Asymptotic Quasi-likelihood Based on Kernel Smoothing for Nonlinear and Non-Gaussian State-Space Models

Raed Alzghool,Yan-Xia Lin *

Abstract—**This paper considers parameter estimation for nonlinear and non-Gaussian state-space models with correlation. We propose an asymptotic quasi-likelihood (AQL) approach which utilises a nonparametric kernel estimator of the conditional variance covariances matrix $\mathbf{\Sigma}_t$ to replace the true $\mathbf{\Sigma}_t$ in the standard quasi-likelihood. The kernel estimation avoids the risk of potential miss-specification of $\mathbf{\Sigma}_t$ and thus make the parameter estimator more robust. This has been further verified by empirical studies carried out in this paper.**

Keywords: asymptotic quasi-likelihood (AQL), kernel smoothing, martingale, Quasi-likelihood (QL), State-Space Models (SSM)

1 Introduction

The class of state space models (SSM) provides a flexible framework for describing a wide range of time series in a variety of disciplines. For extensive discussion on SSM and their applications see Harvey [10] and Durbin and Koopman [8]. A state-space model can be written as

$$y_t = f_1(\alpha_t, \theta) + h_1(y_{t-1}, \theta)\epsilon_t, \quad t = 1, 2, \cdots, T \qquad (1)$$

where $y_1, \ldots, y_T$ represent the time series of observations; θ is an unknown parameter that needs to be estimated; $f_1(.)$ is a known function of state variable α_t and θ; and $\{\epsilon_t\}$ are uncorrelated disturbances with $E_{t-1}(\epsilon_t) = 0$, $Var_{t-1}(\epsilon_t) = \sigma_\epsilon^2$; in which E_{t-1}, and Var_{t-1} denote conditional mean and conditional variance associated with past information updated to time t-1 respectively. State variables $\alpha_1, \ldots, \alpha_T$ are unobserved and satisfy the following model

$$\alpha_t = f_2(\alpha_{t-1}, \theta) + h_2(\alpha_{t-1}, \theta)\eta_t, \quad t = 1, 2 \cdots, T, \qquad (2)$$

where $f_2(.)$ is a function of past state variables and θ; $\{\eta_t\}$ are uncorrelated disturbances with $E_{t-1}(\eta_t) = 0$, $Var_{t-1}(\eta_t) = \sigma_\eta^2$. $h_1(.)$ and $h_2(.)$ are unknown functions.

One special application that we will consider in detail is the case where the time series $y_1, \ldots, y_T$ consist of counts. Here, it might be plausible to model y_t by a Poisson distribution. Models of this type have been used for rare diseases, (Zeger [26]; Chan and Ledolter [5]; Davis, Dunsmuir and Wang [6]).

Another noteworthy application of the SSM that we will consider is Stochastic Volatility Model (SVM), a frequently used model for returns of financial assets. Applications, together with estimation for SVM, can be found in Jacquier, et al [17]; Briedt and Carriquiry [4]; Harvey and Streible [11]; Sandmann and Koopman [24]; Pitt and Shepard [22].

There are several approaches in the literature for estimating the parameters in SSMs by using the maximum likelihood method when the probability structure of underlying model is normal or conditional normal. Durbin and Koopman ([9], [8]) obtained accurate approximation of the log-likelihood for Non-Gaussian state space models by using Monte Carlo simulation. The log-likelihood function is maximised numerically to obtain estimates of unknown parameters. Kuk [18] suggested an alternative class of estimate models based on conjugate latent process and applied it to approximate the likelihood of a time series model for count data. To overcome the complex likelihoods of a time series model with count data, Chan and Ledolter [5] proposed the Monte Carlo EM algorithm that uses a Markov chain sampling technique in the calculation of the expectation in the the E-step of the EM algorithm. Davis and Rodriguez-Yam [7] proposed an alternative estimation procedure which is based on an approximation to the likelihood function. Alzghool and Lin [2] proposed quasi-likelihood (QL) approach for estimation of state space models without full knowledge on the probability structure of relevant state-space system. The QL method relaxes the distributional assumptions and only assumes the knowledge on the first two conditional moments of y_t and α_t associated past information. This weaker assumption makes the QL method widely applicable and become a popular method of estimation. A comprehensive review on the QL method is available in Heyde [16]. A limitation of the QL is that in practice, the conditional second moments of of y_t and α_t might not available. In this paper, we suggest an alternative approach, AQL approach, combining with kernel method

*School of Mathematics and Applied Statistics University of Wollongong NSW 2500, Australia Emails: raed@uow.edu.au, yanxia@uow.edu.au

treatment. This AQL approach provides an alternative method of parameter estimation when unknown form of heteroscedasticity is presented.

This paper is structured as follows. In Section 2, the asymptotic quasi-likelihood based on kernel smoothing is introduced. we apply the AQL approach to SSMs in Section 3. Section 4 report simulation results and covers numerical implementation. An analysis on a real data set by the AQL method is given in Section 5. A summary is given in Section 6.

