ON PARAMETER ESTIMATION FOR EXPONENTIAL DISPERSION ARMA MODELS

By Peter X.-K. Song and Dingan Feng

University of Waterloo and HEC Montreal

Abstract. A class of autoregressive moving-average (ARMA) models proposed by Jørgensen and Song [*Journal of Applied Probability* (1998), vol. 35, pp. 78–92] with exponential dispersion model margins are useful to deal with non-normal stationary time series with high-order autocorrelation. One property associated with the class of models is that the projection process takes the exact form of the classical Box and Jenkins ARMA representation, leading to considerable ease to establish theories. This paper focuses on the issue of parameter estimation for such models, which has not been thoroughly investigated in Jørgensen and Song's paper. The key of the proposed approach is to treat the residual process associated with the projection essentially as a measurement error, which enables us to formulate directly an ARMA representation for the observed time series. The parameter estimation therefore becomes straightforward using the existing methods for the Box and Jenkins ARMA models such as the quasi-likelihood method. The approach is illustrated by simulation studies and by an analysis of myoclonic seizure counts.

Keywords. Exponential dispersion models; measurement error; non-normal time series; quasi-likelihood; thinning; time series of counts.

1. INTRODUCTION

There are many stationary time series that do not follow Box and Jenkins autoregressive moving average (ARMA) models, such as the stationary integer-valued time series. To be more specific, let $\{X_t\}$ be a stationary AR(1) process with Poisson marginal distribution $\mathcal{P}(\mu)$ with mean μ. It is known (e.g. Steutel and van Harn, 1979) that this AR(1) process takes the form

$$X_t = \alpha \circ X_{t-1} + \varepsilon_t \qquad (1)$$

where $\circ$ denotes the thinning operation defined by

$$\alpha \circ X_{t-1} = \sum_{j=1}^{X_{t-1}} B_j \sim \mathcal{P}(\alpha\mu)$$

with independent and identically distributed (i.i.d.) Bernoulli random variables B_j's, the probabilities of success of which equal to $\alpha \in [0, 1]$. Another example is the stationary AR(1) gamma process (e.g. Lewis *et al.*, 1989) defined by

$$X_t = G_t X_{t-1} + \varepsilon_t \tag{2}$$

where X_t follows marginally a gamma distribution and the thinning operation is undertaken by a beta random variable G_t that is independent of both X_t and ε_t with parameters $\alpha \in [0, 1]$ and $1 - \alpha$.

Both examples clearly indicate that when the marginal distribution of a stationary process is known to be non-normal, the Box and Jenkins ARMA representation no longer applies. Joe (1996) proposed a unified definition for ARMA processes within a class of marginal distributions that are infinitely divisible, including the Poisson (1) and gamma (2) processes as special cases. A shortcoming of Joe's approach is the inflexibility in incorporating high-order AR processes. As shown in his paper, his construction of AR stationary processes gets already very complex even at the second order of autocorrelation.

To overcome this, Jørgensen and Song (1998) proposed another unified approach to constructing an alternative class of stationary ARMA time series, the marginal distributions of which are exponential dispersion (ED) models (Jørgensen, 1997). It is known that the ED models serve as the class of error distributions for the generalized linear models, including special distributions such as normal, Poisson, gamma, binomial, negative binomial, inverse Gaussian and compound Poisson. Note that the problem of extending Box and Jenkins Gaussian ARMA models to a non-Gaussian framework has long been of great interest in the statistical literature. To incorporate covariates in such an extension, both observation- and parameter-driven models have been proposed. Examples of the parameter-driven model are those of Zeger (1988) and Jørgensen et al. (1999). Among many observation-driven models, the generalized autoregressive moving average (GARMA) model of Benjamin et al. (2003) presents a general modelling framework to analyze non-Gaussian time-series data, including the models proposed by Zeger and Qaqish (1988) and Li (1994) as special cases. The class of GARMA models is primarily proposed to address nonstationary behaviour for the mean of a non-Gaussian time-series, and it becomes somewhat restrictive to model stationary time series. Given stationarity, a time-series model is deemed to address autocorrelation behaviour, rather than that of marginal moments. This paper deals with a class of stationary ARMA models proposed by Jørgensen and Song (1998), with the focus on the estimation of parameters in the autocorrelation structure.

According to Jørgensen (1997), an ED distribution is infinitely divisible if its index set is $(0, \infty)$ or $(-\infty, \infty)$. An interesting property associated with the Jørgensen and Song (1998) construction is that their stationary ARMA(p, q) process $\{X_t\}$ can be decomposed into the sum of a projection process $\{Y_t\}$ and a residual process $\{\delta_t\}$,

$$X_t = Y_t + \delta_t$$

where the projection process Y_t follows the Box and Jenkins ARMA representation of the form,

$$Y_t - \phi_1 Y_{t-1} - \cdots - \phi_p Y_{t-p} = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \cdots + \psi_q \varepsilon_{t-q}.$$

Such a decomposition sheds light on proceeding the statistical analysis of non-normal stationary time series by borrowing the existing tools from the arsenal of the Box and Jenkins ARMA models. One aspect that this paper deals with is the development of parameter estimation through the use of this decomposition.

To utilize the decomposition, technically we may think of the residual term δ_t essentially as a kind of measurement error for the underlying process Y_t. This observation enables us to express the X_t process itself in a direct form of the Box and Jenkins ARMA representation with a different noise variate ξ_t, following Box and Jenkins (1976). Therefore, parameter estimation can be carried out conveniently by utilizing existing algorithms such as the quasi-likelihood method. Asymptotics are also naturally followed under some mild moment assumptions.

The paper is organized as follows. After giving a brief description of the Jørgensen and Song ARMA models in Section 2, we investigate the properties of the residual process δ_t in Section 3. Section 4 presents details of parameter estimation for general Jørgensen and Song ARMA(p,q) models, and Section 5 focuses on the AR(1) process on which a simulation study is demonstrated. Section 6 presents a data analysis example, and Section 7 provides some discussions.

2. PRELIMINARIES

We start with a brief summary of the theory of exponential dispersion models, and more details about these models can be found from Jørgensen (1997). Let $X \sim \mathrm{ED}^*(\theta, \lambda)$ denote an additive exponential dispersion model with probability density function

$$p(x; \theta, \lambda) = c(x; \lambda) \exp\{\theta x - \lambda \kappa(\theta)\}, \tag{3}$$

for $x \in \mathcal{R}$, with respect to a suitable measure, where the domain of θ, denoted by Φ, is an interval and $\lambda \in \Lambda \subseteq \mathcal{R}_+$. The model is infinitely divisible if and only if the domain of the index parameter λ is $\Lambda = \mathcal{R}_+$. For $\theta \in \mathrm{int}\Phi$, the mean and variance of X are

$$E(X) = \lambda \tau(\theta) \quad \text{and} \quad \mathrm{var}(X) = \lambda \tau'(\theta),$$

where $\tau(\theta) = \kappa'(\theta)$. The additive exponential dispersion model satisfies the convolution formula

$$\mathrm{ED}^*(\theta, \lambda_1) * \mathrm{ED}^*(\theta, \lambda_2) = \mathrm{ED}^*(\theta, \lambda_1 + \lambda_2), \tag{4}$$

which implies that the model is closed under convolution of the members with a common value of the canonical parameter θ. Note that with different cumulant generating functions $\kappa(\cdot)$ in eqn (3), the density functions of the ED* models

correspond to normal, Poisson, gamma, binomial, negative binomial, inverse Gaussian and many other distributions.

