2009

Asymptotic Quasi-likelihood Based on Kernel Smoothing for Multivariate Heteroskedastic Models with Correlation

R. Alzghool
Al-Balqa Applied University

Y. X. Lin
University of Wollongong, yanxia@uow.edu.au

S. X. Chen
Peking University

Centre for Statistical and Survey Methodology

The University of Wollongong

Working Paper

22-09

Asymptotic Quasi-Likelihood Based on Kernel Smoothing for Multivariate Heteroskedastic Models with Correlation

Raed Alzghool, Yan-Xia Lin and Song Xi Chen

Centre for Statistical & Survey Methodology, University of Wollongong, Wollongong NSW 2522. Phone +61 2 4221 5435, Fax +61 2 4221 4845. Email: anica@uow.edu.au

Asymptotic Quasi-likelihood Based on Kernel Smoothing for Multivariate Heteroskedastic Models with Correlation

Raed Alzghool[1] , Yan-Xia Lin[2] and Song Xi Chen[3]

Department of Applied Sciences[1]
Faculty of Prince Abdullah Bin Ghazi for Science & Information Technology
AL-Balqa' Applied University, Al Salt, 19117, Jordan

School of Mathematics and Applied Statistics[2]
University of Wollongong, NSW 2500, Australia

Guanghua School of Management[3]
Peking University, Beijing, 100871, Beijing, China

SYNOPTIC ABSTRACT

This paper considers parameter estimation in multivariate heteroscedastic models with unspecific correlation. In this paper, we propose an asymptotic quasi-likelihood (AQL) approach which utilises a nonparametric kernel estimator of variance covariances matrix Σ to replace the true Σ in the standard quasi-likelihood. The kernel estimation avoids the risk of potential miss-specification of Σ and thus make the parameter estimator more robust. The well developed theory framework for AQL (See Lin, 2000) provides a solid base for ensuring the efficiency of the approach developed in this paper. This has been further verified by empirical studies carried out in this paper.

Key Words and Phrases: multivariate heteroskedastic; asymptotic quasi-likelihood; kernel smoothing; martingale; Quasi-likelihood.

1 INTRODUCTION

Consider the following multivariate heteroskedastic model

$$\mathbf{y}_t = \mathbf{m}(\mathbf{y}_{t-1}, \ldots, \mathbf{y}_{t-q}; \theta) + \mathbf{\Sigma}^{1/2}(\mathbf{y}_{t-1}, \ldots, \mathbf{y}_{t-q}; \theta)\varepsilon_t, \quad t = 1, 2, \cdots, \tag{1}$$

where ε_t are independent and identically distributed random variables with zero mean and variance-covariance matrix $\mathbf{I}_p$; $\mathbf{m} : R^{qp} \to R^p$ is a known autoregressive function with unknown parameter $\theta \in \Theta$ where Θ is an open parameter space in R^d for a positive integer d, and $\mathbf{\Sigma} : R^{qp} \to R^p \times R^p$ is an unknown positive definite matrix. Without further explanation, sometimes we denote $\mathbf{m}(\mathbf{y}_{t-1}, \ldots, \mathbf{y}_{t-q}; \theta)$ and $\mathbf{\Sigma}^{1/2}(\mathbf{y}_{t-1}, \ldots, \mathbf{y}_{t-q}; \theta)$ by $\mathbf{m}_t(\theta)$ and $\mathbf{\Sigma}_t(\theta)$ respectively, or by $\mathbf{m}_t$ and $\mathbf{\Sigma}_t$ respectively. This model is quite general and include popular time series models such as autoregressive, quasilinear ARMA, ARCH and GARCH models. For extensive review on multivariate heteroscedastic models see Bauwens, et al. (2006). Existing methods of estimating θ in model (1) include (i) the maximum likelihood (ML) method by assuming the conditional distribution of $\mathbf{y}_t$ given the past data information up to time $t-1$ is known; (ii) Pseudo maximum likelihood (PML) method by pretending the underlying conditional distribution is Gaussian with mean $\mathbf{m}(\mathbf{y}_{t-1}, \ldots, \mathbf{y}_{t-q}; \theta)$ and variance-covariance matrix $\mathbf{\Sigma}_t$; and (iii) the quasi-likelihood (QL) method by assuming knowledge on $\mathbf{m}(\mathbf{y}_{t-1}, \ldots, \mathbf{y}_{t-q}; \theta)$ and $\mathbf{\Sigma}_t$ given $\mathcal{F}_{t-1}$, where $\mathcal{F}_{t-1}$ is the σ-field generated by $\{\mathbf{y}_s\}_{s \le t-1}$. The ML method is the most efficient method of estimation if the distribution is correctly specified. However, the inference can be misleading if the distribution is wrongly specified. The PML method assumes the conditional distribution is Gaussian with mean $\mathbf{m}(\mathbf{y}_{t-1}, \ldots, \mathbf{y}_{t-q}; \theta)$ and variance-covariance matrix $\mathbf{\Sigma}_t$. As it correctly specifies the first two conditional moments, the estimator is consistent whereas the efficiency of the estimation depends on how far the real underlying conditional distribution is from the Gaussian distribution. The QL method relaxes the distributional assumptions by PML and only assumes the knowledge on $\mathbf{m}(\mathbf{y}_{t-1}, \ldots, \mathbf{y}_{t-q}; \theta)$ and $\boldsymbol{\Sigma}_t$ given $\mathcal{F}_{t-1}$, where $\mathcal{F}_{t-1}$ is the σ-field generated by $\{y_s\}_{s \leq t-1}$. 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 (1997). A limitation of the QL is that the nature of $\boldsymbol{\Sigma}_t$ is not always known. A misspecified $\boldsymbol{\Sigma}_t$ can lead to misleading inference on the unknown parameter. In this paper we propose an asymptotic quasi-likelihood (AQL) approach which essentially carries out the QL method when the form of variance-covariance matrix $\boldsymbol{\Sigma}_t$ is completely unspecified. The idea is to use the nonparametric kernel method to estimate $\boldsymbol{\Sigma}_t$ and then use it to replace $\boldsymbol{\Sigma}_t$ in the standard QL. This AQL approach provides an alternative method of parameter estimation to the ordinary least square (OLS) which simply assumes $\boldsymbol{\Sigma}_t$ is a constant function by ignoring the heteroscedasticity. Although the OLS gives consistent estimation, based on the optimal criterion (see Heyde, 1997), it is not efficient when heteroscedasticity is present. Since the AQL method is developed based on the optimal criterion, it ensures the proposed method is asymptotically efficient in theory. The properties of this asymptotic quasi-likelihood estimator based on kernel smoothing are investigated theoretically and numerically by simulations, which show that without knowing the variance-covariance matrix $\boldsymbol{\Sigma}_t$ and the probability structure of underlying modelling we can still obtain efficient parameter estimation.

This paper is structured as follows. In Section 2, the asymptotic quasi-likelihood based on kernel smoothing is introduced. A uniform consistent result on kernel estimation of a predictable process is presented in Section 3. Section 4 reports simulation results and covers numerical implementation. An analysis on a real data set on foreign exchange markets by the AQL method is given in Section 5. The summary is given in Section 6. All theoretical proofs are detailed in Appendix.