2 Asymptotic quasi-likelihood approach

Consider the following qth-order markovian process model,

$$\mathbf{y}_t = \mathbf{m}_t(y_{t-1}, \ldots, y_{t-q}; \theta) + \delta_t, \quad t = 1, 2, \cdots, \tag{3}$$

where $\mathbf{y}_t$, $\mathbf{m}_t(\theta)$, and δ_t are m-dimension random vectors; $\mathbf{m}_t$ is $\mathcal{F}_{t-1}$ measurable; δ_t is a martingale difference associated with $\mathcal{F}_t$, i.e. $E(\delta_t|\mathcal{F}_{t-1})= E_{t-1}(\delta_t) = 0$; $\mathcal{F}_t$ is a σ-field generated by $\{\mathbf{y}_s\}_{s\leq t}$; and θ is the parameter of interest defined in an open parameter space $\Theta \in R^d$.

Given a sample $\{\mathbf{y}_t\}_{t\leq T}$ drawn from (3), if the expression of $E(\delta_t\delta_t'|\mathcal{F}_{t-1})=E_{t-1}(\delta_t\delta_t') = \mathbf{\Sigma}_t$ is known, the standard quasi-score estimating function in estimating function space

$$\mathcal{G}_{\mathbf{T}} = \{\sum_{t=1}^{T} \mathbf{A}_t(\mathbf{y}_t - \mathbf{m}_t(\theta)); \mathbf{A}_t \text{ is } \mathcal{F}_{t-1}\text{-measureable}\}$$

is

$$\mathbf{G}_T^*(\theta) = \sum_{t=1}^{T} \dot{\mathbf{m}}_t(\theta)\mathbf{\Sigma}_t^{-1}(\mathbf{y}_t - \mathbf{m}_t(\theta)) \tag{4}$$

where $\dot{\mathbf{m}}_t(\theta) = \partial\mathbf{m}_t(\theta)/\partial\theta$. Then the quasi-score normal equation is $\mathbf{G}_T^*(\theta) = 0$, whose root is the quasi-likelihood estimate of θ. For a special scenario, if we only consider sub estimating function spaces of $\mathcal{G}_T$, for example,

$$\mathcal{G}^{(t)} = \{\mathbf{A}_t(\mathbf{y}_t - \mathbf{m}_t); \mathbf{A}_t \text{ is } \mathcal{F}_{t-1}\text{-measureable}\} \subset \mathcal{G}_T, t < T,$$

then, the standard quasi-score estimating function in this space is

$$\mathbf{G}_t^*(\theta) = \dot{\mathbf{m}}_t(\theta)\mathbf{\Sigma}_t^{-1}(\mathbf{y}_t - \mathbf{m}_t(\theta)) \tag{5}$$

and $\mathbf{G}_t^*(\theta) = 0$ will give the quasi-likelihood estimator based on the information provided by $\mathcal{G}^{(t)}$. Under certain regularity conditions, the quasi-likelihood estimator is consistency and achieves optimal efficiency within space $\mathcal{G}_{\mathbf{T}}$ (Heyde, [16]). In particular, under Fisher information criterion, the volume of the confidence region for θ produced by the quasi-score estimating function is smaller than that of any other confidence regions derived from any other estimating functions within the same estimating function space (Lin and Heyde, [19]).

The quasi-score estimating functions (4) and (5) rely on the knowledge of $E_{t-1}(\delta_t\delta_t')$. Such knowledge is not always available in practice considering there is only one sample path of the process being observed. To facilitate QL in a situation where $E_{t-1}(\delta_t\delta_t')$ is unknown, Lin [21] introduced a new concept of asymptotic quasi-score estimation function and suggested an approach, called the asymptotic quasi-likelihood (AQL) approach, replacing the exact quasi-likelihood approach. Let $\mathbf{\Sigma}_{t,n}$ be a sequence of $\mathcal{F}_{t-1}$-measurable random matrices converging to $E_{t-1}(\delta_t\delta_t')$ in probability. Then,

$$\mathbf{G}_{T,n}^*(\theta) = \sum_{t=1}^{T} \dot{\mathbf{m}}_t(\theta)\mathbf{\Sigma}_{t,n}^{-1}(\mathbf{y}_t - \mathbf{m}_t(\theta))$$

forms a sequence of asymptotic quasi-score estimating functions. The corresponding roots of $\mathbf{G}_{T,n}^*(\theta) = 0$ forms a sequence of asymptotic quasi-likelihood estimates $\{\theta_{T,n}^*\}$ which converges to θ under certain conditions. Since $\mathbf{G}_{T,n}^*$ has the following property (Lin, [21])

$$\|(E\dot{\mathbf{G}}_T^*)^{-1}(E\mathbf{G}_T^*\mathbf{G}_T^{*\prime})(E\dot{\mathbf{G}}_T^{*\prime})^{-1}$$

$$-(E\dot{\mathbf{G}}_{T,n}^*)^{-1}(E\mathbf{G}_{T,n}^*\mathbf{G}_T^{*\prime})(E\dot{\mathbf{G}}_{T,n}^{*\prime})^{-1}\| \to 0,$$

as $n \to \infty$, this means that the amount of Fisher Information provided by $\mathbf{G}_{\mathbf{T,n}}^*$ will be close to what provided by the standard QL estimating function $\mathbf{G}_T^*$. Thus, $\mathbf{G}_{T,n}^*$ will be able to provide asymptotic efficient estimation for θ through $\{\theta_{T,n}^*\}$. Thus, using asymptotic quasi-score estimating function to obtain asymptotic efficient estimation for θ is an alternative approach to the QL approach when QL estimating function is not available. The main issue in asymptotic quasi-score approach is about the structure of appropriate asymptotic quasi-score sequence of estimating functions. In this paper, we consider using the kernel smoothing estimator of $\mathbf{\Sigma}_t=: Var(\mathbf{y}_t|\mathcal{F}_{t-1})$ to replace $\mathbf{\Sigma}_t$ in the AQL formulation (4) and (5).

