Scientific
Research
Publishing

Efficiency of Some Estimators for a Generalized Poisson Autoregressive Process of Order 1

Louis G. Doray[1], Andrew Luong[2], El-Halla Najem[1]

[1]Département de Mathématiques et de Statistique, Université de Montréal, Montréal, Canada
[2]École d'Actuariat, Université Laval, Québec, Canada
Email: doray@dms.umontreal.ca, aluong@act.ulaval.ca, najem@dms.umontreal.ca

Abstract

Various models have been proposed in the literature to study non-negative integer-valued time series. In this paper, we study estimators for the generalized Poisson autoregressive process of order 1, a model developed by Alzaid and Al-Osh [1]. We compare three estimation methods, the methods of moments, quasi-likelihood and conditional maximum likelihood and study their asymptotic properties. To compare the bias of the estimators in small samples, we perform a simulation study for various parameter values. Using the theory of estimating equations, we obtain expressions for the variance-covariance matrices of those three estimators, and we compare their asymptotic efficiency. Finally, we apply the methods derived in the paper to a real time series.

Keywords

Discrete Time Series, Autoregressive Process, Moment Estimator, Quasi-Likelihood, Efficiency, Generalized Poisson, Quasi Binomial Distribution

1. Introduction

Time series are used to model various phenomena measured over time. Successive observations are often correlated, since they may depend on some common external factors, but which remain unknown to the analyst. In this case, autoregressive models will be useful to model this dependence.

In some situations, we might be interested in the number of events which occur during a certain period of time. Such observations will necessarily be non-negative and integer-valued. Models which have been used for sequences of dependent discrete random variables include the Poisson autoregressive process of order 1, denoted

$PAR(1)$, introduced by Al-Osh and Alzaid [2] and the generalized Poisson autoregressive process of order 1, denoted $GPAR(1)$ (see Alzaid and Al-Osh [1]). The $PAR(1)$ process, a stationary process with Poisson marginal distributions, is a special case of the $GPAR(1)$.

The paper is organized as follows. In Section 2, for completeness, we review some properties of the generalized Poisson autoregressive process of order 1. In Section 3, we derive the expressions for the moments estimators, the quasi-likelihood and the maximum likelihood estimators of the 3 parameters of the $GPAR(1)$. These methods have appeared in the literature (see Al-Nachawati, Alwasel and Alzaid [3] for the quasilikelihood and moments method and Brännäs [4] for likelihood methods). However, asymptotic properties such as efficiencies of these methods are not discussed in those papers. In this paper (Sections 4 and 5), we study properties of these estimators such as bias and asymptotic efficiency. The last section reanalyzes a real-data example which can be modelled with a $GPAR(1)$ process, where testing is discussed.

We hope that with this study, practitioners will have more information to select one estimation method versus another one and to perform tests concerning values of the parameters.

2 *GPAR*(1) Process

To define the $GPAR(1)$ process, we need first to review the generalized Poisson and the quasi-binomial distributions.

A random variable *X* has a generalized Poisson distribution with parameters λ and θ, denoted $GP(\lambda,\theta)$, if its probability mass function (pmf) is defined by

$$P[X=x]=\begin{cases}\lambda(\lambda+\theta x)^{x-1}\,\mathrm{e}^{-(\lambda+\theta x)}/x! & \text{if } x=0,1,2,\dots \\ 0 & \text{for } x>m, \text{ when } \theta<0,\end{cases}$$

where $\lambda>0$, $\max(-1,-\lambda/m)<\theta\leq1$ and $m(\geq4)$ is the greatest positive integer for which $\lambda+m\theta>0$ when θ is negative. Note that, for $\theta=0$, the random variable *X* becomes a Poisson (λ) distribution. In this paper, we will restrict ourselves to the case where $\theta\geq0$.

Consul [5] has shown that the expected value μ and variance σ^2 of *X* are given, when $0\leq\theta<1$, by

$$\mu=\frac{\lambda}{1-\theta}\quad\text{and}\quad\sigma^2=\frac{\lambda}{(1-\theta)^3},$$

so that, for positive values of θ, we have overdispersion (*i.e.* $\mathrm{Var}[X]>E[X]$).

The sum $X+Y$ of two independent random variables *X* and *Y* with $GP(\lambda_1,\theta)$ and $GP(\lambda_2,\theta)$ distributions, also has a *GP* distribution, with parameters $(\lambda_1+\lambda_2,\theta)$. Ambagaspitiya and Balakrishnan [6] have derived the recurrence formula for the probability function of the compound generalized Poisson distribution, used in risk theory.

A non-negative integer-valued random variable *X* has a quasi-binomial distribution, denoted $QB(p,\theta,n)$, if its pmf is given by

$$P[X=x]=\frac{pq\dbinom{n}{x}(p+x\theta)^{x-1}\left[q+(n-x)\theta\right]^{n-x-1}}{(1+n\theta)^{n-1}},\quad x=0,1,2,\cdots,n,$$

where $0<p<1, q=1-p$ and θ is such that $n\theta<\min(p,q)$. Its mean, equal to pn, is independent of the parameter θ.

The following proposition, proved in Alzaid and Al-Osh [1], shows the relation between the *QB* and *GP* distributions.

Proposition 1: If *X* and $S(n)$ are two independent random variables with $GP(\lambda,\theta)$ and $QB(p,\theta/\lambda,n)$ distributions, then $S(X)$ follows a $GP(p\lambda,\theta)$ distribution.

The $GPAR(1)$ process generalizes the $PAR(1)$ process introduced by Al-Osh and Alzaid [2]. The $PAR(1)$ model, where $\mathrm{Var}[X_t]=E[X_t]$, has been used to model time series in various fields, for example in insurance for short-term workers' compensation because of work-related injuries (Freeland and McCabe [7]) and in medicine for the incidence of infectious diseases (Cardinal, Roy and Lambert [8]).

In practice, many integer-valued series will often exhibit overdispersion, (*i.e.* $\mathrm{Var}[X_t]$ is greater than

$E[X_t]$). The $PAR(1)$ model would therefore not be appropriate for those time series. In cases where the extra variation can be explained in a deterministic way, adding regressors would be adequate (see Freeland and McCabe [7]), but where the extra variation is of a stochastic nature, the $GPAR(1)$ model could be used for modelling overdispersed time series.

The $GPAR(1)$ model, introduced by Alzaid and Al-Osh [1], is defined as

$$X_t = S_t(X_{t-1}) + \varepsilon_t, \quad t = 1, 2, \cdots \tag{1}$$

where

1) $\{S_t(\bullet), t = 1, 2, \cdots\}$ is a sequence of iid random variables with a $QB(p, \theta/\lambda, \bullet)$ distribution.
2) $\{\varepsilon_t\}$ is a sequence of iid random variables with a $GP(q\lambda, \theta)$ distribution.
3) These two sequences are independent of each other.
4) X_0 has a $GP(\lambda, \theta)$ distribution independent of $\{\varepsilon_t\}$ and $\{S_t(\bullet)\}$.