By eqn (4), a general definition of the thinning operation (Joe, 1996) is given as follows. Let X_1 and X_2 be independent ED^* random variables, with $X_i \sim ED^*(\theta, \lambda_i)$, $i = 1, 2$. Their sum is denoted by $W = X_1 + X_2 \sim ED^*(\theta, \lambda)$, where $\lambda = \lambda_1 + \lambda_2$. Clearly, W is sufficient for the parameter θ, so the conditional distribution of X_1 given $W = x$ does not depend on the parameter θ, and is therefore denoted by

$$X_1|X_1 + X_2 = x \sim G(\lambda_1, \lambda_2, x).$$

The distribution $G(\lambda_1, \lambda_2, x)$ is termed the *contraction* corresponding to $ED^*(\theta, \lambda)$ by Jørgensen and Song (1998). This terminology is made according to the fact that when $ED^*(\theta, \lambda)$ is non-negative the support of $G(\lambda_1, \lambda_2, x)$ is between 0 and x. Table 2 of Jørgensen and Song (1998) lists a number of contractions for different ED^* marginals.

A thinning operator, denoted by $A(\cdot; \alpha)$, $\alpha \in [0, 1]$, is a stochastic function of a random variable $X \sim ED^*(\theta, \lambda)$ whose conditional distribution is given by

$$A(X; \alpha)|X = x \sim G(\alpha\lambda, \bar{\alpha}\lambda, x),$$

where $\bar{\alpha} = 1 - \alpha$. It follows immediately from Proposition 3.1 of Jørgensen and Song (1998) that $A(X; \alpha) \sim ED^*(\theta, \alpha\lambda)$, which means $A(X; \alpha)$ is the *thinning of X by the proportion* α. In addition, it is known that $X - A(X; \alpha) \sim ED * (\theta, \bar{\alpha}\lambda)$ is independent of $A(X; \alpha)$. According to Joe (1996),

$$E(A(X; \alpha)|X = x) = \alpha x. \tag{5}$$

In the extreme cases $\alpha = 0, 1$, $A(X; 1) = X$ and $A(X; 0) = 0$, respectively.

Generalizing the definition of the thinning operation for a pair of independent variables, Jørgensen and Song (1998) defined a class of infinite-order moving average processes for infinitely divisible ED^* model marginals. Furthermore, they defined causal ARMA processes with ED^* model marginals through the representation of the infinite-order moving average process. A brief summary is given as follows.

Let ϕ and ψ denote polynomials of degree p and q with no common roots, respectively, given by

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p, \psi(z) = 1 + \psi_1 z + \cdots + \psi_q z^q,$$

and let $\alpha_i \in [0, 1]$, $i = 0, 1, \ldots$, be the coefficients determined by the following power-series expansion

$$\frac{\psi(z)}{\phi(z)} = \sum_{j=0}^{\infty} \alpha_j z^j, \quad |z| \leq 1, \tag{6}$$

where $\phi(z) \neq 0$ for $|z| \leq 1$, $\alpha_0 = 1$ and $\alpha_+ = \sum_j \alpha_j < \infty$. Obviously, with given coefficients ϕ_k's and ψ_k's, the α_j's are determined by the following recursive equations,

$$\alpha_j - \sum_{0<k\leq j} \phi_k \alpha_{j-k} = \psi_j, \quad 0 \leq j < \max(p, q+1),$$

and

$$\alpha_j - \sum_{0<k\leq j} \phi_k \alpha_{j-k} = 0, \quad j \geq \max(p, q+1), \tag{7}$$

with $\psi_0 = 1$, $\psi_j = 0$, $j > q$ and $\alpha_j = 0$, $j < 0$. Assume that the innovations $\{\varepsilon_t, t = 0, \pm 1, \ldots\}$ are i.i.d. random variables with a marginal distribution $ED^*(\theta, \lambda/\alpha_+)$ and that given the innovations $\{\varepsilon_t, t = 0, \pm 1, \ldots\}$ the thinning operators $\{A_{t,j}(\varepsilon_{t-j}; \alpha_j), j = 0, 1, \ldots, t = 0, \pm 1, \ldots\}$ are conditionally independent. According to Jørgensen and Song (1998), an ARMA(p, q) process $\{X_t\}$ with $ED^*(\theta, \lambda)$ margin takes the following form of infinite independent sum

$$X_t = \varepsilon_t + \sum_{j=1}^{\infty} A_{t,j}(\varepsilon_{t-j}; \alpha_j), \tag{8}$$

where the thinning operator $A_{t,j}(\varepsilon_{t-j}; \alpha_j)$ follows marginally $ED^*(\theta, \alpha_j\lambda/\alpha_+)$ and conditionally the contraction distribution $G(\alpha_j\lambda, (1 - \alpha_j)\lambda, \varepsilon_{t-j})$ given ε_{t-j}.

Denote the σ-algebras by $\mathcal{M}_k = \sigma\{\varepsilon_k, \varepsilon_{k-1}, \ldots\}$ and $\mathcal{M}_a = \sigma\{\ldots, \varepsilon_{-1}, \varepsilon_0, \varepsilon_1, \ldots\}$. In the case of finite expectation, the projection process $\{Y_t\}$ is obtained as the projection of the observation process $\{X_t\}$ onto the space generated by the innovations, that is,

$$Y_t = E(X_t|\mathcal{M}_a) = E(X_t|\mathcal{M}_t),$$

and it is easy to see by eqn (5) that the Y_t forms a linear process given by

$$Y_t = \sum_{j=0}^{\infty} \alpha_j \varepsilon_{t-j}. \tag{9}$$

Therefore, with the so-chosen coefficients α_j's in eqn (6), the projection process can be rewritten in the form

$$Y_t - \phi_1 Y_{t-1} - \cdots - \phi_p Y_{t-p} = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \cdots + \psi_q \varepsilon_{t-q}. \tag{10}$$

It is noticeable that this representation is of the same form as the Box and Jenkins ARMA process, where, however, the ε_t's are the ED innovations that have possibly non-zero mean μ_ε and variance σ_ε^2, rather than the regular mean-zero white noise. Moreover, the residual process is yielded as the difference between the observed $\{X_t\}$ process and its projection process $\{Y_t\}$, namely $\delta_t = X_t - Y_t$. As shown in Proposition 5 below, the residual process δ_t turns out to be uncorrelated with the Y_t process, indicating that the stationary process X_t can be decomposed as an *orthogonal* sum of the stationary projection process Y_t and the residual process δ_t,

$$X_t = Y_t + \delta_t.$$

By orthogonality between two random variables, we mean that the two variables are uncorrelated (Stout, 1974, p. 14).