2 ASYMPTOTIC QUASI-LIKELIHOOD APPROACH

For reading convenience, we first briefly introduce the necessary background on the asymptotic quasi-likelihood method. Let us consider the following model which will cover model (1),

$$\mathbf{y}_t = \mathbf{m}_t(\theta) + \varepsilon_t^*, \qquad t = 1, 2, \cdots, \tag{2}$$

where $\mathbf{m}_t$ is $\mathcal{F}_{t-1}$ measurable; ε_t^* is a martingale difference associated with $\mathcal{F}_t$, i.e. $E(\varepsilon_t^*|\mathcal{F}_{t-1})= E_{t-1}(\varepsilon_t^*) = 0$; $\mathcal{F}_t$ 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 (2), if the expression of $E(\varepsilon_t^*\varepsilon_t^{*\prime}|\mathcal{F}_{t-1})=E_{t-1}(\varepsilon_t^*\varepsilon_t^{*\prime}) = \mathbf{\Sigma}_t$ is known, the standard quasi-score estimating function in estimating function space

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

is

$$\mathbf{G}_T^*(\theta) = \sum_{t=1}^{T} \dot{\mathbf{m}}_t(\theta)\mathbf{\Sigma}_t^{-1}(\mathbf{y}_t - \mathbf{m}_t(\theta)) \tag{3}$$

where $\dot{\mathbf{m}}_t(\theta) = \partial\mathbf{m}_t(\theta)/\partial\theta$ (see Heyde, 1997). Then the quasi-score normal equation is $\mathbf{G}_T^*(\theta) = 0$, whose root is the quasi-likelihood estimate of θ. Under certain regularity conditions, the quasi-likelihood estimate is consistency and achieves optimal efficiency within space $\mathcal{G}$ (Heyde, 1997). 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, 1997).

The quasi-score estimating function (3) relies on the knowledge of Σ_t. Such knowledge is not always available in practice considering there is only one sample path of the process $\{\mathbf{y}_t\}$ being observed. To facilitate QL in a situation where $E_{t-1}(\varepsilon_t^*\varepsilon_t^{*\prime})$ is unknown, Lin (2000) 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. The definition of asymptotic quasi-score estimating function is introduced below.

Definition 2.1 *Suppose $\mathbf{G}^*_{T,n}$ are a sequences of estimating functions in $\mathcal{G}$. If for all $\mathbf{G}_T \in \mathcal{G}$,*

$$(E\dot{\boldsymbol{G}}_T)^{-1}(E\boldsymbol{G}_T\boldsymbol{G}_T)'(E\dot{\boldsymbol{G}}'_T)^{-1} - (E\dot{\boldsymbol{G}}^*_{T,n})^{-1}(E\boldsymbol{G}^*_{T,n}\boldsymbol{G}^{*\prime}_T)(E\dot{\boldsymbol{G}}^{*\prime}_{T,n})^{-1}$$

is asymptotically non-negative definite, $\mathbf{G}^*_{T,n}$ is called an asymptotic quasi-score sequence of estimating functions in $\mathcal{G}$, and $\theta_{T,n}$ the solution of the asymptotic quasi-score normal equation $\mathbf{G}^*_{T,n} = 0$, is called the sequence of asymptotic quasi-likelihood estimates.

Let $\boldsymbol{\Sigma}_{t,n}$ be a sequence of $\mathcal{F}_{t-1}$-measurable random matrices converging to $E_{t-1}(\varepsilon^*_t\varepsilon^{*\prime}_t)$ in probability. Then,

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

forms a sequence of asymptotic quasi-score estimating functions. The corresponding roots of $\mathbf{G}^*_{T,n}(\theta) = 0$ form 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, 2000)

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

as $n \to \infty$, this means that the amount of Fisher Information provided by $\mathbf{G}^*_{T,n}$ 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 becomes 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 $\mathbf{\Sigma}_{t,n}$ in the AQL formulation (4).

In this paper, we will apply the asymptotic quasi-likelihood approach to Model (1). The detail discussion appear in next section. The following notation will be used in this paper. Let $\mathbf{x}_t = (\mathbf{y}_{t-1}, \ldots, \mathbf{y}_{t-q})$ be the lagged value of $\mathbf{y}_t = (y_{1t}, y_{2t}, \ldots, y_{pt})'$. Given an initial estimator of θ, say $\hat{\theta}^{(0)}$, which may be the OLS estimator, the Nadaraya-Watson (NW) estimator of $\mathbf{\Sigma}_t$ is $\hat{\mathbf{\Sigma}}_{t,n}$ with elements

$$\hat{\sigma}_n(y_{it}; \hat{\theta}^{(0)}) = \frac{\sum_{s=q+1}^{n} D_{its}(y_{is} - m_{is}(\mathbf{x}_{is}, \hat{\theta}^{(0)}))^2}{\sum_{s=q+1}^{n} D_{its}} \tag{5}$$

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

$i \neq j$, 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 r and h is a smoothing bandwidth such that $h \to 0$ and $nh^q \to \infty$ as $n \to \infty$.

A comprehensive review of the above NW type kernel estimator including the construction of K and the choice of h is available in Härdle (1990) and Wand and Jones (1995).

The estimating function for the model (1) based on the kernel estimators (5) and (6) is

$$\mathbf{G}^*_{T,n}(\theta) = \sum_{t=1}^{T} \dot{\mathbf{m}}_t(\theta) \hat{\mathbf{\Sigma}}_{t,n}^{-1} (\mathbf{y}_t - \mathbf{m}_t(\theta)) \tag{7}$$

and the asymptotic quasi-score normal equation is

$$\mathbf{G}^*_{T,n}(\theta) = \sum_{t=1}^{T} \dot{\mathbf{m}}_t(\theta) \hat{\mathbf{\Sigma}}_{t,n}^{-1}(\hat{\theta}^{(0)}) (\mathbf{y}_t - \mathbf{m}_t(\theta)) = 0. \tag{8}$$

where

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

The solution of (8) is called the AQL estimator of θ. In following sections, we will demonstrate why the solution of (8) can be considered as the AQL estimator of θ and how good it performance is in practice.

In practice to solve (8) iterative procedures are required. It suggests that the iterative procedure starts from the OLS estimator $\hat{\theta}^{(0)}$ and use $\hat{\mathbf{\Sigma}}_{t,n}(\hat{\theta}^{(0)})$ in the above estimating equation (8) to obtain an AQL estimator $\hat{\theta}^{(1)}$. Then update (8) by employing $\hat{\mathbf{\Sigma}}_{t,n}(\hat{\theta}^{(1)})$ and solve for $\hat{\theta}^{(2)}$. Iterate this several times until it converges.

Throughout this paper, $|\cdot|$ denotes the absolute value for number or norms for vector and matrix, respectively. Defined by $|U| = \sum_{i=1}^{q} |u_i|$ for $U = (u_1, u_2, \dots, u_q)' \in R^q$ and $|V| = \sum_{i=1}^{p}\sum_{j=1}^{q} |v_{ij}|$ for $V = (v_{ij})_{p\times q} \in R^p \times R^q$.