Proposition 2: The $GPAR(1)$ process $\{X_t\}$ has a GP marginal distribution.

Proof: See Alzaid and Al-Osh [1]. The $GPAR(1)$ process is obtained from the $QB(p, \theta/\lambda, \bullet)$ and $GP(q\lambda, \theta)$ distributions, and not from the $QB(p, \theta, \bullet)$ and $GP(q\lambda, \theta)$ distributions, as stated in Al-Nachawati *et al.* [3].

The autocorrelation function (acf) of the $GPAR(1)$ process $\{X_t\}$ is equal to

$$\rho_X(k) = \text{Corr}[X_t, X_{t-k}] = p^k, \quad \text{for } k = 0, 1, 2, \cdots.$$

The acf of this process is the same as that of an $AR(1)$ process except that it is always non-negative, since $p \in (0, 1)$. The partial autocorrelation function (pacf) of the $GPAR(1)$ process is equal to

$$\phi_{kk} = \begin{cases} p & \text{if } k = 1 \\ 0 & \text{if } k = 2, 3, \cdots \end{cases}$$

The sample acf and pacf will be useful to identify the $GPAR(1)$ model from an observed time series.

3. Estimation of the Parameters

Estimating the parameters in a $GPAR(1)$ process will present some challenges, since the conditional distribution of X_t, given $X_{t-1} = x_{t-1}$, is the convolution of a $QB(p, \theta/\lambda, x_{t-1})$ and a $GP(q\lambda, \theta)$ distribution.

In this section, we will review three estimation methods for the parameter vector $\Theta = (p, \lambda, \theta)$ of the $GPAR(1)$ process, the methods of moments, quasi-likelihood and conditional maximum likelihood. These methods have been proposed in the literature, see for example, Al-Nachawati *et al.* [3] or Brännäs [4]. However, less emphasis is placed on their asymptotic properties, such as efficiency. In Section 4, we study the bias of these estimators, and in Section 5 their efficiency.

3.1. Method of Moments or Yule-Walker

The first autocovariance of the $GPAR(1)$ process is equal to

$$\text{Cov}[X_{t+1}, X_t] = p\text{Var}[X_0]. \tag{2}$$

By taking the expected value of both sides of the equation given in (1), we find $E[X_t] = E[S(X_{t-1})] + \mu_{\varepsilon_t}$. Since

$$E[S(X_{t-1})] = E_{X_{t-1}}\{E[S(X_{t-1}) \mid X_{t-1} = x_{t-1}]\} = E[pX_{t-1}],$$

we obtain $E[X_t] = pE[X_{t-1}] + \dfrac{q\lambda}{1-\theta}$. \hfill (3)

We also know that

$$\text{Var}[X_t] = \frac{\lambda}{(1-\theta)^3}. \tag{4}$$

From the observations $x_1, x_2, \cdots, x_n$, we estimate the means $E[X_t]$, $E[X_{t-1}]$, the variance $\text{Var}[X_t]$ and the autocovariance $\text{Cov}[X_{t+1}, X_t]$ by their sample analogs

$$\overline{x} = \frac{1}{n}\sum_{t=1}^{n} x_t,$$

$$\overline{x}_0 = \frac{1}{n-1}\sum_{t=1}^{n-1} x_t,$$

$$\mathrm{Var}\left[X_t\right] = \frac{1}{n}\sum_{t=1}^{n}\left(x_t - \overline{x}\right)^2,$$

$$\mathrm{Cov}\left[X_{t+1}, X_t\right] = \frac{1}{n}\sum_{t=1}^{n-1}\left(x_t - \overline{x}\right)\left(x_{t+1} - \overline{x}\right).$$

Solving the system of Equations (2), (3), (4) with $E\left[X_t\right]$, $E\left[X_{t-1}\right]$, $\mathrm{Var}\left[X_t\right]$ and $\mathrm{Cov}\left[X_{t+1}, X_t\right]$ replaced by their sample values, we obtain the moments estimators $\overline{\Theta} = \left(\overline{p}, \overline{\lambda}, \overline{\theta}\right)$ of parameter vector $\Theta = \left(p, \lambda, \theta\right)$,

$$\overline{p} = \frac{\sum_{t=1}^{n-1}\left(x_t - \overline{x}\right)\left(x_{t+1} - \overline{x}\right)}{\sum_{t=1}^{n}\left(x_t - \overline{x}\right)^2}$$

$$\overline{\lambda} = \sqrt{\frac{n\left(\overline{x} - \overline{p}\,\overline{x}_0\right)^3}{\overline{q}^3 \sum_{t=1}^{n}\left(x_t - \overline{x}\right)^2}}$$

$$\overline{\theta} = 1 - \frac{\overline{\lambda}\,\overline{q}}{\left(\overline{x} - \overline{p}\,\overline{x}_0\right)}$$

where $\overline{q} = 1 - \overline{p}$.

We have corrected here misprints in the formulas for the moment estimators of the parameters λ and θ given by Al-Nachawati *et al.* [3].

3.2. Quasi-Likelihood Method

This method, proposed initially by Whittle [9], replaces the true likelihood by the one which assumes that the observations come from a normal distribution with the same conditional mean and variance. Al-Nachawati *et al.* [3] obtained the quasi-likelihood estimators $\tilde{\Theta} = \left(\tilde{p}, \tilde{\lambda}, \tilde{\theta}\right)$ by maximizing

$$L\left(p, \lambda, \theta\right) = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\sigma_t^2}}\, e^{-\left(x_t - \mu_t\right)^2 / 2\sigma_t^2}$$

where μ_t and σ_t^2 are given by

$$\begin{aligned}
\mu_t &= E\left[X_t \mid X_{t-1} = x_{t-1}\right] \\
&= E\left[S\left(x_{t-1}\right)\right] + \mu_{\varepsilon_t} \\
&= p x_{t-1} + \lambda q / \left(1 - \theta\right),
\end{aligned}$$

and

$$\begin{aligned}
\sigma_t^2 &= \mathrm{Var}\left[X_t \mid X_{t-1} = x_{t-1}\right] \\
&= \mathrm{Var}\left[S_t\left(X_{t-1}\right) \mid X_{t-1} = x_{t-1}\right] + \mathrm{Var}\left[\varepsilon_t\right] \\
&= pq\left\{x_{t-1}^2 - \sum_{j=1}^{x_{t-1}-1} \frac{x_{t-1}!\left(\theta/\lambda\right)^{j-1}}{\left(x_{t-1}-j-1\right)!\left[1 + x_{t-1}\left(\theta/\lambda\right)\right]^j}\right\} + \lambda q / \left(1 - \theta\right)^3.
\end{aligned}$$

We have used the expression in Shenton [10] for the formula of the variance of a quasi-binomial distribution,

which is a bit different from the one given in Al-Nachawati *et al.* [3]. Since the $GPAR(1)$ process is restricted to non-negative integers and therefore not symmetrical, one might suspect that the estimators are less efficient than the maximum likelihood estimators, which is indeed the case (see Section 5 for numerical results).

