \omega}\right)^2 = E_{\theta_0}\left[\frac{1}{2\sigma_t^4(\theta_0)}\left(\frac{\partial \sigma_t^2(\theta_0)}{\partial \omega}\right)^2 + \frac{1}{\sigma_t^2(\theta_0)}\left(\frac{\partial \mu_t(\theta_0)}{\partial \omega}\right)^2\right] < \infty.$$

Similarly, we can prove other terms, thus, the first part of (A1) holds. The proof of the second part of (A1) is similar, here we omit the details.

Under (A1), $\left\{\frac{\partial l_t(\theta_0)}{\partial \theta}\right\}$ is a martingale difference sequence with respect to $\{\mathscr{F}_t\}$, it follows that at $\theta = \theta_0$, $E_{\theta_0}\left(\frac{\partial l_t(\theta_0)}{\partial \theta}|\mathscr{F}_{t-1}\right) = 0$, so $E_{\theta_0}\left(\frac{\partial l_t(\theta_0)}{\partial \theta}\right) = 0$. Moreover, we have shown that $\Sigma = E_{\theta_0}\left(\frac{\partial l_t(\theta_0)}{\partial \theta}\frac{\partial l_t(\theta_0)}{\partial \theta^{\mathrm{T}}}\right)$ in Section 3.2. Hence $\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial \tilde{l}_t(\theta_0)}{\partial \theta} \xrightarrow{d} N(0,\Sigma)$ holds by the central limit theorem for martingale difference sequence in Billingsley (1961). Similarly, we have $E_{\theta_0}\left(\frac{\partial l_t^2(\theta_0)}{\partial \theta \partial \theta^{\mathrm{T}}}\right) = -\Sigma$.

Under Assumption 1, $\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 \tilde{l}_t(\theta_0)}{\partial \theta_i \partial \theta_j} \xrightarrow{P} -\Sigma$ follows from the ergodic theorem. Thus, Lemma A1 is proved. □

Before showing Lemma A2, we have

$$\widetilde{T}_n(u) \equiv \tilde{l}_n\left(\theta_0 + \frac{u}{\sqrt{n}}\right) - \tilde{l}_n(\theta_0), \quad u \in \mathbb{R}^{q+1},$$

we use $\widetilde{T}_n$ to derive the asymptotic distribution of $\hat{\theta}_n$.

For any $u \in \mathbb{R}^{q+1}$, the Taylor series expansion of $\widetilde{T}_n(u)$ at θ_0 is

$$\widetilde{T}_n(u) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} u^{\mathrm{T}}\frac{\partial \tilde{l}_t(\theta_0)}{\partial \theta} + \frac{1}{2n}\sum_{t=1}^{n} u^{\mathrm{T}}\frac{\partial^2 \tilde{l}_t(\theta_0)}{\partial \theta \partial \theta^{\mathrm{T}}}u + \frac{1}{2n}\sum_{t=1}^{n} u^{\mathrm{T}}\left[\frac{\partial^2 \tilde{l}_t(\theta^*)}{\partial \theta \partial \theta^{\mathrm{T}}} - \frac{\partial^2 \tilde{l}_t(\theta_0)}{\partial \theta \partial \theta^{\mathrm{T}}}\right]u, \quad \text{(A2)}$$

where $\theta^* = \theta_n^*(u)$ is on the line segment connecting θ_0 and $\theta_0 + \frac{u}{\sqrt{n}}$. For Euclidean distance $\|\cdot\|$ and any compact set $K \subset \mathbb{R}^{q+1}$, $\sup_{u\in K} \| \theta^* - \theta_0 \| \to 0$, as $n \to \infty$.