Under (3), let $\mathbf{x}_t = (\mathbf{y}_{t-1}, \ldots, \mathbf{y}_{t-q})$ be the lagged value of $\mathbf{y}_t = (y_{1t}, y_{2t}, \ldots, y_{mt})'$. Given an initial estimator of θ, say $\hat{\theta}^{(0)}$, the Nadaraya-Watson (NW) estimator of $\mathbf{\Sigma}_t$ is $\hat{\mathbf{\Sigma}}_{\mathbf{t,n}}$ with elements

$$\hat{\sigma}_n(y_{it}) = \frac{\sum_{s=q+1}^{n} D_{its}(y_{is} - m_{is}(\mathbf{x}_{is}, \hat{\theta}^{(0)}))^2}{\sum_{s=q+1}^{n} D_{its}} \tag{6}$$

$$\hat{\sigma}_n(y_{it}, y_{jt}) = \frac{\sum_{s=q+1}^{n} D_{its}D_{jts}(y_{is} - m_{is})(y_{js} - m_{js})}{\sum_{s=q+1}^{n} D_{its}D_{jts}}, i \neq j \tag{7}$$

where $i, j = 1, 2, \ldots, m$, $D_{its} = K\left(\frac{\mathbf{x}_{it} - \mathbf{x}_{is}}{h}\right)$, $\mathbf{x}_{it} = (y_{i,t-1}, \ldots, y_{i,t-q})$, $\mathbf{x}_{is} = (y_{i,s-1}, \ldots, y_{i,s-q})$ and $K(u) = 0.75^q \prod_{l=1}^{q}[(1 - u_l^2)I_{(-1,1)}u_l]$ is a q-dimensional kernel function of order 2 and h is a smoothing bandwidth such that $h \to 0$ and $nh^q \to \infty$ as $n \to \infty$.

ISBN:978-988-98671-2-6

A comprehensive review of the above NW type kernel estimator including the construction of K and the choice of h is available in (Härdle, [13]; Wand and Jones, [25]) . Härdle et al. [14], Härdle and Tsybakov [15] consider the local linear estimator for volatility function for data from a first order Markov process.

The estimating functions (4) and (5) based on the kernel estimators (6) and (7) become

$$\mathbf{G}^*_{T,n}(\theta) = \sum_{t=1}^{T} \dot{\mathbf{m}}_t(\theta)\hat{\mathbf{\Sigma}}^{-1}_{t,n}(\mathbf{y}_t - \mathbf{m}_t(\theta)) \tag{8}$$

$$\mathbf{G}^*_{t,n}(\theta) = \dot{\mathbf{m}}_t(\theta)\hat{\mathbf{\Sigma}}^{-1}_{t,n}(\mathbf{y}_t - \mathbf{m}_t(\theta)) \tag{9}$$

and the asymptotic quasi-score normal equation are

$$\mathbf{G}^*_{T,n}(\theta) = \sum_{t=1}^{T} \dot{\mathbf{m}}_t(\theta)\hat{\mathbf{\Sigma}}^{-1}_{t,n}(\mathbf{y}_t - \mathbf{m}_t(\mathbf{x}_t;\theta)) = 0. \tag{10}$$

$$\mathbf{G}^*_{t,n}(\theta) = \dot{\mathbf{m}}_t(\theta)\hat{\mathbf{\Sigma}}^{-1}_{t,n}(\mathbf{y}_t - \mathbf{m}_t(\mathbf{x}_t;\theta)) = 0. \tag{11}$$

where

$$\hat{\mathbf{\Sigma}}_{t,n}(\hat{\theta}^{(0)}) = \begin{bmatrix} \hat{\sigma}_n(y_{1t}) & \dots & \hat{\sigma}_n(y_{1t}, y_{mt}) \\ \hat{\sigma}_n(y_{2t}, y_{1t}) & \dots & \hat{\sigma}_n(y_{2t}, y_{mt}) \\ \vdots & \ddots & \vdots \\ \hat{\sigma}_n(y_{mt}, y_{1t}) & \dots & \hat{\sigma}_n(y_{mt}) \end{bmatrix}.$$

To solve the above asymptotic quasi-score normal equation, say (10) for example, an iterative procedure can be adapted. It can start from the OLS estimator $\hat{\theta}^{(0)}$ and use $\hat{\mathbf{\Sigma}}_{t,n}(\hat{\theta}^{(0)})$ in equation (10) to obtain an AQL estimator $\hat{\theta}^{(1)}$. Then update (10) by employing $\hat{\mathbf{\Sigma}}_{t,n}(\hat{\theta}^{(1)})$ and solve for $\hat{\theta}^{(2)}$. Iterate this several time until it converges.

For more detail in AQL approach based on kernel smoothing for multivariate heteroskedastic models with correlation see Alzghool, et al. [3].