3. PROPERTIES

We now present some basic properties of the Jørgensen and Song ARMA processes, which are useful for the development of parameter estimation in the next section. Although some of results seem to be obvious, to have a self-contained framework for this paper we keep them here without proofs.

It is easy to see that the projection process $\{Y_t\}$ is the minimum mean squared error prediction for the $\{X_t\}$ process on the basis of the innovations.

PROPOSITION 1. *For an arbitrary function f,*

$$E\{X_t - E(X_t|\mathcal{M}_t)\}^2 \leq E\{X_t - f(\varepsilon_t, \varepsilon_{t-1}, \ldots)\}^2,$$

provided that the expectations exist.

Moreover, by eqn (9) this optimal process is obviously linear and unbiased since $E(X_t - Y_t) = 0$.

PROPOSITION 2. *With the normal margin, the residual process δ_t is marginally normally distributed according to $N(0, \lambda \sum_{j=1}^{\infty} \alpha_j \bar{\alpha}_j)$.*

This result is valid simply because of the fact that the thinning operator

$$A_{t,j}(\varepsilon_{t-j}; \alpha_j) = \alpha_j \varepsilon_{t-j} + \delta_{t,j} \quad \text{with } \delta_{t,j} \sim N\big(0, \lambda \alpha_j \bar{\alpha}_j\big)$$

according to Jørgensen and Song (1998, p. 83). In general, we have the following moment properties.

PROPOSITION 3. *A Jørgensen and Song ARMA(p, q) process (8) and its projection process (10) have the same mean and the same autocovariance function (ACVF) for all lags except zero.*

Proposition 3 implies that although the two processes X_t and Y_t look very different in expressions, they have the same first moment, $\mu_x = \mu_\varepsilon \alpha_+$, and the same second moments (except the variance). This is true because for $h > 0$, their ACVFs are

$$\gamma_x(h) = E\{\text{cov}(X_t, X_{t+h}|\mathcal{M}_a)\} + \text{cov}\{E(X_t|\mathcal{M}_a), E(X_{t+h}|\mathcal{M}_a)\} = \gamma_y(h),$$

where X_t and X_{t+h} are conditionally independent given $\{\varepsilon_t\}$.

Clearly, var(X_t) and var(Y_t) are different and equal to, respectively,

$$\gamma_x(0) = \sigma_\varepsilon^2 \sum_j \alpha_j \quad \text{and} \quad \gamma_y(0) = \sigma_\varepsilon^2 \sum_j \alpha_j^2.$$

Thus, $\gamma_y(0) < \gamma_x(0)$ because of the fact that $\alpha_j \in (0, 1)$, which implies that the observation process X_t is more volatile than the projection process Y_t. Moreover, their autocorrelation functions (ACFs), $\rho_x(h)$ and $\rho_y(h)$, are related in the form of

$$\rho_x(h) = \frac{\sum_j \alpha_j^2}{\alpha_+} \rho_y(h) = \omega \rho_y(h), \quad \text{for } h \geq 1, \tag{11}$$

where the factor $\omega = \sum_j \alpha_j^2 / \alpha_+$ being constant in the interval $(0, 1)$. An interpretation of the factor ω can be given from the perspective of measurement error. In the context of time series with measurement error, the so-called corruption coefficient (Ashley and Vaughan, 1986) is the ratio equal to

$$\frac{\sigma_\delta^2}{\gamma_x(0)} = \frac{\gamma_x(0) - \gamma_y(0)}{\gamma_x(0)} = 1 - \omega,$$

where the first equality is due to Proposition 5(c) in this section. As opposed to corruption, ω indicates the percent of the observed variation attributed to the underlying true process, and therefore it is referred to as the *anticorruption* coefficient. This coefficient may also be taken as to indicate the degree of similarity between the two processes.

Since the two ACFs are proportional by the time-independent anticorruption coefficient ω, both show the same pattern, with the ACF of the Y_t being on the top of the ACF of the X_t. This property is practically useful to select possible orders p and q for the projection process Y_t by the pattern of the empirical (sample) ACF based on the observed process X_t. In some simple cases, this coefficient can easily be obtained. For instance, as shown in Section 5, the Jørgensen and Song AR(1) process gives $\omega = 1/(1 + \phi)$ where ϕ is the autocorrelation coefficient.

It is easy to see that the projection process $\{Y_t\}$ is (weakly) stationary. Thus, the projection transformation preserves not only the stationarity but also the same moments of the mean and autocovariance function.

Let us now consider autoregressive models. A Jørgensen and Song autoregressive process of order p, AR(p), is defined as effectively to be a Jørgensen and Song ARMA(p, 0) process. A justification for such a definition is given by Theorem 6.2 of Jørgensen and Song (1998), i.e. the partial ACF of the so-defined AR(p) process is zero for lags greater than p. An advantage of the Jørgensen and Song AR processes is that they preserve the recursion in computing its ACF. Followed directly from the recursive formula for the ACF of the Box and Jenkins AR process (10), it gives

PROPOSITION 4. *The ACF of a Jørgensen and Song AR(p) process $\{X_t\}$ has a recursive relation given by*

$$\rho_x(h) = \phi_1 \rho_x(h-1) + \cdots + \phi_p \rho_x(h-p), \quad \text{for } h \geq 1,$$

where $\rho_x(0) = \omega^{-1}$ defined in eqn (11).

This recursive relation is useful to compute the autocorrelation function of process $\{X_t\}$ numerically. Consequently, the Yule–Walker equation can be formulated directly from this recursive relation. We now end the section by listing some useful moment properties for the residual term δ_t. Although some of them seem to be trivial, we list them here anyway simply for completeness.

PROPOSITION 5. *Let $\delta_t = X_t - Y_t$ be the residual, where $\{X_t\}$ follows a Jørgensen and Song ARMA(p, q) process and $\{Y_t\}$ is its projection. Then,*
(a) *the residual has zero expectation, $E(\delta_t) = 0$;*
(b) *the residual is uncorrelated with the projection, $\text{cov}(Y_t, \delta_t) = 0$ for all t;*
(c) $\text{var}(\delta_t) = \sigma_x^2 - \sigma_y^2 = \sigma_\varepsilon^2 \sum_j \alpha_j(1 - \alpha_j)$;
(d) *the residual process $\{\delta_t\}$ and the innovation process $\{\varepsilon_t\}$ are uncorrelated, namely, $\text{cov}(\delta_t, \varepsilon_s) = 0$, for all t, s.*
(e) *the residuals are uncorrelated, $\text{cov}(\delta_t, \delta_s) = 0$, $t \neq s$. Moreover, $\{\delta_t\}$ is a white noise.*

The proofs of these results are straightforward and thus are omitted.