Lemma A2. *Under Assumptions 1 and 2, when $n \to \infty$,*

$$\frac{1}{n}\sum_{t=1}^{n}\left[\frac{\partial^2 \tilde{l}_t(\theta^*)}{\partial \theta \partial \theta^{\mathrm{T}}} - \frac{\partial^2 \tilde{l}_t(\theta_0)}{\partial \theta \partial \theta^{\mathrm{T}}}\right] \xrightarrow{P} 0.$$

Proof. Similar to Lemma A1, for any $1 \leq i,j \leq q+1$,

$$\frac{1}{n}\sum_{t=1}^{n}\left\|\frac{\partial^2 l_t(\theta_0)}{\partial \theta_i \partial \theta_j} - \frac{\partial^2 \tilde{l}_t(\theta_0)}{\partial \theta_i \partial \theta_j}\right\| \xrightarrow{P} 0. \quad \text{(A3)}$$

Using arguments similar to the proof of Theorem 2.2 of Francq and Zakoïan (2004) [20], it suffices to show

$$\frac{1}{n}\sum_{t=1}^{n}\left[\frac{\partial^2 l_t(\theta^*)}{\partial \theta_i \partial \theta_j} - \frac{\partial^2 l_t(\theta_0)}{\partial \theta_i \partial \theta_j}\right] \xrightarrow{P} 0. \quad \text{(A4)}$$

By the Taylor series expansion, we have

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t(\theta^*)}{\partial \theta_i \partial \theta_j} = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t(\theta_0)}{\partial \theta_i \partial \theta_j} + \frac{1}{n}\sum_{t=1}^{n}\frac{\partial}{\partial \theta_k}\left(\frac{\partial^2 l_t(\theta^{**})}{\partial \theta_i \partial \theta_j}\right)(\theta^* - \theta_0),$$

here $\theta^{**} = \theta_n^{**}(u)$ is on the line segment connecting θ_0 and θ^*, such that for any u, we have $\| \theta^{**} - \theta_0 \| \to 0$ *a.s.*, $n \to \infty$.

From (A2), $\| \theta^* - \theta_0 \| \to 0$ *a.s*, so

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\partial}{\partial \theta_k}\left(\frac{\partial^2 l_t(\theta^{**})}{\partial \theta_i \partial \theta_j}\right)(\theta^* - \theta_0) \to 0, \ a.s.$$

if

$$\limsup_{n\to\infty}\left\|\frac{1}{n}\sum_{t=1}^{n}\frac{\partial}{\partial\theta_k}\left(\frac{\partial^2 l_t(\theta^{**})}{\partial\theta_i\partial\theta_j}\right)\right\|<\infty,\ a.s. \tag{A5}$$

Then we have

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t(\theta^*)}{\partial\theta_i\partial\theta_j}\to\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t(\theta_0)}{\partial\theta_i\partial\theta_j}\ a.s.,$$

so (A4) is proved.

Using arguments similar to the proof of Theorem 2.2 of Francq and Zakoïan (2004) [20], there exists a neighborhood $\nu(\theta_0)$, that

$$E_{\theta_0}\sup_{\theta\in\nu(\theta_0)\cap\Theta}\left\|\frac{\partial}{\partial\theta_k}\left(\frac{\partial^2 l_t(\theta)}{\partial\theta_i\partial\theta_j}\right)\right\|<\infty,\quad \sup_{\theta\in\nu(\theta_0)}\left\|\frac{1}{n}\sum_{t=1}^{n}\left[\frac{\partial^2 l_t(\theta)}{\partial\theta_i\partial\theta_j}-\frac{\partial^2\tilde{l}_t(\theta)}{\partial\theta_i\partial\theta_j}\right]\right\|\xrightarrow{P}0. \tag{A6}$$

Therefore, by the ergodic theorem, we have

$$\begin{aligned}\limsup_{n\to\infty}\left\|\frac{1}{n}\sum_{t=1}^{n}\frac{\partial}{\partial\theta_k}\left(\frac{\partial^2 l_t(\theta^{**})}{\partial\theta_i\partial\theta_j}\right)\right\| &\le \limsup_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\sup_{\theta\in\nu(\theta_0)\cap\Theta}\left\|\frac{\partial}{\partial\theta_k}\left(\frac{\partial^2 l_t(\theta)}{\partial\theta_i\partial\theta_j}\right)\right\| \\ &= E_{\theta_0}\sup_{\theta\in\nu(\theta_0)\cap\Theta}\left\|\frac{\partial}{\partial\theta_k}\left(\frac{\partial^2 l_t(\theta)}{\partial\theta_i\partial\theta_j}\right)\right\|<\infty,\end{aligned}$$

so (A5) is proved.