3 Parameter estimation

In this section we introduce how to apply the AQL approach to SSM. Consider the following state-space model

$$y_t = f_1(\alpha_t, \theta) + h_1(y_{t-1}, \theta)\epsilon_t, \quad t = 1, 2, \cdots, T \tag{12}$$

$$\alpha_t = f_2(\alpha_{t-1}, \theta) + h_2(\alpha_{t-1}, \theta)\eta_t, \quad t = 1, 2 \cdots, T, \tag{13}$$

where $\{y_t\}$ represents the time series of observations, $\{\alpha_t\}$ the state variables, θ unknown parameter taking value in an open subset Θ of d-dimensional Euclidean space, f_1 and f_2 are known functions of the past information, h_1 and h_2 are unknown functions. Denote $\delta_t = (h_1(y_{t-1}, \theta)\epsilon_t, h_2(\alpha_{t-1}, \theta)\eta_t)'$. Then δ_t is a martingale difference with

$$E_{t-1}(\delta_t) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

and

$$E_{t-1}(\delta_t\delta_t') = \mathbf{\Sigma}_t = \begin{bmatrix} \sigma(y_t;\theta) & \sigma((y_t,\alpha_t);\theta) \\ \sigma((y_t,\alpha_t);\theta) & \sigma(\alpha_t;\theta) \end{bmatrix}.$$

Traditionally, normality or conditional normality condition is assumed and the estimation of parameters are obtained by the ML approach. However, in many applications the normality assumption is not realistic. Further more, the probability structure of the model may not be known. Thus the maximum likelihood method is not applicable or it is too complex to estimate parameters through the ML method as the calculation involved is complex sometimes. In the following the AQL approach for estimating the parameters in SSM is introduced. This approach can be carried out without full knowledge of the system probability structure and Σ_t. It involves in making decision about the initial values of θ, Σ_t and iterative procedure. Each iterative procedure consists of three steps. The first step is to use the AQL method to obtain the optimal estimation for each α_t, say $\hat{\alpha}_t$. The second step is to estimate Σ_t by kernel estimator. The third step is to combine the information of $\{y_t\}$ and $\{\hat{\alpha}_t\}$ to adjust the estimate of θ through the AQL method.

In Step 1, assign an initial value to θ, Σ_t and consider the following martingale difference

$$\delta_t = \begin{bmatrix} h_1(y_{t-1},\theta)\epsilon_t \\ h_2(\alpha_{t-1},\theta)\eta_t \end{bmatrix} = \begin{bmatrix} y_t - E(y_t|\mathcal{F}_{t-1}) \\ \alpha_t - E(\alpha_t|\mathcal{F}_{t-1}) \end{bmatrix}$$

and estimating function space

$$\mathcal{G}^{(t)}_T = \{A_t\delta_t \ \mid \ A_t \text{ is } \mathcal{F}_{t-1} \text{ measurable}\},$$

where α_t is considered as an unknown parameter. A sequence of asymptotic quasi-score estimating functions in this estimating function space is

$$G^*_t(\alpha_t) = E_{t-1}(\frac{\partial\delta_t}{\partial\alpha_t})\hat{\Sigma}^{-1}_{t,n}\delta_t.$$

To obtain the AQL estimate $\hat{\alpha}_t$ of α_t, we let $G^*_t(\alpha_t) = 0$ and solve the equation for α_t. This estimation is as same as the estimation given by Kalman filter approach when the underlying system has a normal probability structure. (For detailed discussion see Lin, [20]).

In Step 2, using kernel estimator (6) and (7) to obtain $\hat{\mathbf{\Sigma}}_{t,n}(\theta^{(0)})$

In Step 3, θ is considered as an unknown parameter and the estimating function space

$$\mathcal{G}_T = \{\sum_{t=1}^{T} A_t\delta_t \ \mid \ A_t \text{ is } \mathcal{F}_{t-1} \text{ measurable}\}$$

is considered. Then a sequence of asymptotic quasi-score estimating functions in this estimating function space is

$$G_T^*(\theta) = \sum_{t=1}^{T} E_{t-1}(\frac{\partial \delta_t}{\partial \theta})\hat{\mathbf{\Sigma}}_{t,n}^{-1}(\theta^{(0)})\delta_t.$$

To obtain the AQL estimate $\hat{\theta}$ for θ we let $G_T^*(\theta) = 0$ and solve the equation while replacing α_t by $\hat{\alpha}_t$ obtained from Step 1. The $\hat{\mathbf{\Sigma}}_{t,n}(\theta^{(0)})$ and $\hat{\theta}$ obtained from Step 2 and 3 respectively will be used as a new initial value for the θ and Σ_t in Step 1 in the next iterative procedure. These three steps will be alternatively repeated until it converges.

In determining the NW type kernel estimate for $\hat{\mathbf{\Sigma}}_{t,n}$, the bandwidths are determined by quick and simple bandwidth selectors i.e. (oversmoothed bandwidth selection rules). The oversmoothed principle relies on the fact that there is a simple upper bound for the asymptotic mean integrated squared error (AMISE-optimal bandwidth). The oversmoothed bandwidth selector is

$$\hat{h}_{os} = (\frac{243R(K)}{35\mu_2(K)^2 n})^{1/5} s \tag{14}$$

where s is the sample standard deviation, $R(K) = \int_{-1}^{1} K(u)^2 du$, and $\mu_2(K) = \int_{-1}^{1} u^2 K(u) du$ (see Wand and Jones, [25]).

In the following we demonstrate the application of the AQL approach. Two simulation studies are presented below. One is based on Poisson Model (PM) and other is based on the basic Stochastic Volatility Model (SVM).