3.3. Conditional Maximum Likelihood Method

To obtain the conditional maximum likelihood estimators (MLE's) $\hat{\Theta} = (\hat{p}, \hat{\lambda}, \hat{\theta})$, we need the conditional distribution of $(X_t \mid X_{t-1} = x_{t-1})$, which is the convolution of a $QB(p, \theta/\lambda, x_{t-1})$ distribution and a $GP(q\lambda, \theta)$ distribution. Given the observations $x_0, x_1, x_2, \cdots, x_n$, we have to maximize the function

$$
\begin{aligned}
L(p, \lambda, \theta) &= \prod_{t=1}^{n} P(X_t = x_t \mid X_{t-1} = x_{t-1}) \\
&= \prod_{t=1}^{n} \left\{ \sum_{r=0}^{\min(x_t, x_{t-1})} P[S_t(x_{t-1}) = r] P[\varepsilon_t = x_t - r] \right\} \\
&= \prod_{t=1}^{n} \left\{ \sum_{r=0}^{\min(x_t, x_{t-1})} \binom{x_{t-1}}{r} \frac{pq\lambda}{\lambda + \theta x_{t-1}} \left[\frac{p\lambda + r\theta}{\lambda + \theta x_{t-1}} \right]^{r-1} \left[\frac{q\lambda + (x_{t-1} - r)\theta}{\lambda + \theta x_{t-1}} \right]^{x_{t-1} - r - 1} \right. \\
&\qquad \left. \times \frac{\lambda q [\lambda q + \theta(x_t - r)]^{x_t - r - 1} \mathrm{e}^{-\lambda q - \theta(x_t - r)}}{(x_t - r)!} \right\}.
\end{aligned}
$$

We will work with the loglikelihood function $l(p, \lambda, \theta)$, equal to $\ln L(p, \lambda, \theta)$, which will have to be maximized numerically to obtain the MLE $\hat{\Theta}$.

Under normal regularity conditions, using likelihood theory (see Gouriéroux and Monfort [11] or Hamilton [12]), the vector $\hat{\Theta}$ has an asymptotic multinormal distribution, *i.e.*

$$
\sqrt{n}(\hat{\Theta} - \Theta) \xrightarrow{\mathcal{L}} N(0, I^{-1}(\Theta)),
$$

where $\xrightarrow{\mathcal{L}}$ denotes convergence in law, 0 is the vector of zeros of dimension 3, and

$$
I(\Theta) = -\left(E\left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} l(\Theta) \right] \right) \text{ is Fisher's expected information matrix, of dimension } 3 \times 3.
$$

4. Bias of Estimators

With simulations, we will study the bias of the moments estimators and the MLE's. Setting the values of the 3 parameters to those in **Table 1**, two series of 50 and 200 observations were generated from model (1) in C++. This experiment was repeated 200 times.

For each series, the moments estimators were calculated, as well as their average, and the bias. The conditional MLE's were calculated using the iterative Downhill Simplex method (see Press, Teukolsky, Vetterling and Flannery [13]), which does not require the calculation of the derivatives of the function to be maximized. As initial values, we used the moments estimators. The results of the simulations appear in **Figures 1-3**.

From **Figures 1-3**, we see that the bias of the MLE's is smaller than that of the moments estimators, and that

Table 1. Values of parameters.

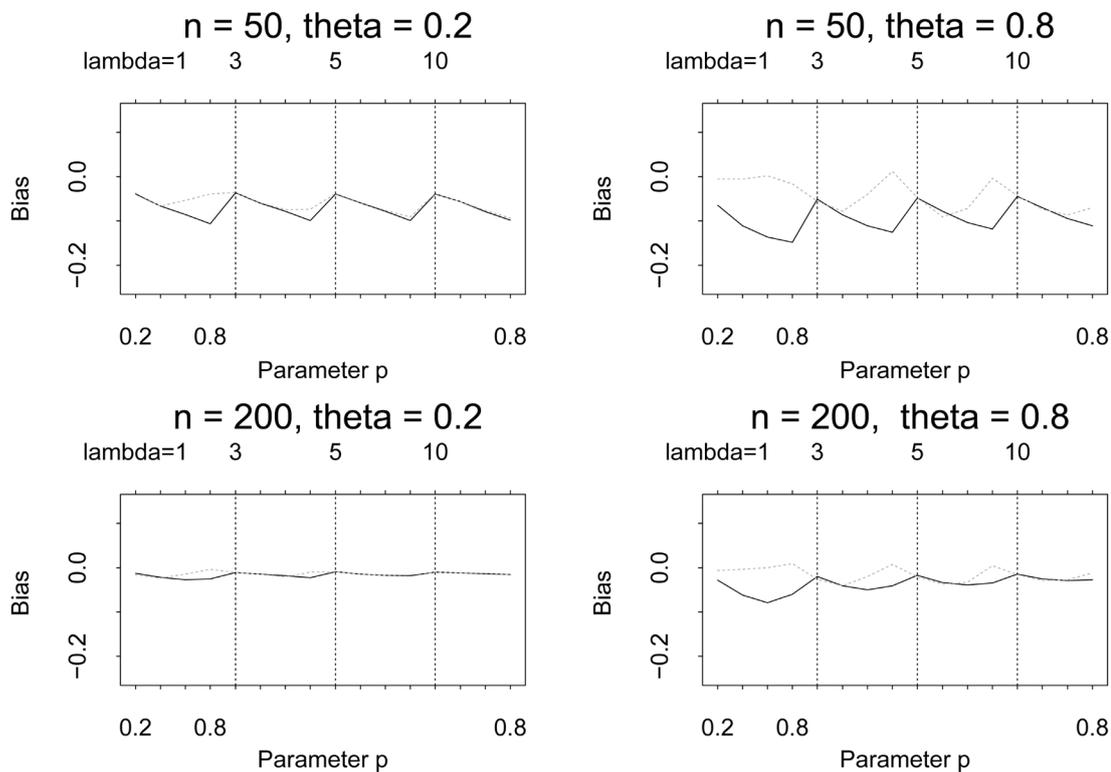| p	λ	θ
0.2	1	0.2
0.4	3	0.4
0.6	5	0.6
0.8	10	0.8