4. ESTIMATION

To begin, let us first formulate Jørgensen and Song ARMA models from the perspective of time series with measurement errors. That is, X_t may be thought of essentially as the observed process of the underlying true process Y_t with measurement errors δ_t, where process Y_t is a latent Box and Jenkins ARMA(p, q) process. Substitution of Y_t by $X_t - \delta_t$ in the ARMA eqn (10) results in

$$(X_t - \delta_t) - \phi_1(X_{t-1} - \delta_{t-1}) - \cdots - \phi_p(X_{t-p} - \delta_{t-p}) = \varepsilon_t + \psi_1\varepsilon_{t-1} + \cdots + \psi_q\varepsilon_{t-q},$$

or equivalently,

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = \sum_{j=0}^{q} \psi_j \varepsilon_{t-j} + \sum_{k=0}^{p} \phi_k \delta_{t-k},$$

with $\psi_0 = 1$ and $\phi_0 = 1$. Without the loss of generality, we assume that both X_t and ε_t in the above expression have zero mean. Otherwise, noting that

$$\alpha_+ = \sum_{j=0}^{\infty} \alpha_j = \frac{1 + \psi_1 + \cdots + \psi_q}{1 - \phi_1 - \cdots - \phi_p},$$

we immediately have, by Proposition 3,

$$(X_t - \mu_x) - \phi_1(X_{t-1} - \mu_x) - \cdots - \phi_p(X_{t-p} - \mu_x) = \sum_{j=0}^{q} \psi_j(\varepsilon_{t-j} - \mu_\varepsilon) + \sum_{k=0}^{p} \phi_k \delta_{t-k}.$$

The right-hand side of the above equation is indeed the sum of two uncorrelated MA processes of orders q and p, respectively. Clearly the sum is a zero-mean stationary process with ACVF that equals to zero for lags larger than $\max(q, p)$. So, it follows directly from the Wold decomposition theorem (Theorem 5.7.1 or Proposition 3.2.1, Brockwell and Davis, 1987) that the sum is purely non-deterministic and forms a moving-average process of order m with $m \leq \max(q, p)$. Therefore, there exists a white noise ξ_t such that the sum can be expressed as of the form $\sum_{k=0}^{m} \chi_k \xi_{t-k}$. This leads to a Box and Jenkins ARMA equation for the X_t process,

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = \xi_t + \chi_1 \xi_{t-1} + \cdots + \chi_m \xi_{t-m}, \qquad (12)$$

where X_t and ξ_t have zero means (also see Granger and Morris, 1976, for similar results).

Estimating parameters in eqn (12) is then straightforward by applying for the existing methods such as Yule–Walker estimation, Burg's algorithm, the innovations algorithm, or the quasi-likelihood estimation developed for the Box and Jenkins ARMA models (e.g. Brockwell and Davis, 1987).

The Yule–Walker estimation approach provides the method of moments estimators of the parameters ϕ_i, $i = 1, \ldots, p$ in an AR(p) model. That is, the $\hat{\phi}_i, i = 1, \ldots, p$ are the solution to the following system of equations:

$$\begin{pmatrix} \rho_0 & \rho_x(1) & \rho_x(2) & \cdots & \rho_x(p-1) \\ \rho_x(1) & \rho_0 & \rho_x(1) & \cdots & \rho_x(p-2) \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \rho_x(p-1) & \rho_x(p-2) & \rho_x(p-3) & \cdots & \rho_0 \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_P \end{pmatrix} = \begin{pmatrix} \rho_x(1) \\ \rho_x(2) \\ \vdots \\ \rho_x(p) \end{pmatrix},$$

where $\rho_0 = \omega^{-1}$ given in eqn (11). Since the Yule–Walker estimation uses only the first two sample moments of a time-series data, the resulting estimators will be robust to the misspecification of the marginal distribution. Moreover, this approach is computationally simple and hence desirable in dealing with large time series data such as high-frequency time series arising from financial applications, where typically millions of tick-by-tick records are collected. It is known that the method-of-moments estimators are usually not very efficient, and when estimation efficiency is of interest, a likelihood-based estimation approach is usually appealing. Moreover, when both autoregressive and moving average components are present, the quasi-likelihood estimation becomes convenient to obtain parameter estimates.

In the context of the Jørgensen and Song ARMA models, it is difficult to obtain the explicit expression of the likelihood function, unless the marginal distribution is normal. On the contrary, for the Box and Jenkins ARMA models, the quasi-likelihood estimation has been widely used to obtain the estimates of the model parameters. With the connection to the Box and Jenkins ARMA model by eqn (12), it seems natural to adopt the quasi-likelihood approach for the parameter estimation in the Jørgensen and Song ARMA models. Note that the quasi-likelihood approach in the Box and Jenkins ARMA models uses the

multivariate normal distribution to construct the so-called quasi-likelihood function, which essentially requires both ACF and the first two moments of the assumed marginal distribution. In other words, the mean–variance relation, $\mathrm{var}(X_t) = \varphi^2 V(\mu)$, that largely characterizes an ED margin takes part in the construction of the variance–covariance matrix of the data needed in the specification of the quasi-likelihood estimation. Here $\varphi^2 = 1/\lambda$ is the dispersion parameter and $V(\cdot)$ is the variance function. It is known that $V(\mu) = \mu$ for Poisson distribution and $V(\mu) = \mu^2$ for gamma distribution.

Now let $\Gamma = \Gamma(\phi_1, \ldots, \phi_p, \chi_1, \ldots, \chi_m)$ denote the autocorrelation matrix given by a Jørgensen and Song ARMA model. Then, the autocovariance matrix will be $\Sigma = \varphi^2 V(\mu)\Gamma$, and moreover the log quasi-likelihood function is

$$\ell_q(\Gamma, \mu, \varphi^2) \propto -\tfrac{1}{2}\ln|\Sigma| - \tfrac{1}{2}(\mathbf{X} - \boldsymbol{\mu})^{\mathrm{T}}\Sigma^{-1}(\mathbf{X} - \boldsymbol{\mu}), \tag{13}$$

where $\mathbf{X} = (X_1, \ldots, X_n)^{\mathrm{T}}$ is the vector of the observed time series and $\boldsymbol{\mu} = \mu\mathbf{1}$ with $\mathbf{1}$ being an n-dimensional vector of all ones. For a stationary time series, the mean parameter μ is constant over time, so we use the sample mean $\bar{X}$ to estimate μ, which is a consistent estimator with stable numerical performance in either short ($n = 100$) or long ($n = 350$) time series, shown in our simulation studies. This leads to a modified version of the quasi-likelihood function:

$$\tilde{\ell}_q(\Gamma, \varphi^2) \propto -\tfrac{1}{2}\ln|\tilde{\Sigma}| - \tfrac{1}{2}(\mathbf{X} - \bar{X}\mathbf{1})^{\mathrm{T}}\tilde{\Sigma}^{-1}(\mathbf{X} - \bar{X}\mathbf{1}), \tag{14}$$

where $\tilde{\Sigma} = \varphi^2 V(\bar{X})\Gamma$, and the estimates are obtained by maximizing the $\tilde{\ell}_q$. Note that the quasi-likelihood allows us to estimate the dispersion parameter φ^2 together with the other parameters in Γ, which is especially useful to deal with overdispersed data.