4 Simulations studies

4.1 Poisson model PM

Let $y_1, y_2, \cdots, y_T$ be observations and $\alpha_1, \alpha_2, \cdots, \alpha_T$ be states. The state-space model is given by

$$y_t \sim \text{Poisson distribution with parameter } e^{\beta+\alpha_t},$$
$$\alpha_t = \phi\alpha_{t-1} + h(\alpha_{t-1}, \theta)\eta_t, \tag{15}$$

where η_t are i.i.d with mean 0 and variance σ_η^2. The study on the generalized form of the above model can be found from Durbin and Koopman [9], Kuk [18], and Davis and Rodriguez-Yam [7]. Here the information on η_t is only given by the first two moments. β, ϕ and σ_η^2 are unknown. Based on this situation, we consider the following martingale difference

$$\delta_t = \begin{bmatrix} \epsilon_t \\ h(\alpha_{t-1}, \theta)\eta_t \end{bmatrix} = \begin{bmatrix} y_t - e^{\beta+\alpha_t} \\ \alpha_t - \phi\alpha_{t-1} \end{bmatrix}.$$

Our estimation consists of three steps. In Step 1, let α_t act as an unknown parameter. A sequence of asymptotic quasi-score estimating functions in the estimating function space determined by

$$\mathcal{G}_t = \{A_t\delta_t | \ A_t \text{ is } \mathcal{F}_{t-1} \text{ measurable } \}$$

is

$$G_t^*(\alpha_t) = (-e^{\beta+\alpha_t}, 1)\Sigma_{t,n}^{-1} \begin{bmatrix} y_t - e^{\beta+\alpha_t} \\ \alpha_t - \phi\alpha_{t-1} \end{bmatrix}$$

To carry out the three steps estimation procedure described in Section 3, the starting value $\theta_0 = (\beta_0, \phi_0)$, $\Sigma_t = I_2$ identity matrix, and the initial value for state process α_t are required. For detail dissection about the impact of the starting value of θ_0 and the issue of the initial value of α_t on parameter estimation see Alzghool and Lin [20]. Initially we assign $\alpha_0 = \hat{\alpha}_0 = 0$. Once the optimal estimate of α_{t-1} is obtained, say $\hat{\alpha}_{t-1}$, the AQL estimate of α_t, will be given by solving equation $G_t^*(\alpha_t) = 0$ through Newton-Raphson algorithm. It gives

$$\alpha_t^{(k+1)} = \alpha_t^{(k)} - \frac{-y_t e^{\beta+\alpha_t^{(k)}} + e^{2(\beta+\alpha_t^{(k)})} + (\alpha_t^{(k)} - \phi\hat{\alpha}_{t-1})}{-y_t e^{\beta+\alpha_t^{(k)}} + 2e^{2(\beta+\alpha_t^{(k)})} + 1}. \tag{16}$$

It starts with $\alpha_t^{(1)} = \hat{\alpha}_{t-1}$ and will be iterative till it is convergent. Then move to Step 2. In Step 2, using kernel estimator (6) and (7) to obtain

$$\hat{\mathbf{\Sigma}}_{t,n}(\theta^{(0)}) = \begin{bmatrix} \hat{\sigma}_n(y_t) & \hat{\sigma}_n(y_t, \alpha_t) \\ \hat{\sigma}_n(\alpha_t, y_t) & \hat{\sigma}_n(\alpha_t) \end{bmatrix}$$

In Step 3, let $\theta = (\beta, \phi)$ act as unknown parameters. We apply the AQL method to estimate θ. In this step, the estimating function space

$$\mathcal{G}_T = \{\sum_{t=1}^{T} A_t\delta_t | \ A_t \text{ is } \mathcal{F}_{t-1} \text{ measurable } \}$$

is considered. The asymptotic quasi-score estimating function related to $\mathcal{G}_T$ is

$$G_T^*(\beta, \phi) = \sum_{t=1}^{T} \begin{bmatrix} -e^{\beta+\alpha_t} & 0 \\ 0 & -\alpha_{t-1} \end{bmatrix} \hat{\mathbf{\Sigma}}_{t,n}^{-1} \begin{bmatrix} y_t - e^{\beta+\alpha_t} \\ \alpha_t - \phi\alpha_{t-1} \end{bmatrix}.$$

Replace α_t by $\hat{\alpha}_t$, $t = 1, 2, \cdots, T$, and the AQL estimate of $\theta = (\beta, \phi)$ will be given by solving $G_T^*(\beta, \phi) = 0$. The above three steps will be iteratively repeated until it converges. The $\hat{\mathbf{\Sigma}}_{t,n}(\theta^{(0)})$ and $\theta = (\beta, \phi)$ obtained from previous Step 2 and 3 will be used as an initial value for Step 1 in next iteration. Our experience showed that the algorithm converged after three iterations.