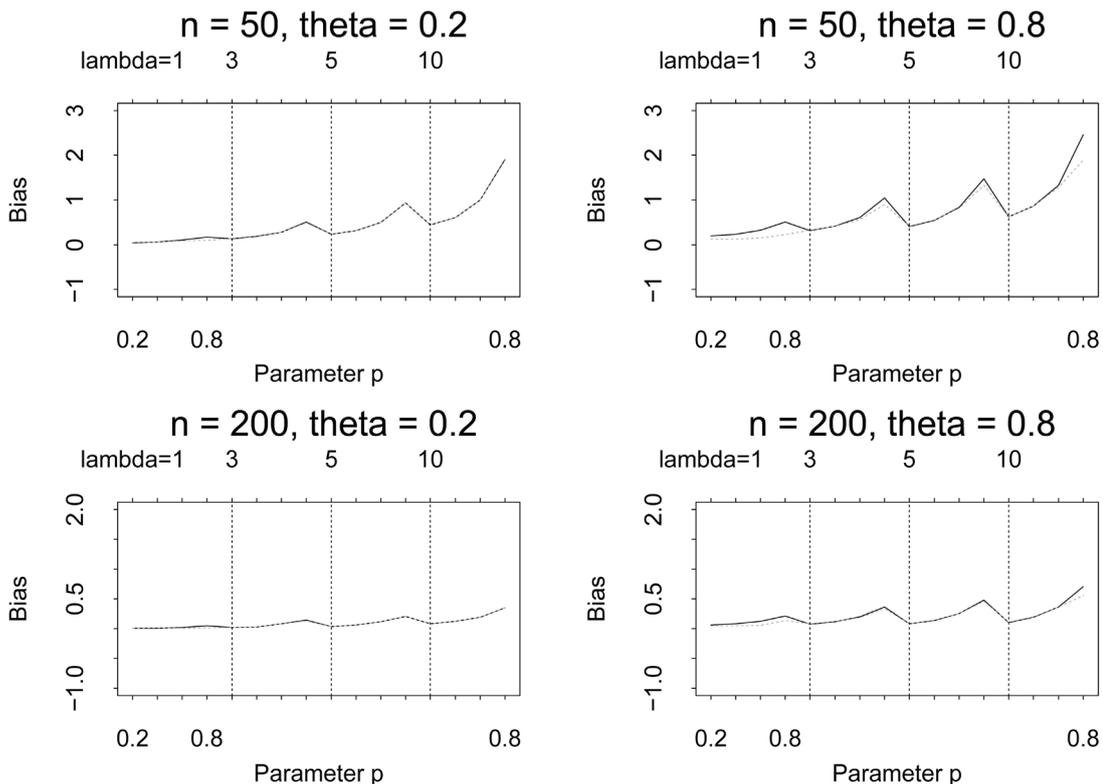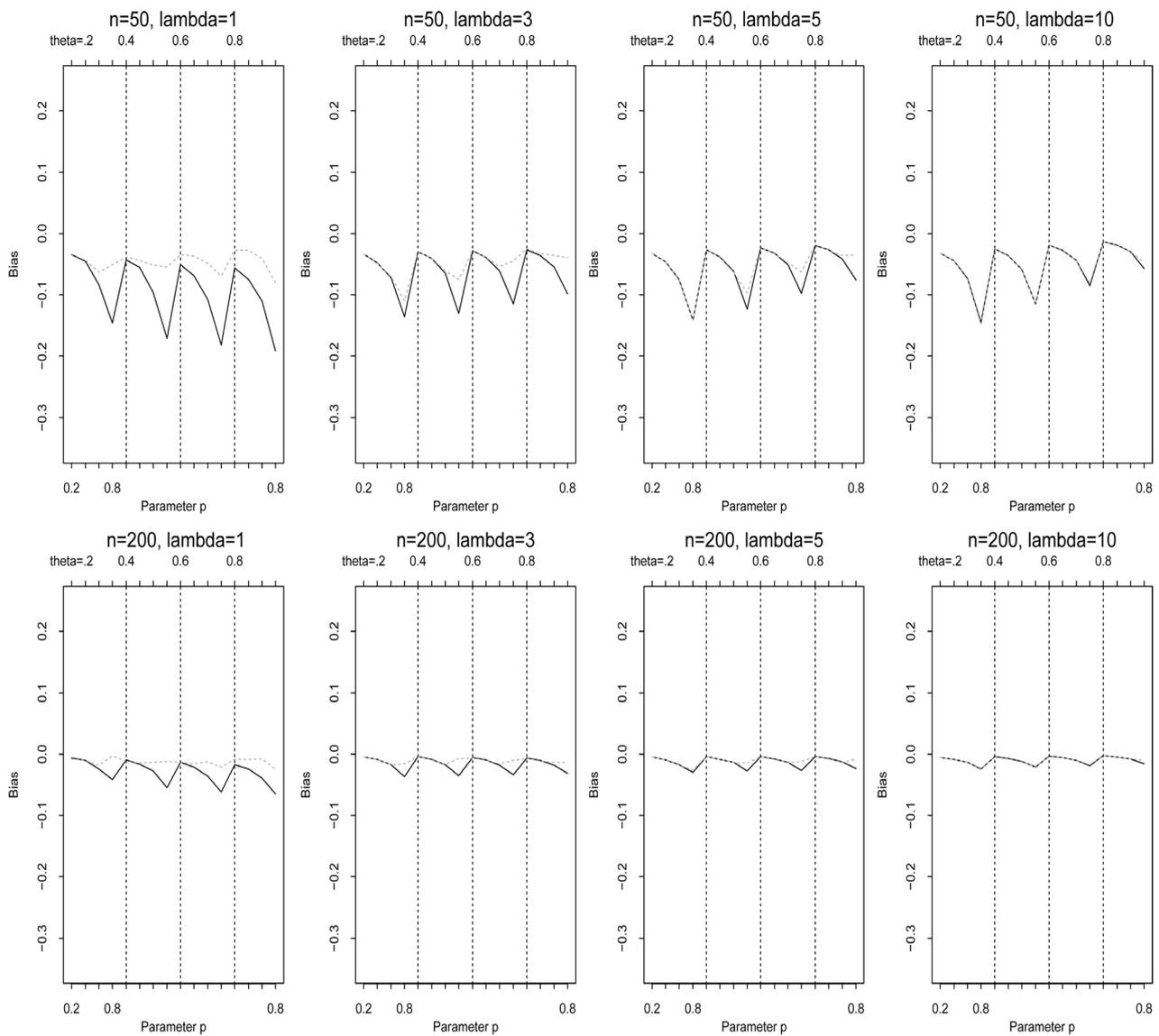**Figure 1.** Bias of the estimators of *p* (Moment: ----- MLE: - - -).