One important feature of the quasi-likelihood inference is that the required maximization can be conveniently carried out by using the existing numerical routines for the Box and Jenkins ARMA model with the normal margin. Consider the Pearson residuals $Z_t = (X_t - \bar{X})/\sqrt{V(\bar{X})}, t = 1, \ldots, n$, and let $\Lambda = \varphi^2\Gamma(\phi_1, \ldots, \phi_p, \chi_1, \ldots, \chi_m)$. Then, the quasi-likelihood function (14) can be rewritten as:

$$\tilde{\ell}_q(\Gamma, \varphi^2) \propto -\tfrac{1}{2}\ln|\Lambda| - \tfrac{1}{2}\mathbf{Z}^{\mathrm{T}}\Lambda^{-1}\mathbf{Z}, \tag{15}$$

where $\mathbf{Z} = (Z_1, \ldots, Z_n)^{\mathrm{T}}$ is the vector of Pearson residuals given under the assumed marginal distribution (effectively under the mean–variance relation). This function (15) is in fact coincident with the quasi-likelihood function derived from the following Box and Jenkins ARMA model,

$$Z_t - \phi_1 Z_{t-1} - \cdots - \phi_p Z_{t-p} = \zeta_t + \chi_1\zeta_{t-1} + \cdots + \chi_m\zeta_{t-m}, \tag{16}$$

with white noise $\zeta_t \sim N(0, \varphi_\zeta^2)$. S-Plus function *arima.mle*() can be readily applied to obtain the estimates of $\phi_1, \ldots, \phi_p, \chi_1, \ldots, \chi_m$ and φ_ζ^2.

For the purpose of forecasting for X_t-process, we need not know the estimates of the 'old' parameters $\psi_j, j = 1, \ldots, q$ and σ_ε^2, where only future values of X_t-process and their associated confidence limits are the target. Such a forecasting

can be carried out by simply utilizing eqn (16). In some cases where the thinning coefficients $\{\alpha_j\}$ need to be determined in order to understand the marginal properties of a process, estimation for the parameters $\psi_j, j = 1, \ldots, q$ is necessary. This may be done by using the following relation between the two ACVFs of the new and old MA processes,

$$\gamma_W(h) = \gamma_U(h) + \gamma_V(h), \quad h = 0, \ldots, q + 1,$$

where the process of the sum

$$W_t = U_t + V_t, \quad \text{with} \quad U_t = \sum_{j=0}^{q} \psi_j \varepsilon_{t-j} \quad \text{and} \quad V_t = \sum_{k=0}^{p} \phi_k \delta_{t-k}.$$

Their ACVFs are given, respectively, by

$$\gamma_W(h) = \begin{cases} \sigma_\xi^2 \sum_{j=0}^{m-|h|} \chi_j \chi_{j+|h|}, & |h| \leq m \\ 0, & |h| > m, \end{cases}$$

$$\gamma_U(h) = \begin{cases} \sigma_\varepsilon^2 \sum_{j=0}^{q-|h|} \psi_j \psi_{j+|h|}, & |h| \leq q \\ 0, & |h| > q, \end{cases}$$

and

$$\gamma_V(h) = \begin{cases} \sigma_\delta^2 \sum_{j=0}^{p-|h|} \phi_j \phi_{j+|h|}, & |h| \leq p \\ 0, & |h| > p. \end{cases}$$

In principle, estimates of $\psi_j, j = 1, \ldots, q$, $\sigma_\xi^2, \sigma_\varepsilon^2$ and σ_δ^2 can be obtained via the above system of equations with given $\hat{\varphi}_\zeta^2$ and $\hat{\phi}_j, j = 1, \ldots, p$ and $\hat{\chi}_j, j = 1, \ldots, m$. A general algorithm for an explicit solution seems to be difficult to establish. However, the solution for given cases, especially of low orders, is relatively easy to obtain.

5. AR(1) PROCESS

We now focus on the AR(1) process with non-normal margins and use this special case to demonstrate the estimation theory through a simulation study.

When $\phi(z) = 1 - \phi z$ with $\phi \in (0, 1)$, $\alpha_j = \phi^j, j = 0, 1, \ldots$, and therefore the corresponding Jørgensen and Song MA(∞) takes the form

$$X_t = \varepsilon_t + \sum_{j=1}^{\infty} A_{t,j}(\varepsilon_{t-j}; \phi^j)$$

where the marginal distribution of thinning operator $A_{t,j}(\varepsilon_{t-j}; \phi^j)$ is $ED^*(\theta, (1 - \phi)\phi^j \lambda)$ and $\varepsilon_t \sim ED^*(\theta, (1 - \phi)\lambda)$.

It follows immediately that $X_t = Y_t + \delta_t$ where the projection process is of the Box and Jenkins AR(1) form

$$Y_t = \phi Y_{t-1} + \varepsilon_t,$$

and δ_t is a white noise with zero mean and variance equal to $\sigma_\varepsilon^2 \phi / (1 - \phi^2)$. It is noted that the white noise variation is monotonically increasing in ϕ; it drops down to zero when ϕ approaches zero and diverges to ∞ when ϕ tends to 1.

With the anticorruption coefficient $\omega = 1/(1 + \phi)$, it is easy to see that the ACVF of the AR(1) is

$$\gamma_x(0) = \frac{\sigma_\varepsilon^2}{1 - \phi}, \quad \gamma_x(h) = \frac{\phi^h \sigma_\varepsilon^2}{1 - \phi^2}, \quad h \geq 1$$

and the ACF of the process is

$$\rho_x(1) = \frac{\phi}{1 + \phi}, \quad \rho_x(h) = \phi^{h-1} \rho_x(1) = \frac{\phi^h}{1 + \phi}, \quad h \geq 2.$$

It follows from the Yule–Walker equation that the consistent estimators of ϕ and σ_ε^2 are given by

$$\hat{\phi} = \frac{\hat{\rho}_x(1)}{1 - \hat{\rho}_x(1)} \quad \text{and} \quad \hat{\sigma}_\varepsilon^2 = (1 - \hat{\phi})\hat{\gamma}_x(0), \tag{17}$$

respectively, where both sample ACVF $\hat{\gamma}_x(h)$ and sample ACF $\hat{\rho}_x(h)$ are strongly consistent for each given lag h because of the ergodicity and stationarity of the process.

There is a constraint in the application of the Yule–Walker equation method (17) to obtain $\hat{\phi}$. That is, the sample autocorrelation $\hat{\rho}_x(1)$ at lag 1 has to be bounded by $1/2$ in order to ensure $\hat{\phi} < 1$. When a time series is highly autocorrelated, $\hat{\rho}_x(1)$ may exceed such an upper bound, possibly because of the problem that the sample variance underestimated the variance parameter $\gamma_x(0)$. According to our simulation studies in this section, this underestimation problem appeared frequently in cases of short time series ($n = 100$), but disappeared when time series was long ($n = 350$). To overcome this problem, we propose an alternative estimator that takes the ratio of the first two sample autocorrelations $\hat{\rho}_x(h), h = 1, 2$, which hence dodges the use of the sample variance in the estimation, namely,

$$\hat{\phi} = \frac{\hat{\rho}_x(2)}{\hat{\rho}_x(1)} = \frac{\hat{\gamma}_x(2)}{\hat{\gamma}_x(1)}. \tag{18}$$

Estimates of ϕ using the two methods (17) and (18), referred to as YW1 and YW2, respectively, are reported and compared in Tables I and II in simulation studies, for short time series with both Poisson and gamma margins.