To demonstrate the above estimation procedures we carried out a simulation study on model (15). In our simulation study, $h(\alpha_{t-1}, \theta)$ is assigned as 1. The main reason for doing is that, given $h(\alpha_{t-1},\theta)=1$, Σ_t can be easily evaluated. Thus, the QL method can be applied to simulated data, and it is possible to compare the QL estimation with the estimations given by the AQL approach, in which Σ_t is pretended to be unknown. Our simulation was carried as follows: Firstly, independently simulate 1000 samples with size 500 from (15) based on a true parameter $\theta=(\beta,\phi)$. After series $\{y_t\},\{\alpha_t\}$ are generated, we pretend that α_t are unobserved and ϕ and β are unknown. Then apply the above estimation procedure to y_t only to obtain the estimation of α_t, ϕ and β. We consider different parameter settings for $\theta=(\phi,\beta)$ which are the same as the layout considered in Rodriguez-Yam [23]. For the simulation, we compute mean and root mean squared errors for $\hat{\beta}$ and $\hat{\phi}$ based on N=1000 independent samples. Result are shown in Table 1. In Table 1, AQL denotes the asymptotic quasi-likelihood estimate, QL denotes the quasi-likelihood estimate.

Table 1: Comarison of AQL and QL estimates for PM based on 1000 replication. Root mean square error of estimates are reported below each estimate.

	$\sigma_\eta=0.675$		$\sigma_\eta=0.484$		$\sigma_\eta=0.308$	
	γ	ϕ	γ	ϕ	γ	ϕ
true	-0.613	0.90	-0.613	0.95	-0.613	0.98
AQL	-0.620	0.990	-0.615	0.990	-0.616	0.990
	0.046	0.090	0.031	0.040	0.048	0.011
QL	-0.610	0.890	-0.611	0.939	-0.616	0.969
	0.004	0.025	0.007	0.021	0.023	0.017
	$\sigma_\eta=0.312$		$\sigma_\eta=0.223$		$\sigma_\eta=0.142$	
true	0.15	0.90	0.15	0.95	0.15	0.98
AQL	0.155	0.939	0.153	0.957	0.153	0.968
	0.008	0.057	0.007	0.037	0.009	0.035
QL	0.149	0.898	0.149	0.945	0.147	0.974
	0.005	0.021	0.009	0.017	0.021	0.012
	$\sigma_\eta=0.111$		$\sigma_\eta=0.079$		$\sigma_\eta=0.051$	
true	0.373	0.90	0.373	0.95	0.373	0.98
AQL	0.374	0.872	0.374	0.901	0.373	0.941
	0.002	0.067	0.004	0.079	0.002	0.061
QL	0.372	0.898	0.345	0.946	0.345	0.973
	0.011	0.019	0.030	0.015	0.033	0.013

The result in Table (1) show that AQL performed as well as QL in the state space model parameters estimation. In some cases the AQL more efficient than QL with smaller root mean square error, because true Σ_t is not a diagonal matrix. But, for simplicity purpose assumed to be a diagonal matrix when the QL method is applied.

4.2 Stochastic Volatility Models (SVM)

For the second simulation example, we consider the stochastic volatility process, which is often used for modelling log-returns of financial assets, defined by

$$y_t=\sigma_t\xi_t=e^{\alpha_t/2}\xi_t,\quad t=1,2,\cdots,T, \tag{17}$$

and

$$\alpha_t=\gamma+\phi\alpha_{t-1}+h(\alpha_{t-1},\theta)\eta_t,\quad t=1,2,\cdots,T, \tag{18}$$

where both ξ_t and η_t i.i.d respectively; η_t has mean 0 and variance σ_η^2. A key feature of the SVM in (17) is that it can be transformed into a linear model by taking the logarithm of the square of observations

$$\ln(y_t^2)=\alpha_t+\ln{\xi_t}^2,\quad t=1,2,\cdots,T. \tag{19}$$

If ξ_t were standard normal, then $E(\ln\xi_t^2)=-1.2704$ and $Var(\ln\xi_t^2)=\pi^2/2$ (see Abramowitz and Stegun [1], p943). Let $\varepsilon_t=\ln\xi^2+1.2704$. The disturbance ε_t is defined so as to have zero mean. Based on this situation, we consider the following martingale difference

$$\delta_t=\begin{bmatrix}\epsilon_t\\ h(\alpha_{t-1},\theta)\eta_t\end{bmatrix}=\begin{bmatrix}\ln(y_t^2)-\alpha_t+1.2704\\ \alpha_t-\gamma-\phi\alpha_{t-1}\end{bmatrix}.$$

In Step 1, let α_t act as an unknown parameter. A sequence of asymptotic quasi-score estimating function determined by the estimating function space

$$\mathcal{G}_t=\{A_t\delta_t|\ A_t \text{ is } \mathcal{F}_{t-1} \text{ measurable }\}$$

is

$$G^*_{(t)}(\alpha_t)=(-1,1)\mathbf{\Sigma}_{t,n}^{-1}\begin{bmatrix}\ln(y_t^2)-\alpha_t+1.2704\\ \alpha_t-\gamma-\phi\alpha_{t-1}\end{bmatrix}$$

Let $\hat{\alpha_0}=0$ and starting values $\theta_0=(\gamma_0,\phi_0)$, $\mathbf{\Sigma}_{t,n}^{(0)}=I_2$. Given $\hat{\alpha}_{t-1}$ the optimal estimation of α_{t-1}, the AQL estimate of α_t, i.e. the optimal estimation of α_t, will be given by solving $G^*_{(t)}(\alpha_t)=0$, i.e.

$$\hat{\alpha_t}=\frac{\ln(y_t^2)+1.2704+\phi\hat{\alpha}_{t-1}+\gamma}{2},\quad t=1,2,\cdots,T. \tag{20}$$