Figure 2. Bias of the estimators of λ (Moment: ----- MLE: - - -).

Figure 3. Bias of the estimators of θ (Moment: ----- MLE: - - -).

it decreases when the size of the series increases. **Figure 1** shows that the bias of $\hat{p}$ is much smaller than that of $\overline{p}$, except when $n = 200$ and $\theta = 0.2$ where they are almost equal to 0. The bias of the two estimators is negative. In **Figure 2**, we see that the bias of $\hat{\lambda}$ and $\overline{\lambda}$ is close to 0 when $\lambda = 1$; as λ increases, $\hat{\lambda}$ and $\overline{\lambda}$ are more biased. In all cases, the bias of the estimator of λ is positive. The bias of the estimator of θ behaves like that of p (**Figure 3**); for the two estimation methods, it is similar for $\lambda = 5$ or 10.

Since the moments estimators and the conditional MLE's are almost unbiased for large *n*, we study their asymptotic efficiency in the next section.

5. Asymptotic Efficiency of Estimators

We will first discuss the techniques by which we can obtain the asymptotic variance-covariance matrix of the estimators under the three estimation methods. To study efficiencies, we calculate, in subsection 5.4, the ratios of the variances of the estimators and the ratio of the determinants of their variance-covariance matrix using observations simulated from a $GPAR(1)$ process for various values of the parameters. The results are summarized in **Table 2** and **Table 3** of this section.

λ	p	θ	$\mathrm{Var}(\bar{p})/\mathrm{Var}(\hat{p})$	$\mathrm{Var}(\bar{\lambda})/Var(\hat{\lambda})$	$\mathrm{Var}(\bar{\theta})/\mathrm{Var}(\hat{\theta})$	$|\bar{\Theta}|/|\hat{\Theta}|$
1	0.2	0.2	2.27	1.10	1.33	1.79
		0.4	4.12	1.38	1.84	3.75
		0.6	8.55	2.07	2.11	8.18
		0.8	19.9	4.88	2.78	23.6
	0.4	0.2	5.11	1.19	1.44	2.46
		0.4	9.61	1.47	1.63	4.92
		0.6	17.5	1.96	1.65	10.8
		0.8	38.1	3.88	1.94	32.7
	0.6	0.2	16.9	1.21	1.56	3.81
		0.4	28.2	1.48	1.44	7.48
		0.6	57.4	2.14	1.53	18.8
		0.8	151.6	5.08	1.81	69.2
	0.8	0.2	76.4	1.37	2.69	6.68
		0.4	153.9	2.09	2.32	13.9
		0.6	295.3	4.56	2.09	25.2
		0.8	688.3	16.3	2.20	58.1
3	0.2	0.2	1.42	1.03	1.10	1.10
		0.4	1.96	1.23	1.34	1.68
		0.6	3.37	1.71	1.69	3.19
		0.8	8.02	3.38	2.23	10.4
	0.4	0.2	2.85	1.07	1.02	1.21
		0.4	4.52	1.24	1.21	1.97
		0.6	7.88	1.63	1.41	3.99
		0.8	18.5	2.99	1.70	14.1
	0.6	0.2	9.65	1.19	1.21	1.79
		0.4	16.2	1.35	1.27	3.15
		0.6	30.6	1.75	1.36	7.60
		0.8	78.2	3.25	1.51	30.6
	0.8	0.2	84.1	1.55	2.99	5.93
		0.4	126.4	1.90	2.42	10.5
		0.6	215.2	2.62	1.99	20.8
		0.8	501.4	7.21	2.01	67.1

5.1. Method of Moments

By using an asymptotically equivalent factor of $1/n$ instead of $1/n-1$ in Equation (3), moments estimators $\bar{\Theta}=\left(\bar{p},\bar{\lambda},\bar{\theta}\right)$ are given as solutions of the system of equations

$$\sum_{t=1}^{n-1}\left[\frac{(n-1)\left(X_t-\bar{X}\right)\left(X_{t+1}-\bar{X}\right)}{\sum_{t=1}^{n}\left(X_t-\bar{X}\right)^2}-p\right]=0$$

Table 3. Efficiency of quasi-likelihood estimators.

λ	p	θ	$\mathrm{Var}(\tilde{p})/\mathrm{Var}(\hat{p})$	$\mathrm{Var}(\tilde{\lambda})/\mathrm{Var}(\hat{\lambda})$	$\mathrm{Var}(\tilde{\theta})/\mathrm{Var}(\hat{\theta})$	$|\tilde{\Theta}|/|\hat{\Theta}|$
1	0.2	0.2	2.06	1.07	1.18	2.25
		0.4	3.16	1.25	1.73	5.09
		0.6	7.17	2.30	1.82	15.3
		0.8	12.4	7.39	2.48	48.5
	0.4	0.2	2.03	1.08	1.35	2.49
		0.4	3.07	1.36	1.67	5.14
		0.6	5.79	2.73	1.99	15.9
		0.8	14.7	15.4	3.22	98.4
	0.6	0.2	1.94	1.30	1.66	3.63
		0.4	2.46	1.46	1.97	5.95
		0.6	5.91	4.08	2.83	23.0
		0.8	30.6	19.8	5.62	38.2
	0.8	0.2	2.27	2.14	2.94	12.0
		0.4	4.03	2.94	4.45	19.0
		0.6	14.8	13.7	6.61	90.6
		0.8	120.1	44.4	16.1	41.9
3	0.2	0.2	1.37	1.03	1.09	1.40
		0.4	1.79	1.16	1.25	2.12
		0.6	2.72	1.52	1.50	4.24
		0.8	6.23	3.04	1.94	17.6
	0.4	0.2	1.39	1.08	1.12	1.37
		0.4	1.82	1.19	1.23	2.12
		0.6	2.72	1.58	1.52	4.55
		0.8	5.47	3.29	2.03	17.5
	0.6	0.2	1.76	1.27	1.41	1.88
		0.4	2.32	1.44	1.49	3.21
		0.6	3.01	1.87	1.82	6.87
		0.8	6.36	4.15	2.69	26.0
	0.8	0.2	3.67	1.88	3.01	6.50
		0.4	3.92	2.19	2.52	12.6
		0.6	6.61	2.94	2.97	32.4
		0.8	12.0	8.34	5.39	58.5

$$\sum_{t=1}^{n}\left[X_t(1-p)-\frac{(1-p)\lambda}{1-\theta}\right]=0$$

$$\sum_{t=1}^{n}\left[(X_t-\overline{X})^2-\frac{\lambda}{(1-\theta)^3}\right]=0.$$