In the meanwhile, from Section 4, we may write the X_t-process to be a Box and Jenkins ARMA(1,1) process

$$X_t - \phi X_{t-1} = \xi_t + \chi \xi_{t-1}$$

TABLE I

THE SAMPLE AVERAGES AND STANDARD DEVIATIONS IN PARENTHESES OF THE QUASI-LIKELIHOOD (QL) ESTIMATOR AND THE TWO YULE–WALKER (YW) ESTIMATORS OVER 100 REPLICATIONS FOR POISSON AR(1) SERIES OF LENGTH $n = 100$

| | Mean parameter μ | | | | | | | | |
| | 5 | | | 10 | | | 20 | | |
True ϕ	QL	YW2	YW1	QL	YW2	YW1	QL	YW2	YW1
0.30	0.22 (0.39)	0.28 (0.86)	0.28 (0.17)	0.17 (0.37)	0.35 (1.30)	0.29 (0.17)	0.20 (0.37)	0.25 (1.10)	0.29 (0.15)
0.50	0.39 (0.28)	0.41 (0.40)	0.50 (0.22)	0.34 (0.31)	0.41 (0.39)	0.47 (0.20)	0.39 (0.26)	0.46 (0.57)	0.47 (0.23)
0.70	0.61 (0.21)	0.66 (0.30)	0.62 (0.31)	0.59 (0.25)	0.68 (0.30)	0.62 (0.30)	0.59 (0.24)	0.66 (0.28)	0.71 (0.32)
0.90	0.78 (0.18)	0.93 (0.40)	0.70 (0.34)	0.80 (0.15)	0.91 (0.32)	0.71 (0.43)	0.79 (0.18)	0.88 (0.39)	0.73 (0.38)

TABLE II

THE SAMPLE AVERAGES AND STANDARD DEVIATIONS IN PARENTHESES OF THE QUASI-LIKELIHOOD (QL) ESTIMATOR AND THE TWO YULE–WALKER (YW) ESTIMATORS OVER 100 REPLICATIONS FOR GAMMA AR(1) SERIES OF LENGTH $n = 100$

| | Index parameter λ | | | | | | | | |
| | 2 | | | 5 | | | 10 | | |
True ϕ	QL	YW2	YW1	QL	YW2	YW1	QL	YW2	YW1
0.30	0.16 (0.39)	0.23 (1.76)	0.31 (0.18)	0.13 (0.38)	0.15 (0.71)	0.30 (0.17)	0.14 (0.39)	−0.08 (1.84)	0.29 (0.15)
0.50	0.34 (0.28)	0.39 (0.36)	0.54 (0.26)	0.41 (0.28)	0.54 (0.95)	0.61 (0.27)	0.39 (0.27)	0.40 (0.30)	0.61 (0.25)
0.70	0.62 (0.19)	0.66 (0.19)	0.96 (0.43)	0.63 (0.14)	0.66 (0.17)	1.02 (0.42)	0.62 (0.18)	0.63 (0.19)	1.24 (0.49)
0.90	0.82 (0.16)	0.85 (0.16)	1.56 (0.77)	0.84 (0.09)	0.83 (0.09)	2.03 (1.01)	0.86 (0.07)	0.86 (0.08)	2.75 (1.20)

where ξ_t is a white noise with variance σ_ξ^2. Let $Z_t = (X_t - \bar{X})/\sqrt{V(\bar{X})}$. Then the estimates of parameters ϕ, χ and φ^2 can be obtained by maximizing the quasi-likelihood function arising from the following Box and Jenkins ARMA(1, 1) model:

$$Z_t - \phi Z_{t-1} = \zeta_t + \chi \zeta_{t-1}, \tag{19}$$

where $\zeta_t \sim N(0, \varphi_\zeta^2)$. Because of the fact that a linear transformation on Z_t does not change ACF, it is easy to see that

$$\hat{\varphi}^2 = \frac{(1 + \hat{\chi}^2)(1 + \hat{\phi})}{1 + \hat{\phi} - \hat{\phi}^2 + \hat{\phi}^3} \hat{\varphi}_\zeta^2,$$

and moreover, $\hat{\sigma}_\xi^2 = \hat{\varphi}_\zeta^2 V(\bar{X})$.

Simulation studies based on the AR(1) model with Poisson and gamma margins were conducted to illustrate the numerical computation and to compare two estimation methods, YW1/YW2 and quasi-likelihood. The focus of the comparison was only on estimation of the autocorrelation parameter ϕ. In connection with $ED^*(\theta, \lambda)$, the Poisson case had the dispersion parameter $\varphi^2 = 1/\lambda = 1$ and the mean parameter $\mu = e^\theta$. For a given value of parameter $\phi \in \{0.3, 0.5, 0.7, 0.9\}$, we generated an AR(1) time series of n counts, $n \in \{100, 350\}$ at different mean parameters $\mu \in \{5, 10, 20\}$, according to the Jørgensen and Song model (8) on the basis of 100 terms:

$$X_t = \varepsilon_t + \sum_{j=1}^{100} A_{t,j}(\varepsilon_{t-j}; \phi^j), \tag{20}$$

because ϕ^j, $j > 100$ were virtually zero. For each simulated series, we computed two Yule–Walker estimates based on eqns (17) and (18), and quasi-likelihood estimate based on eqn (19), respectively. With every chosen setting of parameter combination, these procedures were replicated 100 times, and the sample averages and standard deviations of the resulting 100 estimates were tabulated in Tables I and III.

What we learned from Table I was that the performances of both the Yule–Walker and quasi-likelihood methods were not very stable for short time series ($n = 100$) of counts. We observed that as far as bias is concerned, the YW1 mostly seemed to perform better than the other two, YW2 and QL, with small-or medium-sized ϕ, whereas the YW2 mostly appeared to outperform the other two, YW1 and QL, with large-sized ϕ. As far as the standard deviation is concerned, the quasi-likelihood method apparently produced better estimation precision than the other two, YW1 and YW2, with large-sized ϕ; but with small-or medium-sized ϕ the YW1 was actually the best.

When the length of time series increased to $n = 350$, as shown in Table III where the YW2 was not considered, both YW1 and quasi-likelihood methods worked very well with little bias. It was clear that the quasi-likelihood was more efficient than the YW method, with smaller standard deviations especially with

TABLE III

THE SAMPLE AVERAGES AND STANDARD DEVIATIONS IN PARENTHESES OF THE QUASI-LIKELIHOOD (QL)
ESTIMATOR AND THE YULE–WALKER (YW1) ESTIMATOR OVER 100 REPLICATIONS FOR POISSON AR(1)
SERIES OF LENGTH $n = 350$

	Mean parameter μ					
	5		10		20	
True ϕ	QL	YW1	QL	YW1	QL	YW1
0.30	0.25 (0.23)	0.26 (0.22)	0.28 (0.24)	0.31 (0.24)	0.29 (0.21)	0.28 (0.27)
0.50	0.50 (0.13)	0.50 (0.13)	0.47 (0.14)	0.48 (0.15)	0.46 (0.14)	0.47 (0.14)
0.70	0.66 (0.10)	0.66 (0.12)	0.67 (0.10)	0.68 (0.13)	0.67 (0.09)	0.68 (0.11)
0.90	0.88 (0.05)	0.86 (0.11)	0.88 (0.04)	0.89 (0.10)	0.88 (0.05)	0.89 (0.11)

large ϕ. Our conclusion was that for long integer-valued series like $n = 350$ the quasi-likelihood method was recommended, while for short integer-valued series like $n = 100$, the YW1 was recommended for small-or medium-sized ϕ and the YW2 for large-sized ϕ.