3 MAIN RESULTS

The key in the proposed AQL formulation (8) is the kernel estimation of $\mathbf{\Sigma}_t$. In this section we establish results on the consistency of the kernel estimator as it is needed by the AQL. Let the process $\{\mathbf{y}_t\}$ satisfy model (1) and for a given t

$$g(\mathbf{x}_{it}) = E(v(y_{it},\mathbf{x}_{it})|\mathbf{x}_{it} = \mathbf{x}_i),$$

and

$$g(\mathbf{x}_{it},\mathbf{x}_{jt}) = E(v((y_{it},\mathbf{x}_{it}),(y_{jt},\mathbf{x}_{jt}))|\mathbf{x}_{it} = \mathbf{x}_i, \mathbf{x}_{jt} = \mathbf{x}_j)$$

where v is a general continuous function of y_{it} and $\mathbf{x}_{it}$. The functions $g(\mathbf{x}_{it})$ and $g(\mathbf{x}_{it},\mathbf{x}_{jt})$ are $\mathcal{F}_{t-1}$-measurable and are predictable functions associated with the stochastic processes $\{y_{it}\}$ and $\{y_{jt}\}$. To simplify expression, we denote $v(y_{it},\mathbf{x}_{it})$ as $v_{i\mathbf{x}_t}$, and $v((y_{it},\mathbf{x}_{it}),(y_{jt},\mathbf{x}_{jt}))$ as $v_{ij\mathbf{x}_t}$. Clearly, for model (1), the case of $v_{i\mathbf{x}_t} = (y_{it} - m_{it}(\theta, \mathbf{x}_{it}))^2$ gives $g_t(\mathbf{x}_{it}) = \sigma_{iit}$ and $v_{ij\mathbf{x}_t} = (y_{i,t} - m_{it}(\theta, \mathbf{x}_{it}))(y_{jt} - m_{jt}(\theta, \mathbf{x}_{jt}))$ gives $g_t(\mathbf{x}_{it}, \mathbf{x}_{jt}) = \sigma_{ijt}$. Sometimes, for simplicity reasons, we might use $g(\mathbf{x}_{it})$ and $g(\mathbf{x}_{it}, \mathbf{x}_{jt})$ instead of $g_t(\mathbf{x}_{it})$ and $g_t(\mathbf{x}_{it}, \mathbf{x}_{jt})$ respectively.

Following seven key conditions are needed to assure the process being geometrically ergodic:

C1: Distribution of i.i.d. random vectors ε_t's is absolutely continuous (with respect to μ_p) and has a density $\psi(\cdot)$ which is positive a.e.(μ_p) with $E|\varepsilon_t| < \infty$ where μ_p denotes the Lebesgue measure in R^p;

C2: There exists constants $a_{ij} \geq 0$, $b_{ij} \geq 0$, $1 \leq i \leq p$, $1 \leq j \leq q$ such that for $|\mathbf{x}| \to \infty$,

$$|m(\mathbf{x})| \leq \sum_{i=1}^{p} \sum_{j=1}^{q} a_{ij}|x_{ij}| + o(|\mathbf{x}|),$$

$$|\Sigma^{1/2}(\mathbf{x})| \leq \sum_{i=1}^{p} \sum_{l=1}^{q} b_{ij}|x_{ij}| + o(|\mathbf{x}|);$$

C3: Functions $m(\mathbf{x})$ and $\Sigma(\mathbf{x})$ are bounded on any bounded Borel measurable set of R^{pq}. In addition, the matrix function $\Sigma(\mathbf{x})$ is symmetric for any $\mathbf{x} \in R^{pq}$, and satisfies either (1) $\inf_{\mathbf{x}\in S} \lambda_{min}\{\Sigma(\mathbf{x})\} > \lambda_S > 0$, for any compact set $S \subset R^{pq}$, or (2) $\lambda_{min}\{\Sigma(\mathbf{x})\} > 0$ for any $\mathbf{x} \in R^{pq}$ and $\Sigma(\mathbf{x})$ is continuous on R^{pq}, where $\lambda_{min}\{\Sigma(\mathbf{x})\}$ denotes the minimal eigenvalue of $\Sigma(\mathbf{x})$;

C4:

$$\max_{1\leq i\leq p}\{\sum_{j=1}^{q} a_{ij} + E|\varepsilon_1| \sum_{l=j}^{q} b_{ij}\} < 1$$

.

C5: K is a bounded Lipschitz continuous symmetric r-th order kernel and $\int |u|K(u)du < \infty$, and $h = cn^{-1/(2r+q)}$ for a positive constant c and $2r > q$;

C6: Let f_{ij} be the density of the ergodic measure $\pi_{ij}(\cdot)$ and $\phi_{ij}(\cdot) = g_{ij}(\cdot)f_{ij}(\cdot)$. We assume that both $f_{ij}(\cdot)$ and $\phi_{ij}(\cdot)$ are Lipschitz continuous, and S be a compact set in R^{pq} such that $c_0 \leq \inf_{\mathbf{x}\in S} f_{ij}(\mathbf{x}) \leq \sup_{\mathbf{x}\in S} f_{ij}(\mathbf{x}) \leq c_1$ for some positive c_0 and c_1, and for $i, j = 1, 2, 3, \cdots, p$;

C7: There exists a constant $a_0 > 0$ such that $\int \exp(a_0 u) f_{ij}(u) du < \infty$ and for all t, $E\{\exp(a_0 v_{ij\mathbf{x}})\} \le C_3 < \infty$.

Lu and Jiang (2001) proved under C1-C4 that the Markov chain $\{\mathbf{y}_t\}$ is geometrically ergodic, that is, it is ergodic with a stationary measure $\pi(\cdot)$ such that for almost every $\mathbf{y}$

$$\sum_{n=1}^{\infty} \rho^{-n} ||P^n(\cdot|\mathbf{y}) - \pi(\cdot)||_{\mathbf{tv}} < \infty \tag{9}$$

where $\rho \in (0,1)$, $P^n(B|\mathbf{y}) = P(\mathbf{y}_n \in B|\mathbf{y}_0 = \mathbf{y})$ for any Borel set B in R^p and $||\cdot||_{tv}$ is the total variation norm which is defined as

$$||\mu||_{tv} = \sup_{A \in \mathcal{B}(\mathcal{X})} \mu(A) - \inf_{A \in \mathcal{B}(\mathcal{X})} \mu(A)$$

for a measure μ defined on a σ-filed $\mathcal{B}(\mathcal{X})$ on a space $\mathcal{X}$.

C5-C7 are standard conditions in nonparametric curve estimation. In particular, the choice of $h = cn^{-1/(2r+q)}$ is fully compatible with the optimal bandwidth that minimises the mean integrated square error of the kernel estimator based on a r-th order kernel; see Härdle (1990) for details.