In view of (A3), (A4) and (A6), we obtain Lemma A2. □

Lemma A3. *For any compact set $K\in\mathbb{R}^{q+1}$ and any $\varepsilon>0$,*

$$\lim_{\sigma\to 0}\limsup_{n\to\infty}\mathbb{P}\left(\sup_{u,v\in K,\|u-v\|<\sigma}\left|\widetilde{T}_n(u)-\widetilde{T}_n(v)\right|\ge\varepsilon\right)=0.$$

Proof. For any $\epsilon>0$, by (A2) we have

$$\begin{aligned}&\lim_{\delta\to 0}\limsup_{n\to\infty}\mathbb{P}\left(\sup_{u,v\in K,\|u-v\|<\delta}\left|\widetilde{T}_n(u)-\widetilde{T}_n(v)\right|\ge\varepsilon\right)\\ &\le\lim_{\delta\to 0}\limsup_{n\to\infty}\mathbb{P}\left(\sup_{u,v\in K,\|u-v\|<\delta}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(u-v)^{\mathrm{T}}\frac{\partial\tilde{l}_t(\theta_0)}{\partial\theta}\right|\ge\frac{\epsilon}{3}\right)\\ &+\lim_{\delta\to 0}\limsup_{n\to\infty}\mathbb{P}\left(\sup_{u,v\in K,\|u-v\|<\delta}\left|\frac{1}{n}\left(\sum_{t=1}^{n}u^{\mathrm{T}}\frac{\partial^2\tilde{l}_t(\theta_0)}{\partial\theta\partial\theta^{\mathrm{T}}}u-\sum_{t=1}^{n}v^{\mathrm{T}}\frac{\partial^2\tilde{l}_t(\theta_0)}{\partial\theta\partial\theta^{\mathrm{T}}}v\right)\right|\ge\frac{2\epsilon}{3}\right)\\ &+\lim_{\delta\to 0}\limsup_{n\to\infty}\mathbb{P}\left\{\sup_{u,v\in K,\|u-v\|<\delta}\left|\frac{1}{n}\left[\sum_{t=1}^{n}u^{\mathrm{T}}\left(\frac{\partial^2\tilde{l}_t(\theta^*)}{\partial\theta\partial\theta^{\mathrm{T}}}-\frac{\partial^2\tilde{l}_t(\theta_0)}{\partial\theta\partial\theta^{\mathrm{T}}}\right)u\right.\right.\right.\\ &\left.\left.\left.-\sum_{t=1}^{n}v^{\mathrm{T}}\left(\frac{\partial^2\tilde{l}_t(\theta^*)}{\partial\theta\partial\theta^{\mathrm{T}}}-\frac{\partial^2\tilde{l}_t(\theta_0)}{\partial\theta\partial\theta^{\mathrm{T}}}\right)v\right]\right|\ge\frac{2\epsilon}{3}\right\}.\end{aligned}$$

Because of Lemmas A1 and A2, we have

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial\tilde{l}_t(\theta_0)}{\partial\theta}=O_p(1),\quad\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2\tilde{l}_t(\theta_0)}{\partial\theta\partial\theta^{\mathrm{T}}}=O_p(1),$$

$$\frac{1}{n}\sum_{t=1}^{n}\left[\frac{\partial^2\tilde{l}_t(\theta^*)}{\partial\theta\partial\theta^{\mathrm{T}}}-\frac{\partial^2\tilde{l}_t(\theta_0)}{\partial\theta\partial\theta^{\mathrm{T}}}\right]=o_p(1),$$

where $O_p(1)$ and $o_p(1)$ for vector and matrix means $O_p(1)$ and $o_p(1)$ for every elements. By the compactness of K, we have