A similar simulation study was conducted to compare the Yule-Walker and quasi-likelihood methods for the case of gamma margins. The gamma $ED^*(\theta, \lambda)$ gave the mean $\mu = 1/\theta$, which was fixed as 1 in the simulation, and the dispersion parameter $\varphi^2 = 1/\lambda$ where $\lambda \in \{2, 5, 10\}$. Similarly, we took $\phi \in \{0.3, 0.5, 0.7, 0.9\}$ and generated AR(1) series using the finite truncation representation (20). Again, 100 replications were run, and related summary statistics were listed in Tables II and IV.

Similar conclusions can be drawn from this second simulation study. Based on Table II, we learned that when ϕ was < 0.5, the YW1 performed better than the other two for short time series, as it gave the smallest bias and standard deviation. However, when ϕ was ≥ 0.7, the YW2 and the quasi-likelihood were essentially comparable, and they performed significantly better than the YW1. It is important to point out that the application of the YW1 method should be very cautious as it may give an estimate of ϕ exceeding the upper bound 1, as indicated by the boldfaced numbers in Table II.

TABLE IV

THE SAMPLE AVERAGES AND STANDARD DEVIATIONS OF THE QUASI-LIKELIHOOD (QL) ESTIMATES AND
THE YULE–WALKER (YW1) ESTIMATES WITH 100 REPLICATIONS FOR GAMMA AR(1) SERIES OF
LENGTH $n = 350$

	Index parameter λ					
	2		5		10	
True ϕ	QL	YW1	QL	YW1	QL	YW1
0.30	0.24 (0.23)	0.25 (0.24)	0.27 (0.22)	0.28 (0.26)	0.27 (0.21)	0.28 (0.22)
0.50	0.46 (0.16)	0.50 (0.15)	0.48 (0.11)	0.49 (0.13)	0.47 (0.13)	0.48 (0.13)
0.70	0.67 (0.09)	0.68 (0.11)	0.67 (0.06)	0.68 (0.07)	0.69 (0.06)	0.70 (0.07)
0.90	0.88 (0.04)	0.89 (0.06)	0.88 (0.04)	0.88 (0.05)	0.89 (0.04)	0.89 (0.04)

However, when the length of time series increased to $n = 350$, such an problematic estimation given by the YW1 disappeared. Table IV summarizes the results of the simulation study based on long gamma time series.

From Table IV, both methods performed very well, and the quasi-likelihood method appeared to be slightly more efficient than the YW1 method. Again, in this case, bias became marginal, which implied that both methods provided consistent estimates of ϕ when the time series was long.

6. ANALYSIS OF MYOCLONIC SEIZURES

We now apply the proposed model to analyze the daily myoclonic seizure counts. The data have previously been analysed by Le *et al.* (1992) using a two-state Markov mixture model and MacDonald and Zucchini (1997) using a hidden Markov model. (Also see Albert, 1991 for a similar analysis of epileptic seizure counts using a two-state Markov mixture model.) Hopkins *et al.* (1985) pointed out that the behaviour of an individual's susceptibility to seizures may be addressed reasonably by a Markov process. Franke and Seligmann (1993) used a Poisson AR(1) analogue based on the binomial thinning to model time series of seizure counts. However, their method was developed based only on the first-order Markov process, and the extension of their method to high-order processes seemed to be intricate. In contrast, our method is flexible and applicable to different ARMA models, with little restriction on the orders of the chosen models. This flexibility gives rise to a great deal of ease in fitting a non-normal time-series data to various models and consequently in selecting suitable models for data analysis. We used the S-Plus function *arima.mle*() to obtain the parameter estimates in this data analysis.

Assume that the time series of 204 daily seizure counts of an individual follows a Jørgensen and Song AR(p) model. Refer to Figure 1 of Le *et al.* (1992) for the time-series plot of the raw data. Based on sample ACF and partial ACF of the data, four candidate AR models with orders $p = 1, 2, 3$ and 4 were chosen to fit the data, and the quasi-likelihood approach enabled us to compute Akaike's information criterion (AIC) for the corresponding Box and Jenkins ARMA(p, p). The four AIC values were 617.2786, 613.6754, 608.4537 and 607.9489, respectively, starting from $p = 1$. In the light of these AIC values, the Jørgensen and Song AR(4) model having the smallest AIC value was selected for the further analysis.

The condition of causality was confirmed since the four roots of the resulting AR polynomial $\phi(z) = 1 - 0.098z - 0.507z^2 - 0.547z^3 + 0.258z^4$ all stayed outside the unitary circle, as shown in Figure 1(a). This implied that the coefficients $\alpha_j, j \geq 0$ existed, which were obtained by the recursions (7). Moreover, the anticorruption coefficient was estimated as $\hat{\omega} = 0.354$ based on eqn (11), where we used 500 $\hat{\alpha}_j$ terms to calculate $\sum \alpha_j^2$. Note that these 500 α_j's were sufficient to attain the desirable precision.
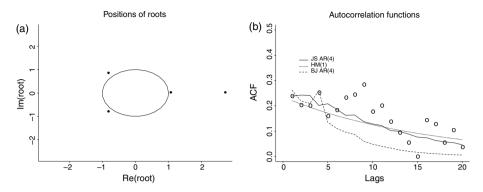
FIGURE 1. (a) The positions of the four roots from the projection AR(4) models relative to the unitary circle. The roots are indicated by dots. (b) The circles indicate the sample ACF of the myoclonic seizure data, the solid line stands for the ACF of the Jørgensen and Song AR(4) model, the dotted line represents the ACF of the two-state hidden Markov model, and the broken line denotes the ACF of the Box and Jenkins AR(4) model.

To compare our approach with the two-state Poisson-hidden Markov model used by MacDonald and Zucchini (1997, Sect. 4.3) to fit the data, we plotted the autocorrelation functions of the three fitted models in Figure 1(b) over the first 20 lags, together with the sample autocorrelation function indicated by circles. Overall, the two ACFs of the Jørgensen and Song AR(4) (the solid line) and MacDonald and Zucchini's hidden Markov model (the dotted line) are fairly comparable, and the small difference at the first five lags between them is partly caused by the inclusion of the moving average structure induced from the residual δ_t (or the measurement error) term for the parameter estimation in the Jørgensen and Song AR(4) model.