The r-th order kernel K can be constructed as follows. Let $k(\cdot)$ be a univariate kernel of r-th order such that

$$\int k(u)du = 1, \quad \int u^{\gamma} k(u) du = 0 \quad \text{for} \quad \gamma = 1, \ldots, r-1, \quad \int u^r k(u) du = k_r \ne 0.$$

Then $K(x_1, \cdots, x_q) = \prod_{l=1}^{q} k(x_i)$ is a q-dimensional r-th order kernel.

A NW kernel estimator for $g(\mathbf{x}_{it})$ is

$$\hat{g}_{n,h}(\mathbf{x}_{it}) = \frac{\sum_{s=q+1}^{n} K\left(\frac{\mathbf{x}_{it} - \mathbf{x}_{is}}{h}\right) v_{i\mathbf{x}_s}}{\sum_{s=q+1}^{n} K\left(\frac{\mathbf{x}_{it} - \mathbf{x}_{is}}{h}\right)} = \frac{\hat{\phi}_{n,h}(\mathbf{x}_{it})}{\hat{f}_{n,h}(\mathbf{x}_{it})}, \tag{10}$$

and for $g(\mathbf{x}_{it}, \mathbf{x}_{jt})$ is

$$\hat{g}_{n,h}(\mathbf{x}_{it}, \mathbf{x}_{jt}) = \frac{\sum_{s=q+1}^{n} K\left(\frac{\mathbf{x}_{it} - \mathbf{x}_{is}}{h}\right) K\left(\frac{\mathbf{x}_{jt} - \mathbf{x}_{js}}{h}\right) v_{ij\mathbf{x}_s}}{\sum_{s=q+1}^{n} K\left(\frac{\mathbf{x}_{it} - \mathbf{x}_{is}}{h}\right) K\left(\frac{\mathbf{x}_{jt} - \mathbf{x}_{js}}{h}\right)} = \frac{\hat{\phi}_{n,h}(\mathbf{x}_{it}, \mathbf{x}_{jt})}{\hat{f}_{n,h}(\mathbf{x}_{it}, \mathbf{x}_{jt})}, \tag{11}$$

where

$$\hat{\phi}_{n,h}(\mathbf{x}_{it}) = (nh^q)^{-1} \sum_{s=q+1}^{n} K\left(\frac{\mathbf{x}_{it} - \mathbf{x}_{is}}{h}\right) v_{i\mathbf{x}_s},$$

$$\hat{f}_{n,h}(\mathbf{x}_{it}) = (nh^q)^{-1} \sum_{s=q+1}^{n} K\left(\frac{\mathbf{x}_{it} - \mathbf{x}_{is}}{h}\right),$$

$$\hat{\phi}_{n,h}(\mathbf{x}_{it}, \mathbf{x}_{jt}) = (nh^{2q})^{-1} \sum_{s=q+1}^{n} K\left(\frac{\mathbf{x}_{it} - \mathbf{x}_{is}}{h}\right) K\left(\frac{\mathbf{x}_{jt} - \mathbf{x}_{js}}{h}\right) v_{ij\mathbf{x}_s},$$

and

$$\hat{f}_{n,h}(\mathbf{x}_{it}, \mathbf{x}_{jt}) = (nh^{2q})^{-1} \sum_{s=q+1}^{n} K\left(\frac{\mathbf{x}_{it} - \mathbf{x}_{is}}{h}\right) K\left(\frac{\mathbf{x}_{jt} - \mathbf{x}_{js}}{h}\right).$$

Clearly, $\hat{\phi}_{n,h}(\mathbf{x}_{it})$, $\hat{f}_{n,h}(\mathbf{x}_{it})$, $\hat{\phi}_{n,h}(\mathbf{x}_{it}, \mathbf{x}_{jt})$, and $\hat{f}_{n,h}(\mathbf{x}_{it}, \mathbf{x}_{jt})$ are respectively kernel estimators of $\phi(\mathbf{x}_{it})$, $f(\mathbf{x}_{it})$, $\phi(\mathbf{x}_{it}, \mathbf{x}_{jt})$, and $f(\mathbf{x}_{it}, \mathbf{x}_{jt})$.

To facilitate the application of the AQL based on the kernel method, we establish the uniformly convergence in probability of nonparametric kernel estimator of $\hat{g}(\mathbf{x}_{it})$, and $\hat{g}(\mathbf{x}_{it}, \mathbf{x}_{jt})$ in the following theorem.

Theorem 3.1 *: Under conditions C1-C7, as $n \to \infty$,*

$$\sup_{\mathbf{x}_t \in S} |\hat{\mathbf{\Sigma}}_n^h(\mathbf{x}_t) - \mathbf{\Sigma}(\mathbf{x}_t)| \xrightarrow{p} 0 \tag{12}$$

for a compact set S in R^{pq} defined in C3.

where $$\hat{\mathbf{\Sigma}}_n^h(\mathbf{x}_t) = \begin{pmatrix} \hat{g}_{n,h}(\mathbf{x}_{1t}) & \hat{g}_{n,h}(\mathbf{x}_{1t}, \mathbf{x}_{2t}) & \dots & \hat{g}_{n,h}(\mathbf{x}_{1t}, \mathbf{x}_{mt}) \\ \hat{g}_{n,h}(\mathbf{x}_{2t}, \mathbf{x}_{1t}) & \hat{g}_{n,h}(\mathbf{x}_{2t}) & \dots & \hat{g}_{n,h}(\mathbf{x}_{2t}, \mathbf{x}_{mt}) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{g}_{n,h}(\mathbf{x}_{mt}, \mathbf{x}_{1t}) & \hat{g}_{n,h}(\mathbf{x}_{mt}, \mathbf{x}_{2t}) & \dots & \hat{g}_{n,h}(\mathbf{x}_{mt}) \end{pmatrix}.$$

The proof of theorem 3.1 is given in Appendix.

Theorem 3.1 will play an important role in the approached studied in this paper. $\hat{\mathbf{\Sigma}}_n^h(\mathbf{x}_t)$ will act as the $\mathbf{\Sigma}_{t,n}$ in (8) to obtain the AQL estimate of θ, and

$$\mathbf{G}^*_{T,h,n}(\theta) = \sum_{t=1}^{T} \dot{\mathbf{m}}_t(\theta) \hat{\mathbf{\Sigma}}_n^h(\mathbf{x}_t)^{-1} (\mathbf{y}_t - \mathbf{m}_t(\theta)) \tag{13}$$

forms a sequence of estimating functions. In the following, we prove that $\{\mathbf{G}^*_{T,h,n}\}$ is a sequence of asymptotic quasi-score estimating functions. Before the proof, we need Theorem 3.2 below by Lin (2000).