Let us define the functions

$$f_t(p,\lambda,\theta)=\frac{(n-1)(X_t-\overline{X})(X_{t+1}-\overline{X})}{\sum_{t=1}^{n}(X_t-\overline{X})^2}-p$$

$$g_t(p, \lambda, \theta) = X_t(1-p) - \frac{(1-p)\lambda}{1-\theta}$$

$$h_t(p, \lambda, \theta) = (X_t - \bar{X})^2 - \frac{\lambda}{(1-\theta)^3}$$

and the vector

$$F_n(\Theta) = \left(\sum_{t=1}^{n-1} f_t(p, \lambda, \theta), \sum_{t=1}^{n} g_t(p, \lambda, \theta), \sum_{t=1}^{n} h_t(p, \lambda, \theta) \right).$$

The expected values $E[f_t(p, \lambda, \theta)]$, $E[g_t(p, \lambda, \theta)]$ and $E[h_t(p, \lambda, \theta)]$ are asymptotically equal to 0. Using a Taylor series expansion around $\Theta_0 = (p_0, \lambda_0, \theta_0)$, the true parameter value, we obtain

$$F_n(\bar{\Theta}) = F_n(\Theta_0) + \nabla F_n(\Theta_0)(\bar{\Theta} - \Theta_0) + \varepsilon, \tag{5}$$

where $\sqrt{n}\varepsilon \xrightarrow{P} 0$, with $\xrightarrow{P}$ denoting convergence in probability.

Since $\bar{\Theta}$ is a solution of $F_n(\Theta) = 0$, Equation (5) can rewritten as

$$\nabla F_n(\Theta_0)(\bar{\Theta} - \Theta_0) = -F_n(\Theta_0) - \varepsilon,$$

$$\text{or } \frac{1}{n}\nabla F_n(\Theta_0)\sqrt{n}(\bar{\Theta} - \Theta_0) = \frac{-1}{\sqrt{n}} F_n(\Theta_0) + o_p(1).$$

Using Slutsky's theorem, we find that

$$\frac{1}{n}\nabla F_n(\Theta_0)\sqrt{n}(\bar{\Theta} - \Theta_0) \xrightarrow{\mathcal{L}} \frac{1}{\sqrt{n}} F_n(\Theta_0),$$

$$\text{or } \sqrt{n}(\bar{\Theta} - \Theta_0) \xrightarrow{\mathcal{L}} N\left(0, A^{-1}\text{Var}(Y)(A^{-1})' \right), \tag{6}$$

where, with probability 1,

$$A = \lim_{n\to\infty} \frac{1}{n}\nabla F_n(\Theta) \text{ and } Y = \lim_{n\to\infty} -\frac{1}{\sqrt{n}} F_n(\Theta_0).$$

Matrix A evaluated at Θ_0 can be estimated by

$$\hat{A} = \frac{1}{n}\begin{vmatrix} -n+1 & 0 & 0 \\ -n\bar{x} + \dfrac{n\lambda_0}{1-\theta_0} & -\dfrac{n(1-p_0)}{1-\theta_0} & -\dfrac{n\lambda_0(1-p_0)}{(1-\theta_0)^2} \\ 0 & -\dfrac{n}{(1-\theta_0)^3} & -\dfrac{3n\lambda_0}{(1-\theta_0)^4} \end{vmatrix}.$$

If θ_0, λ_0 and p_0 are unknown, they can be replaced by appropriate estimates. The variance-covariance matrix of Y is equal to

$$\frac{1}{n}\begin{pmatrix} \text{Var}\left(\sum_{i=1}^{n} f_t\right) & \text{Cov}\left(\sum_{i=1}^{n} f_t, \sum_{i=1}^{n} g_t\right) & \text{Cov}\left(\sum_{i=1}^{n} f_t, \sum_{i=1}^{n} h_t\right) \\ \text{Cov}\left(\sum_{i=1}^{n} g_t, \sum_{i=1}^{n} f_t\right) & \text{Var}\left(\sum_{i=1}^{n} g_t\right) & \text{Cov}\left(\sum_{i=1}^{n} g_t, \sum_{i=1}^{n} h_t\right) \\ \text{Cov}\left(\sum_{i=1}^{n} h_t, \sum_{i=1}^{n} f_t\right) & \text{Cov}\left(\sum_{i=1}^{n} h_t, \sum_{i=1}^{n} g_t\right) & \text{Var}\left(\sum_{i=1}^{n} h_t\right) \end{pmatrix}.$$

Let us consider the first element of this matrix:

$$\frac{1}{n}\text{Var}\left(\sum_{t=1}^{n-1} f_t\right) = \frac{1}{n}\sum_{t=1}^{n-1}\sum_{k=1}^{n-1} E[f_t f_k],$$

since asymptotically $E[f_t f_k] = \text{Cov}(f_t, f_k)$ (because $E[f_t] = 0$ as $n \to \infty$). In practice, we truncate these expressions, since $\text{Cov}(f_t, f_k) \to 0$, as $|t - k| \to \infty$. If we limit ourselves to a difference of $|t - k| \leq 5$, the last equality becomes

$$\frac{1}{n}\text{Var}\left(\sum_{t=1}^{n-1} f_t\right) \simeq \frac{1}{n}\sum\sum_{|t-k|\leq 5} E[f_t f_k].$$

Using the law of large numbers, we can estimate this last term by

$$\frac{1}{n}\sum\sum_{|t-k|\leq 5} f_t f_k.$$

The other elements of the matrix can be estimated in the same way.

5.2. Quasi-Likelihood Method

To determine the quasi-likelihood estimator $(\tilde{\Theta})$, we have to maximize

$$l(p, \lambda, \theta) = -0.5n\ln(2\pi) - 0.5\sum_{t=1}^{n}\left[\ln(\sigma_t^2) + \frac{(x_t - \mu_t)^2}{\sigma_t^2}\right]. \tag{7}$$

Let us define the quasi-score vector

$$S_t(\Theta) = \nabla l(\Theta) = \left(S_t^{(1)}, S_t^{(2)}, S_t^{(3)}\right)' = \left(\frac{\partial}{\partial p}l(\Theta), \frac{\partial}{\partial \lambda}l(\Theta), \frac{\partial}{\partial \theta}l(\Theta)\right)'.$$