In addition, the quasi-likelihood estimate of the scale parameter φ^2 from the Jørgensen amd Song AR(4) model was 1.15, indicating no overdispersion. This implied that the Poisson margin assumption seemed to be suitable for the data.We also obtained $\hat{\sigma}_\delta^2 = 0.597$, so the estimated standard deviation between the observed process and the projection process was $\sqrt{\hat{\sigma}_\delta^2} = 0.773$, which seemed to be moderate. When the residual term δ_t was ignored and the data were directly fitted to the Box and Jenkins AR(4) model, the resulting ACF of this model was plotted as the broken line in Figure 1(b). This ACF seemed only able to capture the sample ACF at the first five lags, and it then quickly departed away from the sample ACF at the rest of lags. This suggests that recognizing data nature (which is the count data in this example) and utilizing such information in the analysis can help us to better capture the overall pattern of the ACF.

Although the AIC is useful to select a desirable model from a pool of candidate models, some other aspects of model assumptions should be checked in a data analysis. Because of the connection between the Jørgensen and Song models and the Box and Jenkins models, some available model diagnostics with the Box and Jenkins ARMA models may be applied for the Jørgensen and Song

models. In this data analysis example, we used the S-Plus function *arima.diag*() to confirm the stationarity and the order p of the AR(4) model through the Ljung–Box test.

7. DISCUSSION

Both Yule–Walker estimation and the quasi-likelihood estimation proposed in this paper are carried out by only using the first two moments of the Jørgensen and Song ARMA(p, q) processes. As pointed out above, the quasi-likelihood approach requires a full specification of the autocovariance matrix that pertains to the mean–variance relation, which makes the connection of the quasi-likelihood approach to the assumed marginal distribution. In contrast, a full likelihood approach relies on the joint distributions and hence on entirely the assumed marginal distribution, not just its first two moments. Since the Jørgensen and Song models are defined by linear processes in a unified framework, the closed-form expressions of the joint density functions are generally unavailable, unless the margin is the normal distribution. Thus, the full likelihood method seems to be feasible only for the model with the normal margin. More studies are necessary for improving the estimation efficiency.

In general, the quasi-likelihood estimation is the method recommended for inference in the Jørgensen and Song models, because it has some useful properties as far as data analysis is concerned. First, this method works for general ARMA models, as opposed to that Yule–Walker method works mainly for AR models; second, the quasi-likelihood method allows us to jointly estimate the dispersion parameter and autocorrelation parameters; third, this method enables us to carry out the model selection and some model diagnostics. In addition, the numerical calculations associated with this method are straightforward and can be done using the existing software packages.

We found from the simulation studies that the marginal mean μ_x can be reasonably estimated by the sample mean $\bar{X}$, regardless of the length of time series, but this was not the case for the marginal variance parameter $\gamma_x(0)$ especially when time series is short. Instead of using the sample variance to estimate $\gamma_x(0)$, a better way is first to estimate the dispersion parameter via the quasi-likelihood approach, and then to estimate the $\gamma_x(0)$ via the mean–variance relation for a given ED model.

Checking model assumptions is an important part of statistical analysis. There are several aspects of the Jørgensen and Song models that can be checked by certain model diagnostics. In the analysis of the time series of seizure counts, the quasi-likelihood method produced an estimate of the overdispersion parameter that was used to select a margin between Poisson and negative binomial distributions, respectively corresponding to the absence and the presence of overdispersion. For a positive continuous time series, we may start with the Tweedie class of distributions (Jørgensen, 1997, Ch. 4) with the mean–variance

relation $\text{var}(X_t) = \varphi^2 \mu_x^r, r \in R \setminus (0, 1)$, and then plot the AIC values against the shape parameter r in a given interval range. Technically, the Tweedie model (or the value of shape parameter r) with the smallest AIC would be selected as the margin in a further analysis of the data.

To address the selection of the orders for the autoregression and moving average structure, we may follow the current time series literature. This is because the quasi-likelihood method essentially utilizes the Box–Jenkins' ARMA model in the parameter estimation. For instance, the S-Plus function *arima.diag*() provides a few model diagnostics that may be used for this purpose, as shown in the data analysis example of this paper.

NOTE

Corresponding author: Peter X.-K. Song, Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1. E-mail: song@uwaterloo.ca

REFERENCES

ALBERT, P. S. (1991) A two-state Markov mixture model for a time series of epileptic seizure counts. *Biometrics* 47, 1371–81.

ASHLEY, R. and VAUGHAN, D. (1986) Measuring measurement error in economic time series. *Journal of Business and Economic Statistics* 4, 95–103.

BENJAMIN, M. A., RIGBY, R. A. and STASINOPOULOS, D. M. (2003) Generalized autoregressive moving average models. *Journal of the American Statistical Association* 98, 214–23.

BOX, G. E. P. and JENKINS, G. M. (1976) *Time Series Analysis: Forecasting and Control*, revised edition. Holden-Day, San Francisco, CA.

BROCKWELL, P. J. and DAVIS, R. A. (1987) *Time Series: Theory and Methods*. Springer-Verlag, New York.

FRANKE, J. and SELIGMANN, T. (1993) Conditional maximum-likelihood estimates for INAR(1) processes and their application to modelling epileptic seizure counts. In *Developments in Time Series Analysis* (ed. T. SUBBA RAO). Chapman and Hall, London, pp. 310–30.

GRANGER, C. W. J. and MORRIS, M. J. (1976) Time series modelling and interpretation. *Journal of the Royal Statististical Society Series* A 139, 246–57.

HOPKINS, A., DAVIES, P. and DOBSON, C. (1985) Mathematical models of patterns of seizures: their use in the evaluation of drugs. *Archives of Neurology* 42, 463–7.

Joe, H. (1996) Time series models with univariate margins in the convolution-closed infinitely divisible class. *Journal of Applied Probability* 33, 664–77.

Jørgensen, B. (1997) *The Theory of Dispersion Models*. Chapman and Hall, London.

Jørgensen, B. and Song, P. X.-K. (1998) Stationary time-series models with exponential dispersion model margins. *Journal of Applied Probability* 35, 78–92.

Jørgensen, B., Lundbye-Christensen, S., Song, P. X.-K. and Sun, L. (1999) A state space model for multivariate longitudinal data. *Biometrika* 86, 169–81.

Le, N. D., Leroux, B. G. and Puterman, M. L. (1992) Exact likelihood evaluation in a Markov mixture model for time series of seizure counts. *Biometrics* 48, 317–23.

Lewis, P. A. W., McKenzie, E. and Hugus, D. K. (1989) Gamma processes. *Communications in Statistics – Stochastic Models* 5(1), 1–30.

Li, W.K. (1994) Time series models based on generalized linear models: some further results. *Biometrics* 50, 506–11.

MacDonald, I. L. and Zucchini, W. (1997) *Hidden Markov and Other Models for Discrete-valued Time Series*. Chapman and Hall, London.

Steutel, F. W. and van Harn, K. (1979) Discrete analogues of self-decomposability and stability. *Annals of Probability* 7, 893–9.

Stout, W. F. (1974) *Almost Sure Convergence*. Academic Press, New York.

Zeger, S. L. (1988) A regression model for time series of counts. *Biometrika* 75, 822–35.

Zeger, S. L. and Qaqish, B. (1988) Markov regression models for time series: a quasi-likelihood approach. *Biometrics* 44, 1019–32.