In Step 2, using kernel estimator (6) and (7) to obtain

$$\hat{\mathbf{\Sigma}}_{t,n}(\theta^{(0)})=\begin{bmatrix}\hat{\sigma}_n(y_t) & \hat{\sigma}_n(y_t,\alpha_t)\\ \hat{\sigma}_n(\alpha_t,y_t) & \hat{\sigma}_n(\alpha_t)\end{bmatrix}$$

In Step 3, based on $\{\hat{\alpha_t}\}$ and $\{y_t\}$, let $\theta=(\gamma,\phi)$ act as unknown parameters, and use the AQL approach to estimate them. A sequence of asymptotic quasi-score estimating function related to the estimating function space

$$\mathcal{G}=\{\sum_{t=1}^{T}A_t\begin{bmatrix}\epsilon_t\\ \eta_t\end{bmatrix}|\ A_t \text{ is } \mathcal{F}_{t-1} \text{ measurable }\}$$

is

$$G_T(\gamma,\phi)=\sum_{t=1}^{T}\begin{bmatrix}0 & -1\\ 0 & -\alpha_{t-1}\end{bmatrix}\hat{\mathbf{\Sigma}}_{t,n}^{-1}\begin{bmatrix}\ln(y_t^2)-\alpha_t+1.2704\\ \alpha_t-\gamma-\phi\alpha_{t-1}\end{bmatrix}.$$

Replace α_t by $\hat{\alpha}_t$,, $t=1,2,\cdots,T$, the AQL estimate of γ and ϕ will be given by solving $G_T(\gamma,\phi)=0$.

The above three steps will be iteratively repeated until it converges. The $\hat{\mathbf{\Sigma}}_{t,n}$ and $\theta = (\gamma, \phi)$ obtained from previous step will be used as an initial value for next iterative.

The format for this simulation study is the same as the layout considered by Rodriguez-Yam [23]. From empirical studies (e.g Harvey and Shepard [12]; Jacquier et, al. [17]) the values of ϕ between 0.9 and 0.98 are of primary interest. For this simulation study, 1000 independent samples with size 1000 simulated from (17) and (18) where $h(\alpha_{t-1}, \theta) = 1$, we compute mean and root mean squared errors for $\hat{\phi}$, $\hat{\gamma}$. The results are shown in Table (2). AQL denotes the asymptotic quasi-likelihood estimate, QL denotes the quasi-likelihood estimate. The

Table 2: Comarison of AQL and QL estimates for SVM based on 1000 replication. Root mean square error of estimates are reported below each estimate.

	$\sigma_\eta = 0.675$		$\sigma_\eta = 0.484$		$\sigma_\eta = 0.308$	
	γ	ϕ	γ	ϕ	γ	ϕ
true	-0.821	0.90	-0.411	0.95	-0.6134	0.98
AQL	-0.716	0.988	-0.369	0.978	-0.161	0.98
	0.155	0.091	0.047	0.028	1.356	0.173
QL	-0.989	0.867	-0.563	0.921	-0.213	0.95
	0.254	0.039	0.202	0.035	0.075	0.031
	$\sigma_\eta = 0.363$		$\sigma_\eta = 0.260$		$\sigma_\eta = 0.166$	
true	-0.736	0.90	-0.368	0.95	-0.147	0.98
AQL	-0.696	0.968	-0.318	0.950	-0.096	0.948
	0.047	0.068	0.052	0.010	0.086	0.221
QL	-0.835	0.898	-0.416	0.931	-0.155	0.970
	0.153	0.015	0.083	0.022	0.030	0.012
	$\sigma_\eta = 0.135$		$\sigma_\eta = 0.096$		$\sigma_\eta = 0.061$	
true	-0.706	0.90	-0.353	0.95	-0.141	0.98
AQL	-0.639	0.895	-0.386	0.988	-0.122	0.989
	0.405	0.548	0.034	0.038	0.020	0.010
QL	-0.721	0.891	-0.353	0.946	-0.143	0.979
	0.070	0.014	0.037	0.007	0.012	0.002

results in Table (2) farther confirm that AQL performed as well as QL in the state space model parameters estimation.

5 Application to real data

The data set consists of the observed time series $y_1, \ldots, y_{168}$ of monthly number of U.S. cases of poliomyelitis for 1970 to 1983 that was first considered by Zeger [26]. We adopt the same model used by Zeger in which the distribution of Y_t, given the state α_t, is Poisson with rate $\lambda = e^{\mathbf{x}_t'\beta + \alpha_t}$. Where $\beta = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6)$, and $\mathbf{x_t}$ is the vector of covariates given by

$\mathbf{x}_t' = (1, \frac{t}{1000}, \cos(\frac{2\pi t}{12}), \sin(\frac{2\pi t}{12}), \cos(\frac{2\pi t}{6}), \sin(\frac{2\pi t}{6}))$, and the state process is assumed to follow the AR(1) model given by $\alpha_t = \phi\alpha_{t-1} + \epsilon_t$, $t = 1, \ldots, n$.

Table (3) contains the AQL and QL estimates. The results in (3) are slightly different. In the AQL approach, we assume there is correlation between series, but in the QL approach, we do not assume that. The second and third columns in table (3) give the mean of residuals squares and the standard deviation of the residuals squares. Both values indicate that the AQL approach catches more information from data than the QL approach does.