Theorem 3.2 *Assume that there exists a quasi-score estimating function G_T^* in $\mathcal{G}$. If there is a sequence of estimating functions $\mathbf{G}_{T,n}^* \in \mathcal{G}$, $n = 1, 2, \cdots$, which satisfies the condition that*

$$\underline{\lim}_{n\to\infty} |E(\mathbf{G}_{T,n}^* \mathbf{G}_{T,n}^*{}')| \geq \alpha > 0.$$

and, for all $G_T \in \mathcal{G}$,

$$\lim_{n\to\infty} (E\dot{\mathbf{G}}_T)^{-1} E(\mathbf{G}_T \mathbf{G}_{T,n}^*{}') = W = \lim_{n\to\infty} (E\dot{\mathbf{G}}_{T,n}^*)^{-1} E(\mathbf{G}_{T,n}^* \mathbf{G}_{T,n}^*{}'),$$

where W is nonsingular, then $\{\mathbf{G}_{T,n}^*\}$ is an asymptotic quasi-score sequence of estimating functions in $\mathcal{G}$.

Theorem 3.2 gives a sufficient condition for asymptotic quasi-score estimating functions. Now, we are able to show that, under fairly weaker conditions, (13) is a sequence of asymptotic quasi-score estimating functions.

Consider Model (1),

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

is a standard quasi-score estimating function in space

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

The following theorem shows that, under conditions C1-C7 and some other weak conditions on the quasi-score estimating function $\mathbf{G}_T^*$, $\{\mathbf{G}_{T,h,n}^*\}$ is a sequence of asymptotic quasi-score estimating functions.

Theorem 3.3 *Under conditions C1-C7, if*

(i) there are real numbers c_{ij} and d_{ij} such that $0 < c_{ii} \leq |\sigma_{t,(ii)}/\hat{g}_n(\mathbf{x}_{it})| < d_{ii} < \infty$, and $0 < c_{ij} \leq |\sigma_{t,(ij)}/\hat{g}_n(\mathbf{x}_{it}, \mathbf{x}_{jt})| < d_{ij} < \infty$, for all t,

(ii) there is a positive real number $d > 0$ such that $|E[\mathbf{G}_T^*(\theta)\mathbf{G}_T^{*\prime}(\theta)]| > d > 0$,

then,

$$\lim_{n\to\infty} |E[\mathbf{G}^*_{T,h,n}(\theta)\mathbf{G}^{*\prime}_{T,h,n}(\theta)]| \geq d > 0$$

and, for any $G_T \in \mathcal{G} = \{\sum_{t=1}^T \alpha_t(\mathbf{y}_t - \mathbf{m}_t(\theta); \alpha_t$ *is* $\mathcal{F}_{t-1}$*-measureable*$\}$,

$$\lim_{n\to\infty} (E\dot{\mathbf{G}}_T)^{-1} E(\mathbf{G}_T \mathbf{G}^{*\prime}_{T,h,n}) = \mathbf{I},$$

where $\mathbf{I}$ *is an identity matrix.*

Furthermore, $\mathbf{G}^*_{T,h,n}(\theta)$ *forms a sequence of asymptotic quasi-score estimating functions.*

The proof of Theorem 3.3 is straightforward and briefly described below.

Without loss the generality, we may assume that process $\mathbf{y}_t$ has a compact domain. Under condition C1-C7, Theorem 3.1 shows that $\hat{\mathbf{\Sigma}}^h_n(\mathbf{x}_t)$ uniformly p-converges to $\mathbf{\Sigma}_t$ as $n \to \infty$. Therefore, we have

$$\mathbf{G}^*_{T,h,n}(\theta) \xrightarrow{p} \mathbf{G}^*_T(\theta), \qquad \text{as } n \to \infty.$$

Under conditions (i) and (ii), by applying dominated convergence Theorem (Loève, 1963, p125), we obtain $lim_{n\to\infty}|E[\mathbf{G}^*_{T,h,n}(\theta)\mathbf{G}^{*\prime}_{T,h,n}(\theta)]| > d > 0$,

$$\lim_{n\to\infty} (E\dot{\mathbf{G}}_T)^{-1} E(\mathbf{G}_T \mathbf{G}^{*\prime}_{T,h,n}) = (E\dot{\mathbf{G}}_T)^{-1} E(\mathbf{G}_T \mathbf{G}^{*\prime}_T)$$

and

$$\lim_{n\to\infty} (E\dot{\mathbf{G}}^*_{T,h,n})^{-1} E(\mathbf{G}^*_{T,h,n} \mathbf{G}^{*\prime}_{T,h,n}) = (E\dot{\mathbf{G}}^*_T)^{-1} E(\mathbf{G}^*_T \mathbf{G}^{*\prime}_T),$$

for any $\mathbf{G}_T \in G = \{\sum_{t=1}^T \alpha_t(\mathbf{y_t} - m_t(\theta); \alpha_t$ is $\mathcal{F}_{t-1}$-measureable$\}$. Since

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

is a standard quasi-score estimating function in $\mathcal{G}$, we have

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

for any estimating function $\mathbf{G}_t \in \mathcal{G}$ (see Heyde, 1997). From Theorem 3.2, $\mathbf{G}^*_{T,h,n}(\theta)$ forms a sequence of asymptotic quasi-score estimating functions.

4 SIMULATIONS STUDY

In this section we report results from simulation studies which design to evaluate the empirical performance of the proposed kernel based AQL approach for parameter estimation. Consider model

$$y_{1t} \sim \text{Poisson distribution with parameter } e^{\beta+\alpha_t} \tag{14}$$

$$y_{2t} = \alpha_{t+1} = \phi\alpha_t + c\alpha_t\eta_t = \phi\alpha_t + \epsilon_{2t}, \qquad t \le T, \tag{15}$$

where $E(\eta_t) = 0$, $Var(\eta_t) = \sigma_\eta^2$. Thus,

$$\begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix} = \begin{pmatrix} y_{1t} - e^{\beta+\alpha_t} \\ y_{2t} - \phi\alpha_t \end{pmatrix} \qquad t \le T,$$

forms a martingale difference vector. The parameter of interest is $\theta = (\phi, \beta)$, and the constant c identifies two levels of variance in the model. From (13), the AQL estimate of θ is the solution of the following asymptotic quasi-score normal equation

$$\mathbf{G}^*_{T,h,n}(\beta,\phi) = \sum_{t=1}^{T} \begin{pmatrix} -e^{\beta+\alpha_t} & 0 \\ 0 & -\alpha_t \end{pmatrix} \hat{\mathbf{\Sigma}}_{t,n}^{h^{-1}} \begin{pmatrix} y_{1t} - e^{\beta+\alpha_t} \\ y_{2t} - \phi\alpha_t \end{pmatrix} = 0. \tag{16}$$

in which

$$\hat{\mathbf{\Sigma}}_{t,n}^{h} = \begin{pmatrix} \hat{\sigma}_n(y_{1t}) & \hat{\sigma}_n(y_{1t}, y_{2t}) \\ \hat{\sigma}_n(y_{2t}, y_{1t}) & \hat{\sigma}_n(y_{2t}) \end{pmatrix}$$

is the NW type kernel estimator based on an initial estimate $\hat{\theta}^{(0)}$. An iterative algorithm is used to solve the AQL normal equations (16). The initial parameter used by the kernel estimator $\hat{\mathbf{\Sigma}}_{t,n}^{h}$ is the OLS estimate of θ. After obtaining an AQL estimate, say $\hat{\theta}^{(1)}$, the normal equation and the kernel estimator $\hat{\mathbf{\Sigma}}_{t,n}^{h}$ are updated and the iteration continuous until it converges. Our experience shows that the algorithm usually converges after three or four iterations.