From Hamilton [12], using quasi-likelihood theory, we conclude that

$$\sqrt{n}\left(\tilde{\Theta} - \Theta_0\right)' \xrightarrow{\mathcal{L}} N\left(0, D^{-1}SD^{-1}\right), \tag{8}$$

where with probability 1, D and S are limits in probability matrices. They are defined as

$$D = \lim_{n\to\infty} -\frac{1}{n}\begin{pmatrix} \sum_{t=1}^{n}E\left[\frac{\partial}{\partial p}\left(S_t^{(1)} \mid X_{t-1}\right)\right] & \sum_{t=1}^{n}E\left[\frac{\partial}{\partial \lambda}\left(S_t^{(1)} \mid X_{t-1}\right)\right] & \sum_{t=1}^{n}E\left[\frac{\partial}{\partial \theta}\left(S_t^{(1)} \mid X_{t-1}\right)\right] \\ \sum_{t=1}^{n}E\left[\frac{\partial}{\partial p}\left(S_t^{(2)} \mid X_{t-1}\right)\right] & \sum_{t=1}^{n}E\left[\frac{\partial}{\partial \lambda}\left(S_t^{(2)} \mid X_{t-1}\right)\right] & \sum_{t=1}^{n}E\left[\frac{\partial}{\partial \theta}\left(S_t^{(2)} \mid X_{t-1}\right)\right] \\ \sum_{t=1}^{n}E\left[\frac{\partial}{\partial p}\left(S_t^{(3)} \mid X_{t-1}\right)\right] & \sum_{t=1}^{n}E\left[\frac{\partial}{\partial \lambda}\left(S_t^{(3)} \mid X_{t-1}\right)\right] & \sum_{t=1}^{n}E\left[\frac{\partial}{\partial \theta}\left(S_t^{(3)} \mid X_{t-1}\right)\right] \end{pmatrix},$$

and

$$S = \lim_{n\to\infty} \frac{1}{n}\begin{pmatrix} \sum_{t=1}^{n}S_t^{(1)}S_t^{(1)} & \sum_{t=1}^{n}S_t^{(1)}S_t^{(2)} & \sum_{t=1}^{n}S_t^{(1)}S_t^{(3)} \\ \sum_{t=1}^{n}S_t^{(2)}S_t^{(1)} & \sum_{t=1}^{n}S_t^{(2)}S_t^{(2)} & \sum_{t=1}^{n}S_t^{(2)}S_t^{(3)} \\ \sum_{t=1}^{n}S_t^{(3)}S_t^{(1)} & \sum_{t=1}^{n}S_t^{(3)}S_t^{(2)} & \sum_{t=1}^{n}S_t^{(3)}S_t^{(3)} \end{pmatrix},$$

evaluated at Θ_0, the true parameter. We can obtain estimates for $\hat{D}$ and $\hat{S}$, where matrix $\hat{D}$ is defined as

$$\hat{D} = -\frac{1}{n}\begin{pmatrix} \sum_{t=1}^{n}\frac{\partial}{\partial p}\left(S_t^{(1)} \mid X_{t-1}\right) & \sum_{t=1}^{n}\frac{\partial}{\partial \lambda}\left(S_t^{(1)} \mid X_{t-1}\right) & \sum_{t=1}^{n}\frac{\partial}{\partial \theta}\left(S_t^{(1)} \mid X_{t-1}\right) \\ \sum_{t=1}^{n}\frac{\partial}{\partial p}\left(S_t^{(2)} \mid X_{t-1}\right) & \sum_{t=1}^{n}\frac{\partial}{\partial \lambda}\left(S_t^{(2)} \mid X_{t-1}\right) & \sum_{t=1}^{n}\frac{\partial}{\partial \theta}\left(S_t^{(2)} \mid X_{t-1}\right) \\ \sum_{t=1}^{n}\frac{\partial}{\partial p}\left(S_t^{(3)} \mid X_{t-1}\right) & \sum_{t=1}^{n}\frac{\partial}{\partial \lambda}\left(S_t^{(3)} \mid X_{t-1}\right) & \sum_{t=1}^{n}\frac{\partial}{\partial \theta}\left(S_t^{(3)} \mid X_{t-1}\right) \end{pmatrix},$$

and $\hat{S}$ is the finite version of S evaluated at Θ_0; the elements of $\hat{D}$ are evaluated numerically using

expression (7). Packages such as MATHEMATICA can handle these derivatives calculations numerically. Consequently, the variance-covariance matrix of $\tilde{\Theta}$ can be estimated by $\hat{D}^{-1}\hat{S}\hat{D}^{-1}$.

5.3. Conditional Maximum Likelihood

Using the true loglikelihood function from section 4.3, we define the score vector

$$S_t(\Theta) = \nabla l(\Theta) = \left(S_t^{(1)}, S_t^{(2)}, S_t^{(3)}\right)'.$$

From Hamilton [12], using likelihood theory, we find that

$$\sqrt{n}\left(\hat{\Theta} - \Theta_0\right) \xrightarrow{\mathcal{L}} N\left(0, S^{-1}\right), \tag{9}$$

where matrix S is defined analogously as in the previous section, but with a different loglikelihood function.

5.4. Numerical Comparisons

Table 2 and **Table 4** give the estimate of the asymptotic efficiency of the moment and the quasi-likelihood estimators compared to the MLE, calculated from 20,000 observations (10 series of 2000 observations) generated from a $GPAR(1)$ process with various parameter values.

Comparing **Table 2** and **Table 3**, the quasi-likelihood estimator for p has a smaller variance than the moments estimator; for λ, it depends on the values of the parameters. The moments estimator of θ has a smaller variance than the quasi-likelihood estimator, except when $p = 0.2$, where $\hat{\theta}$ is better than $\bar{\theta}$.

The estimated determinant of the variance-covariance matrix of $\hat{\Theta}$ using the average of the determinants is always smaller than that of $\bar{\Theta}$ and $\tilde{\Theta}$ (last column of **Table 2** and **Table 3**). The MLE is more efficient than the moment or the quasi-likelihood estimator, and the moment estimator more efficient than the quasi-likelihood estimator, in general.

6. Applications: Number of Computer Breakdowns

In this section, we perform some tests on a real time series presented by Al-Nachawati *et al.* [3] on the number of weekly computer breakdowns for 128 consecutive weeks. This series is overdispersed, since its mean and variance are equal to 4.016 and 14.504. In **Figure 4**, the acf function is seen to decrease with the lag, while the pacf is high for lag 1 and low thereafter; a $GPAR(1)$ model could therefore be appropriate for this series. We use the $GPAR(1)$ model in the analysis.

Since the MLE was shown to be the best asymptotic estimator in the previous section, the parameters were estimated with this method; the estimates appear in **Table 4**, with the estimated variance-covariance matrix.

With the estimated variance-covariance matrices based on expressions (6), (8) and (9) of Section 5, Wald tests can be performed quite easily depending on which estimator has been chosen.