Table 3: Parameter estimates for polio data by AQL (second row) and QL (third row) approaches

$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$	$\hat{\phi}$	mean	S.d
0.18	-4.38	-0.12	-0.42	0.20	-0.44	0.76	2.18	3.94
0.20	-3.22	0.08	-0.51	0.39	-0.11	0.75	2.46	4.86

6 Conclusion

In this paper an alternative approach, the AQL method, for estimating the parameters in nonlinear and non-Gaussian State-Space Models with unspecific correlation is given. Results from the simulation study indicates that the AQL method is an efficient estimation procedure. The study also shows that the AQL estimating procedure is easy to implement, especially when the system probability structure can not be fully specified. By utilising the nonparametric kernel estimator of conditional variance covariances matrix $\mathbf{\Sigma}_t$ to replace the true $\mathbf{\Sigma}_t$ in the standard quasi-likelihood, the AQL method avoids the risk of potential miss-specification of $\mathbf{\Sigma}_t$ and thus make the parameter estimator more efficient.

References

[1] M. Abramovitz, and N. Stegun, *Handbook of Mathematical Functions*, Dover Publication, New York, 1970.

[2] R. Alzghool and Y.-X. Lin, Estimation for State-Space Models: Quasi-likelihood, (submited to *J. of Applied statistics*) 2007.

[3] R. Alzghool and Y.-X. Lin, and S. X., Chen, Asymptotic Quasi-likelihood Based on Kernel Smoothing for Multivariate Heteroskedastic Models with Correlation, (submited to *The Australian and New Zealand Journal of Statistics*) 2007.

[4] F.J. Breidt, and A.L. Carriquiry, Improved quasi-maximum likelihood estimation for stochastic volatility models. In: Zellner, A. and Lee, J.S. (Eds.), *Modelling and Prediction: Honouring Seymour Geisser*, Springer, New York, 228-247, 1996.

[5] K. S. Chan and J. Ledolter, Monte Carlo EM estimation for time series models involving counts, *J. Amer. Statist. Assoc.*, **90**, 242-252, 1995.

[6] R. A. Davis, and W. T. M. Dunsmuir, and Y. Wang, *Modelling time series of count data.* In Asymptotics, Nonparametrics and Time Series (ed Subir Ghosh), Marcel Dekker, 1988.

[7] R. A. Davis, and G. Rodriguez-Yam, Estimation for State-Space Models: an approximate likelihood approach, *Statistica Sinica*, **15**, 381-406, 2005.

[8] J. Durbin, and S. J. Koopman, *Time Series Analysis by State Space Methods*, Oxford, New York, 2001.

[9] J. Durbin, and S. J. Koopman, Monte Carlo maximum likelihood estimation for non-gaussian state space models, *Biometrika*, **84**, 669-684, 1997.

[10] A. C. Harvey, *Forecasting, Structural Time Series Moodels and the Kalman Filter*, Cambridge University Press, 1989.

[11] A. C. Harvey, and M. Streible, Testing for a slowly changing level with special reference to stochastic volatility, *J. Econometrics*, **87**, 167-189, 1998.

[12] A. C. Harvey, and N. Shepard, *Estimation and testing of stochastic variance models*, Unpublished manuscript, The London School of Economics, 1993.

[13] W. Härdle, *Applied Nonparametric Regression*, Cambridge, University Press, 1990.

[14] W. Härdle, and A., Tsybakov, and L., Yang, Nonparametric vector autoregression, *J. Statist. Planning Inference*, **68**, 221-245, 1998.

[15] W. Härdle, and A., Tsybakov, Local polynomial estimators of the volatility function in nonparametric autoregression, *J. Econometrics*, **81**, 223-242, 1997.

[16] C. Heyde, *Quasi-likelihood and its Application: a general Approach to Optimal Parameter Estimation*, Springer, New York, 1997.

[17] E. Jacquire, and N. G. Polson, and P. E. Rossi, Bayesian analysis of stochastic volatility models (with discussion), *J. Bus. Econom. Statist*, **12**, 371-417, 1994.

[18] A. Y. Kuk, The use of approximating models in Monte Carlo maximum likelihood estimation, *Statist. Probab. Letter*, **45**, 325-333, 1999.

[19] Y.-X. Lin, and C.C. Heyde, On spaces of estimating functions, *Journal of Statistical Planning and Inference*, **63**, 255-264, 1997.

[20] Y.-X. Lin, An alternative derivation of the Kalman filter using the quasi-likelihood method, *J. of Statistical Planning and Inference*, **137**, 1627-1633, 2007.

[21] Y.-X. Lin, A new kind of asymptotic quasi-score estimating function, *Scandinavian Journal of Statistics*, **27**, 97-109, 2000.

[22] M. K. Pitt, and N. Shepard, Filtering via simulation: auxiliary particle filters, *J.Amer. Statist. Assoc.*, **94**, 590-599, 1999.

[23] G. Rodriguez-Yam, *Estimation for State-Space Models and Baysian regression analysis with parameter constraints*, Ph.D. Thesis, Colorado State University, 2003.

[24] G. Sandmann, and S. J. Koopman, Estimation of stochastic volatility models via Monte Carlo maximum likelihood, *J. Econometrics*, **87**, 271-301. 1998.

[25] M.P. Wand, and M.C. Jones, *Kernel Smoothing*, Chapman & Hall, New York, 1995.

[26] S. L. Zeger, A Regression Model for Time Series of Counts, *Biometrika*, **75**, 621-629, 1988.