To demonstrate the described above estimation procedure, we carried out a simulation study on (14) and (15). The simulation is carried as follows:

Firstly, independently simulate 1000 samples with size 200 from (14) and (15) based on the true value of parameter $\theta = (\phi, \beta)$. After series $\{y_{1t}\}$ and $\{y_{2t}\}$ are generated, we pretend that the value of $\theta = (\phi, \beta)$ is unknown. Then apply the above estimation procedure to $\{y_{1t}\}$ and $\{y_{2t}\}$ to obtain the estimation of $\theta = (\phi, \beta)$. In the following studies, we consider several different parameter settings for $\theta = (\phi, \beta)$ for c=0.01 and 0.1 respectively. The mean and root mean squared errors for $\hat{\phi}$, $\hat{\beta}$ based on the 1000 independent samples are shown in Tables 1 and 2. We notice that the AQL method tends to produces efficient estimate for all set of parameters, especially for c=0.1. In Table 3, the simulation results also indicate that, the larger the sample size is, the smaller the root mean squared error will be.

5 APPLICATION TO FOREIGN EXCHANGE DATA

We consider an application of the proposed AQL method to a real data set in this section. The data set contains the daily return of $z_{1,t} = USD/AUD$ (US Dollar/Australian Dollar) and $z_{2,t} = GBP/AUS$ (British Pound/Australian Dollar) for the period from 1/1/2003 to 1/1/2006, 921 observations in total (see $http://www.rba.gov.au/Statistics/exchange-rates.html$). Both $z_{1,t}$ and $z_{2,t}$ appear not to be stationary, as indicated in Figures 1 and 2. We took the nature logarithm of $z_{i,t}$, and let $y_{i,t} = \log(z_{i,t}/z_{i,t-1})$, $i = 1, 2$ and $t = 1, 2, 3, \ldots, 921$.

GARCH(1,1) is a popular econometric model for finance data (see Gourié -roux, 1997). Interesting applications of the exact quasi-likelihood approach to real data via a GARCH model can be found in Li and Turtle (2000). Existing techniques for parameter estimation in GARCH models are mainly maximum likelihood based. This means that the probability structure of $\{y_{i,t}\}$ has to be known. Usually it assume $\{y_{i,t}\}$ has conditional Gaussian distribution. This concern is very valid in finance as empirical data reveal fat-tailness and skewness which contradicts to the conditional normality. Therefore, it might lead estimation procedure to be exposed to modelling errors.

FIGURE 1: The plot of the daily returns of $z_{1,t} = USD/AUD$ (US Dollar/Australian Dollar)

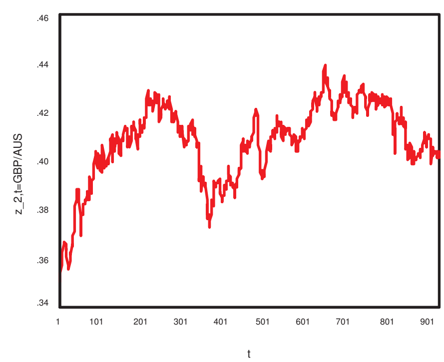

FIGURE 2: The plot of the daily returns of $z_{2,t} = GBP/AUS$ (British Pound/Australian Dollar)

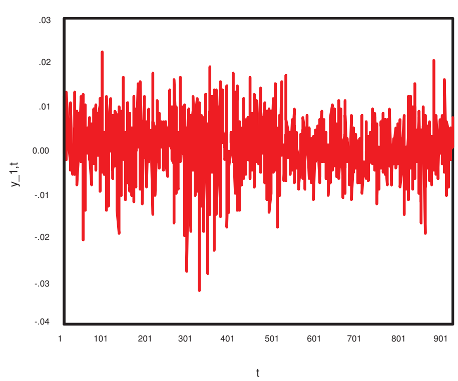

FIGURE 3: The plot of $y_{1,t} = \log(z_{1,t}/z_{1,t-1})$

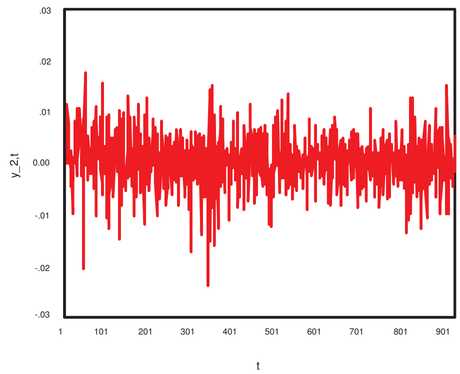

FIGURE 4: The plot of $y_{2,t} = \log(z_{2,t}/z_{2,t-1})$

We used the S+FinMetrics function archTest to carry out Lagrange Multiplier test for the presence of ARCH effects in the residuals (see; Zivot and Wang 2006). For both $z_{1,t}, z_{2,t}$ the p-value are significant ($<$ 0.05 level), so reject the null hypothesis that there are no ARCH effects and we fit $\{y_{i,t}\}$ by following models:

$$y_{1,t} = \theta_{1,0} + \theta_{1,1} y_{1,t-1} + f_1(y_{1,t-1})\varepsilon_{1,t}, \tag{17}$$

$$y_{2,t} = \theta_{2,0} + \theta_{2,1} y_{2,t-1} + f_2(y_{2,t-1})\varepsilon_{2,t}, \tag{18}$$

where $E(\varepsilon_{i,t}) = 0$, $Var(\varepsilon_{i,t}) = \sigma^2_{\varepsilon_i}$. Let $\epsilon_{i,t} = f_i(y_{i,t-1})\varepsilon_{i,t}, i = 1, 2$, and consider the following martingale difference

$$\begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} = \begin{pmatrix} y_{1,t} - \theta_{1,0} + \theta_{1,1} y_{1,t-1} \\ y_{2,t} - \theta_{2,0} + \theta_{2,1} y_{2,t-1} \end{pmatrix}, t = 1, 2, \cdots, 921,$$

where the parameter of interest is $\theta = (\theta_{1,0}, \theta_{1,1}, \theta_{2,0}, \theta_{2,1})$. In the following we apply the proposed AQL method to (17) and (18) without assuming conditional normality on $(\varepsilon_{1,t}, \varepsilon_{2,t})'$. The AQL estimate of θ is the solution of the following asymptotic quasi-score normal equation

$$G_t(\theta) = \sum_{t=1}^{T} \begin{pmatrix} 1 & 0 \\ y_{1,t-1} & 0 \\ 0 & 1 \\ 0 & y_{2,t-1} \end{pmatrix} \hat{\mathbf{\Sigma}}^{h^{-1}}_{t,n} \begin{pmatrix} y_{1,t} - \theta_{1,0} + \theta_{1,1} y_{1,t-1} \\ y_{2,t} - \theta_{2,0} + \theta_{2,1} y_{2,t-1} \end{pmatrix} = 0. \tag{19}$$

in which

$$\hat{\boldsymbol{\Sigma}}^{h}_{t,n} = \begin{pmatrix} \hat{\sigma}_n(y_{1,t}) & \hat{\sigma}_n(y_{1,t}, y_{2,t}) \\ \hat{\sigma}_n(y_{2,t}, y_{1,t}) & \hat{\sigma}_n(y_{2,t}) \end{pmatrix}$$

is the NW type kernel estimator based on an initial assigned parameter $\hat{\theta}^{(0)}$. Iterative algorithm is used to solve the AQL normal equations (19). The initial assigned parameter is always chosen as the OLS estimate of θ. After obtaining an AQL estimate, say $\hat{\theta}^{(1)}$, the normal equation and the kernel estimator $\hat{\boldsymbol{\Sigma}}^{h}_{t,n}$ will be updated. The iteration continuous until it converged. Our experience shows that the algorithm converges after three iterations.