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\mathbb{P}\left(\sup_{u,v\in K,\|u-v\|<\delta}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(u-v)^{\mathrm{T}}\frac{\partial\tilde{l}_t(\theta_0)}{\partial\theta}\right|\geq\frac{\epsilon}{3}\right)=0,$$

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\mathbb{P}\left(\sup_{u,v\in K,\|u-v\|<\delta}\left|\frac{1}{n}\left(\sum_{t=1}^{n}u^{\mathrm{T}}\frac{\partial^2\tilde{l}_t(\theta_0)}{\partial\theta\partial\theta^{\mathrm{T}}}u-\sum_{t=1}^{n}v^{\mathrm{T}}\frac{\partial^2\tilde{l}_t(\theta_0)}{\partial\theta\partial\theta^{\mathrm{T}}}v\right)\right|\geq\frac{2\epsilon}{3}\right)=0,$$

$$\begin{aligned}\lim_{\delta\to 0}\limsup_{n\to\infty}\mathbb{P}\Bigg\{\sup_{u,v\in K,\|u-v\|<\delta}\Bigg|\frac{1}{n}\Bigg[&\sum_{t=1}^{n}u^{\mathrm{T}}\left(\frac{\partial^2\tilde{l}_t(\theta^*)}{\partial\theta\partial\theta^{\mathrm{T}}}-\frac{\partial^2\tilde{l}_t(\theta_0)}{\partial\theta\partial\theta^{\mathrm{T}}}\right)u\\&-\sum_{t=1}^{n}v^{\mathrm{T}}\left(\frac{\partial^2\tilde{l}_t(\theta^*)}{\partial\theta\partial\theta^{\mathrm{T}}}-\frac{\partial^2\tilde{l}_t(\theta_0)}{\partial\theta\partial\theta^{\mathrm{T}}}\right)v\Bigg]\Bigg|\geq\frac{2\epsilon}{3}\Bigg\}=0,\end{aligned}$$

which completes our proof. □

Appendix A.6. Proof of Theorem 2

Proof. Let $T(u)=u^{\mathrm{T}}N(0,\Sigma)-\frac{1}{2}u^{\mathrm{T}}\Sigma u$, where N is a multivariate Gaussian random vector with mean 0 and covariance matrix Σ. By Lemmas A1 and A2, for any $u\in\mathbb{R}^{q+1}$ and $n\to\infty$, the finite dimensional distributions of $\widetilde{T}_n$ converge to those of T: $\widetilde{T}_n(u)\to T(u)$.

By Lemma A3, similar to Hu (2016) [9], $\widetilde{T}_n(u)$ is tight on the continuous function space $C(K)$ for any compact set $K\in\mathbb{R}^{q+1}$. So by Theorem 7.1 in Billingsley (1999) [25], $\widetilde{T}_n(\cdot)\to T(\cdot)$ on $C(K)$. From Appendix A.4 and Lemma A1, Σ is positive finite and invertible, meanwhile, $T(\cdot)$ is concave with the unique maximum $\Sigma^{-1}N(0,\Sigma)=N(0,\Sigma^{-1})$. $\widetilde{T}_n(\cdot)$ is maximized at $u_{\max}=\sqrt{n}(\hat{\theta}_n-\theta_0)$. Thus, the result of Theorem 2 can be proved by the proof of Lemma 2.2 and Remark 1 in Davis et al. (1992) [26]. □