For example, to test $H_0 : \theta = \theta_0$ using $\tilde{\theta}$, the quasilikelihood estimator, the statistic can be based on the statistic $Z = \left(\tilde{\theta} - \theta_0\right)\big/\sqrt{V\left(\tilde{\theta}\right)}$, where $V\left(\tilde{\theta}\right)$ is an estimate of the variance of $\tilde{\theta}$, which can be obtained from the corresponding diagonal element of $\hat{V} = \hat{D}^{-1}\hat{S}\hat{D}^{-1}$. Since $\left(\tilde{\theta} - \theta_0\right)\big/\sqrt{V\left(\tilde{\theta}\right)}$ is asymptotically $N(0,1)$, we reject H_0 at level α if $|Z|$ is greater than $z_{1-\alpha/2}$.
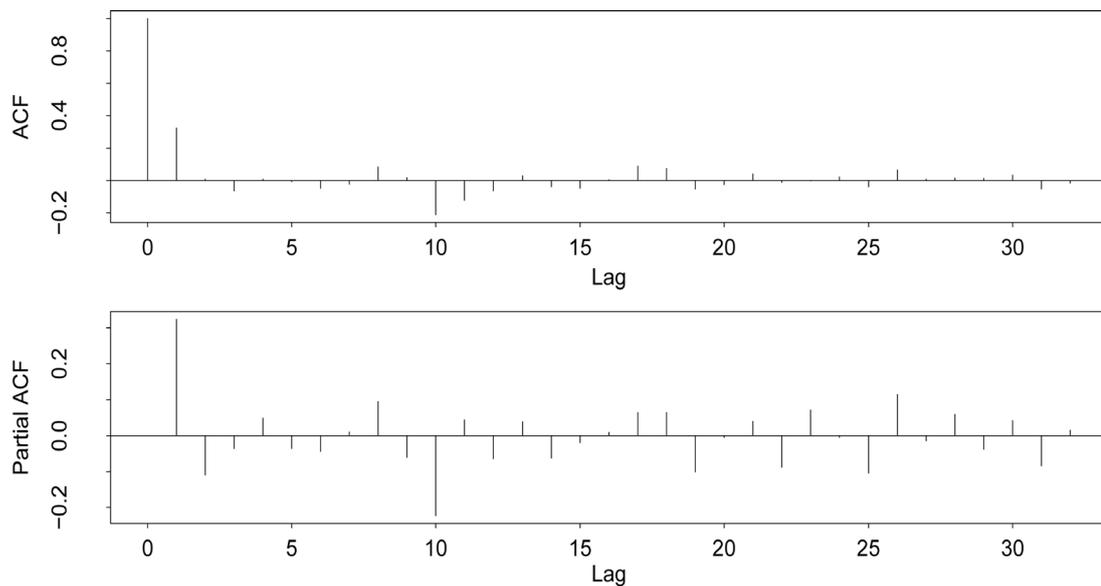
To test $H_0 : p = p_0, \lambda = \lambda_0, \theta = \theta_0$, the test statistic can be based on

$$\left[\tilde{\Theta} - \Theta_0\right]\hat{V}^{-1}\left[\tilde{\Theta} - \Theta_0\right]',$$

which follows a $\chi^2(3)$ distribution asymptotically. It is expected that the more efficient the estimator is, the

Table 4. MLE's of the parameters.

$\hat{p}$	$\hat{\lambda}$	$\hat{\theta}$	Variance-covariance matrix		
0.323	2.125	0.471	0.0055	0.0017	0.0008
			0.0017	0.0493	−0.0040
			0.0008	−0.0040	0.0026

Figure 4. Acf and pacf.

more powerful the test will be.

With the estimated parameters, we can test the $GPAR(1)$ model versus the simpler $PAR(1)$ model. Since the conditional MLE $\hat{\theta}$ equals 0.471, with a variance of 0.0026, performing the test $H_0 : \theta = 0$ vs $H_1 : \theta > 0$ gives $Z = 0.471/\sqrt{0.0026} = 9.24$. This leads us to reject H_0 and to conclude that the $GPAR(1)$ model is more appropriate: there is overdispersion in the observations.

Acknowledgements

References

[1] Alzaid, A.A. and Al-Osh, M.A. (1993) Some Autoregressive Moving Average Processes with Generalized Poisson Marginal Distributions. *Annals of the Institute of Mathematical Statistics*, **45**, 223-232. http://dx.doi.org/10.1007/BF00775809

[2] 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. http://dx.doi.org/10.1111/j.1467-9892.1987.tb00438.x

[3] Al-Nachawati, H., Alwasel, I. and Alzaid, A.A. (1997) Estimating the Parameters of the Generalized Poisson AR(1) process. *Journal of Statistical Computation and Simulation*, **56**, 337-352. http://dx.doi.org/10.1080/00949659708811798

[4] Brännäs, K. (1994) Estimation and Testing in Integer-Valued AR(1) Models. Umeå Economic Studies 335, Umeå, Sweden.

[5] Consul, P.C. (1989) Generalized Poisson Distribution: Properties and Applications. Marcel Dekker Inc., New York.

[6] Ambagaspitiya, R.S. and Balakrishnan, N. (1994) On the Compound Generalized Poisson Distributions. *ASTIN Bulletin*, **24**, 255-263. http://dx.doi.org/10.2143/AST.24.2.2005069

[7] Freeland, R.K. and McCabe, B. (2004) Analysis of Low Count Time Series Data by Poisson Autoregression. *Journal of Time Series Analysis*, **25**, 701-722. http://dx.doi.org/10.1111/j.1467-9892.2004.01885.x

[8] Cardinal, M., Roy, R. and Lambert, J. (1999) On the Application of Integer-Valued Time Series Models for the Analysis of Disease Incidence. *Statistics in Medicine*, **18**, 2025-2039. http://dx.doi.org/10.1002/(SICI)1097-0258(19990815)18:15<2025::AID-SIM163>3.0.CO;2-D

[9] Whittle, P. (1961) Gaussian Estimation in Time Series. *Bulletin of the International Statistical Institute*, **39**, 1-26.

[10] Shenton, L.R. (1986) Quasi-Binomial Distributions. Encyclopedia of Statistical Sciences, John Wiley & Sons, New York, Vol. 7, 458-460.

[11] Gouriéroux, C. and Monfort. A. (1983) Cours de Hamilton, J.D. (1994) Time Series Analysis. Princeton University Press, New Jersey.

[12] Hamilton, J.D. (1994) Time Series Analysis. Princeton University Press, New Jersey.

[13] Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P. (2002) Numerical Recipies in C++. The Art of Scientific Computing. Cambridge University Press, Cambridge.