In determining the NW type kernel estimate for $\hat{\boldsymbol{\Sigma}}^{h}_{t,n}$, 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

$$\hat{h}_{os} = (\frac{243R(K)}{35\mu_2(K)^2 n})^{1/5} s \tag{20}$$

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, 1995). Estimates of the parameters in (17) and (18) are presented in Table 4. We also present the standardized residual for both methods AQL and OLS.

We can see from the third column in Table 4 that AQL gives smaller standardized residual when bandwidth $h \leq \hat{h}_{os}$, and has the similar values of standardized residual given by OLS when the bandwidth $h > \hat{h}_{os}$. That means, the AQL method tends to be more efficient than OLS method especially when an appropriate bandwidth is chosen.

The above studies show that the asymptotic quasi-likelihood method combining with kernel smoothing technique will provide an efficient approach for estimating unknown parameter θ when the exactly probability structure of underlying model is unknown. It will provide a robust tool for obtaining optimal point estimate of parameters in heteroscedastic models, like GARCH models.

6 CONCLUSION

In this paper an alternative approach the AQL method to estimate the parameters in multivariate heteroscedastic models with unspecific correlation is given. Results from the simulation and empirical studies indicate that the AQL method can provide efficient estimate of parameter. The study also shows that the AQL estimating procedure for model (1) is easy to be implemented comparing with other approaches, especially when the system probability structure can not be fully specified. By utilising the nonparametric kernel estimator of variance-covariance 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.

REFERENCE

Bauwens, L. Laurent, S. and Rombouts, J. V. K. (2006). Multivariate GARCH models: a survey, *Journal of Applied Econometrics*, **21**, 79-109.

Bosq, D. (1998). Nonparametric Statistics for Stochastic Processes, *Lecture Note in Statistics*, **110**. Heidelberg: Springer-Verlag.

Gouriéroux, C. (1997). *ARCH Models and Financial Applications*, Springer-Verlag.

Härdle, W. (1990). *Applied Nonparametric Regression*, Cambridge, University Press.

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

Li, D. X. and Turtle, H.L. (2000). Semiparametric ARCH models: an estimating function approach, *Journal of Business & Economic Statistics*, **12**, 174-186.

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

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

Loève, M. (1963). *Probability theory*, D. Van Nostrand Company, New Jersey.

Lu, Z. D. and Jiang, Z. (2001). L_1 geometric ergodicity of a multivariate nonlinear AR model with an ARCH term, *Statist. Probab. Lett.*, **51**, 121-130.

Meyn, S. P. and Tweedie, R. L. (1993). *Markov Chains and Stochastic Stability*. Springer-Verlag.

Wand, M.P. and Jones, M.C.(1995). *Kernel Smoothing*, Chapman & Hall, New York.

Zivot, E. and Wang, J. (2006). *Modeling financial time series with S-PLUS*, Springer, New York.

A APPENDIX

In this section, the proof of Theorem 1 is given. In the proofs we need the mixing properties of geometric ergodicity of Markov chain.

In Lu and Jiang (2001) result, under Conditions C1-C4, the process $\mathbf{y}_t$ satisfying model (1) is geometrically ergodic. That means that there is a ergodic stationary measure $\pi(\cdot)$ on R^P given in (9). Let $\{\mathbf{y}_t^*\}_{t=0}^{\infty}$ is a stationary process according to the model (1) and the initial $\mathbf{y}_0$ is distributed with the ergodic stationary measure $\pi(\cdot)$ given in (9). Then, $\{\mathbf{y}_t^*\}$ is strictly stationary with invariance measure $\pi(\cdot)$ (Meyn and Tweedie, 1993, Chapter 10). Let $\mathbf{x}_t^* = (\mathbf{y}_{t-q}^*, \ldots, \mathbf{y}_{t-1}^*)$, and f be the a joint probability density function of $\mathbf{x}_t^*$. Furthermore, let $\phi(\mathbf{x}_{it}, \mathbf{x}_{jt}) = g(\mathbf{x}_{it}, \mathbf{x}_{jt}) f(\mathbf{x}_{it}, \mathbf{x}_{jt})$ and $\hat{\phi}_{n,h}^*(\mathbf{x}_{it}, \mathbf{x}_{jt})$ and $\hat{f}_{n,h}^*(\mathbf{x}_{it}, \mathbf{x}_{jt})$ be versions of $\hat{\phi}(\mathbf{x}_{it}, \mathbf{x}_{jt})$ and $\hat{f}(\mathbf{x}_{it}, \mathbf{x}_{jt})$ based on $\{y_{it}^*\}$ and $\{y_{jt}^*\}$ instead of the original $\{y_{it}\}$ and $\{y_{jt}\}$, that is,

$$\begin{aligned} \hat{\phi}_{n,h}^*(\mathbf{x}_{it}, \mathbf{x}_{jt}) &= (nh^q)^{-1} \sum_{s=q+1}^{n} K\left(\frac{\mathbf{x}_{it} - \mathbf{x}_{is}^*}{h}\right) {v^*}_{i\mathbf{x}_s} \\ \hat{f}_{n,h}^*(\mathbf{x}_{it}, \mathbf{x}_{jt}) &= (nh^q)^{-1} \sum_{s=q+1}^{n} K\left(\frac{\mathbf{x}_{it} - \mathbf{x}_{is}^*}{h}\right) \end{aligned} \tag{21}$$

and for $i \neq j$,

$$\begin{aligned} \hat{\phi}_{n,h}^*(\mathbf{x}_{it}, \mathbf{x}_{jt}) &= (nh^{2q})^{-1} \sum_{s=q+1}^{n} K\left(\frac{\mathbf{x}_{it} - \mathbf{x}_{is}^*}{h}\right) K\left(\frac{\mathbf{x}_{jt} - \mathbf{x}_{js}^*}{h}\right) {v^*}_{ij\mathbf{x}_s} \\ \hat{f}_{n,h}^*(\mathbf{x}_{it}, \mathbf{x}_{jt}) &= (nh^{2q})^{-1} \sum_{s=q+1}^{n} K\left(\frac{\mathbf{x}_{it} - \mathbf{x}_{is}^*}{h}\right) K\left(\frac{\mathbf{x}_{jt} - \mathbf{x}_{jt}^*}{h}\right). \end{aligned} \tag{22}$$

Lemma A.1 *Under Conditions C1-C7*

$$\sup_{\mathbf{x}_{it}, \mathbf{x}_{jt} \in S} |\hat{\phi}_{n,h}^*(\mathbf{x}_{it}, \mathbf{x}_{jt}) - \phi(\mathbf{x}_{it}, \mathbf{x}_{jt})| \xrightarrow{p} 0$$

and

$$\sup_{\mathbf{x}_{it}, \mathbf{x}_{jt} \in S} |\hat{f}_{n,h}^*(\mathbf{x}_{it}, \mathbf{x}_{jt}) - f(\mathbf{x}_{it}, \mathbf{x}_{jt})| \xrightarrow{p} 0,$$

as $n \to \infty$.