References

1. McKenzie, E. Some simple models for discrete variate time series. *Water Resour. Bull.* **1985**, *21*, 645–650. [CrossRef]
2. Al-Osh, M.A.; Alzaid, A.A. First-order integer-valued autoregressive (INAR(1)) process. *J. Time Ser. Anal.* **1987**, *8*, 261–275. [CrossRef]
3. Al-Osh, M.A.; Alzaid, A.A. Integer-valued moving average (INMA) process. *Stat. Pap.* **1988**, *29*, 281–300. [CrossRef]
4. McKenzie, E. Some ARMA models for dependent sequences of Poisson counts. *Adv. Appl. Probab.* **1988**, *20*, 822–835. [CrossRef]
5. Ferland, R.; Latour, A.; Oraichi, D. Integer-valued GARCH process. *J. Time Ser. Anal.* **2006**, *27*, 923–942. [CrossRef]
6. Steutel, F.W.; van Harn, K. Discrete analogues of self-decomposability and stability. *Ann. Probab.* **1979**, *7*, 893–899. [CrossRef]
7. Qian, L.; Zhu, F. A new minification integer-valued autoregressive process driven by explanatory variables. *Aust. N. Z. J. Stat.* **2022**, *64*, 478–494. [CrossRef]
8. Huang, J.; Zhu, F.; Deng, D. A mixed generalized Poisson INAR model with applications. *J. Stat. Comput. Simul.* 2023, *forthcoming*. [CrossRef]
9. Hu, X. Volatility Estimation for Integer-Valued Financial Time Series. Ph.D. Thesis, Northwestern University, Evanston, IL, USA, 2016. [CrossRef]
10. Liu, M.; Zhu, F.; Zhu, K. Modeling normalcy-dominant ordinal time series: An application to air quality level. *J. Time Ser. Anal.* **2022**, *43*, 460–478. [CrossRef]
11. Weiß, C.H.; Zhu, F.; Hoshiyar, A. Softplus INGARCH models. *Stat. Sin.* **2022**, *32*, 1099–1120.
12. Weiß, C.H. *An Introduction to Discrete-Valued Time Series*; John Wiley & Sons: Chichester, UK, 2018. [CrossRef]
13. Davis, R.A.; Fokianos, K.; Holan, S.H.; Joe, H.; Livsey, J.; Lund, R.; Pipiras, V.; Ravishanker, N. Count time series: A methodological review. *J. Am. Stat. Assoc.* **2021**, *116*, 1533–1547. [CrossRef]
14. Aknouche, A.; Scotto, M. A multiplicative Thinning-Based Integer-Valued GARCH Model. Working Paper. 2022. Available online: https://mpra.ub.uni-muenchen.de/112475 (accessed on 17 January 2023).
15. Daniels, H.E. Saddlepoint approximations in statistics. *Ann. Math. Stat.* **1954**, *25*, 631–650. [CrossRef]
16. Field, C.; Ronchetti, E. Small sample asymptotics. In *Institute of Mathematical Statistics Lecture Notes—Monograph Series*; Institute of Mathematical Statistics: Hayward, CA, USA, 1990.
17. Jensen, J.L. *Saddlepoint Approximations*; Oxford University Press: Oxford, UK, 1995. [CrossRef]
18. Butler, R.W. *Saddlepoint Approximations with Applications*; Cambridge University Press: Cambridge, UK, 2007.

19. Pedeli, X.; Davison, A.C.; Fokianos, K. Likelihood estimation for the INAR(p) model by saddlepoint approximation. *J. Am. Stat. Assoc.* **2015**, *110*, 1229–1238.
20. Francq, C.; Zakoïan, J.M. Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* **2004**, *10*, 605–637.
21. Aknouche, A.; Francq, C. Two-stage weighted least squares estimator of the conditional mean of observation-driven time series models. *J. Econom.* 2023, *forthcoming*. [CrossRef]
22. Xu, Y.; Zhu, F. A new GJR-GARCH model for $\mathbb{Z}$-valued time series. *J. Time Ser. Anal.* **2022**, *43*, 490–500. [CrossRef]
23. Straumann, D. *Estimation in Conditionally Heteroscedastic Time Series Models*; Springer: Berlin, Germany, 2005. [CrossRef]
24. Hu, X.; Andrews, B. Integer-valued asymmetric GARCH modeling. *J. Time Ser. Anal.* **2021**, *42*, 737–751. [CrossRef]
25. Billingsley, P. *Convergence of Probability Measures*, 2nd ed.; Wiley: New York, NY, USA, 1999.
26. Davis, R.A.; Knight, K.; Liu, J. M-estimation for autoregressions with infinite variance. *Stoch. Process. Their Appl.* **1992**, *40*, 145–180. [CrossRef]