Proof: The first part of the lemma is proved if we can establish

$$\sup_{\mathbf{x}_{it},\mathbf{x}_{jt}\in S} |\hat{\phi}^*_{n,h}(\mathbf{x}_{it},\mathbf{x}_{jt}) - E\{\hat{\phi}^*_{n,h}(\mathbf{x}_{it},\mathbf{x}_{jt})\}| \xrightarrow{p} 0, \tag{23}$$

and

$$\sup_{\mathbf{x}_{it},\mathbf{x}_{jt}\in S} |E\{\hat{\phi}^*_{n,h}(\mathbf{x}_{it},\mathbf{x}_{jt})\} - \phi(\mathbf{x}_{it},\mathbf{x}_{jt})| \to 0, \tag{24}$$

as $n \to \infty$. To prove (24), when $i = j$, we notice that

$$E\{\hat{\phi}^*_{n,h}(\mathbf{x}_{it})\} - \phi(\mathbf{x}_{it})$$

$$= \frac{n-q-1}{n} \int K(\mathbf{z}_{it})\{\phi(\mathbf{x}_{it} - h\mathbf{z}_{it}) - \phi(\mathbf{x}_{it})\}d\mathbf{z}_{it} - \frac{q+1}{n}\phi(\mathbf{x}_{it}).$$

As ϕ is Lipchitz continuous,

$$\sup_{\mathbf{x}_{it}\in S} |E\{\hat{\phi}^*_{n,h}(\mathbf{x}_{it})\} - \phi(\mathbf{x}_{it})| \le C[h \int |\mathbf{z}_{it}| K(\mathbf{z}_{it}) d\mathbf{z}_{it} + \frac{q+1}{n}] \to 0$$

as $n \to \infty$ for some constant $C > 0$. When $i \ne j$, we notice that

$$E\{\hat{\phi}^*_{n,h}(\mathbf{x}_{it},\mathbf{x}_{jt})\} - \phi(\mathbf{x}_{it},\mathbf{x}_{jt}) = \frac{n-q-1}{n} \int\int K(\mathbf{z}_{it})K(\mathbf{z}_{jt})\{\phi(\mathbf{x}_{it} - h\mathbf{z}_{it},$$

$$\mathbf{x}_{jt} - h\mathbf{z}_{jt}) - \phi(\mathbf{x}_{it},\mathbf{x}_{jt})\}d\mathbf{z}_{it}d\mathbf{z}_{jt} - \frac{q+1}{n}\phi(\mathbf{x}_{it},\mathbf{x}_{jt}).$$

As ϕ is Lipchitz continuous, and (by Lemma 1.3 Bosq, 1998),

$$\sup_{\mathbf{x}_{it},\mathbf{x}_{jt}\in S} |E\{\hat{\phi}^*_{n,h}(\mathbf{x}_{it},\mathbf{x}_{jt})\} - \phi(\mathbf{x}_{it},\mathbf{x}_{jt})|$$

$$\le C[h \int\int (|\mathbf{z}_{it}|^2 + |\mathbf{z}_{jt}|^2)^{1/2} K(\mathbf{z}_{it})K(\mathbf{z}_{jt}) d\mathbf{z}_{it} d\mathbf{z}_{jt} + \frac{q+1}{n}] \to 0$$

as $n \to \infty$ for some constant $C > 0$. This proves (24).

When $i = j$, let $M_n = b_0 \log(n)$,

$$\begin{aligned} I^+(\mathbf{x}_{it}) &= (nh^q)^{-1} \sum_{s=q+1}^{n} K\left(\frac{\mathbf{x}_{it} - \mathbf{x}^*_{is}}{h}\right) v^*_{i\mathbf{x}_s} I(v^*_{i\mathbf{x}_s} \ge M_n) \quad \text{and} \\ I^-(\mathbf{x}_{it}) &= (nh^q)^{-1} \sum_{s=q+1}^{n} K\left(\frac{\mathbf{x}_{it} - \mathbf{x}^*_{is}}{h}\right) v^*_{i\mathbf{x}_s} I(v^*_{i\mathbf{x}_s} \le M_n). \end{aligned}$$

As $\sup_{\mathbf{x}_{it}\in S}|I^+(\mathbf{x}_{it})| \le C(nh^q)^{-1}\sum_{s=q+1}^n |v^*_{i\mathbf{x}_s}|I(v^*_{i\mathbf{x}_s} \ge M_n)$ for some $C > 0$, the Cauchy-Schwartz inequality implies that

$$E\left[\sup_{\mathbf{x}_{it}\in S}|I^+(\mathbf{x}_{it}) - E\{I^+(\mathbf{x}_{it})\}|\right] \le 2C(nh^q)^{-1}\sum_{s=q+1}^{n}\{E(|v^*_{i\mathbf{x}_s}|^2)P(v^*_{i\mathbf{x}_s} \ge M_n)\}^{1/2}.$$

From the Markov inequality and C4, for a positive constant η_0,

$$P\left[M_n^{-1}(nh^q)^{1/2}\sup_{\mathbf{x}_{it}\in S}|I^+(\mathbf{x}_{it}) - E\{I^+(\mathbf{x}_{it})\}| \ge \eta_0)\right]$$
$$\le 2C\eta_0^{-1}n^{-1/2}h^{-q/2}M_n^{-1}(n-q)\exp\{-a_0b_0\log(n)/2\}.$$

By properly choosing b_0, the RHS converges to zero as $n \to \infty$. This means that

$$\sup_{\mathbf{x}_{it}\in S}|I^+(\mathbf{x}_{it}) - E\{I^+(\mathbf{x}_{it})\}| = o_p\{(nh^q)^{-1/2}\log(n)\}. \tag{25}$$

Let $\phi_{t,n}(\mathbf{x}_{it}) = K(\frac{\mathbf{x}_{it}-\mathbf{x}^\star_{is}}{h})v^*_{is}I(v^*_{is} < M_n)$, $Z_{t,n}(\mathbf{x}_{it}) = \phi_{t,n}(\mathbf{x}_{it}) - E\{\phi_{t,n}(\mathbf{x}_{it})\}$. Clearly, at each fixed $\mathbf{x}_{it}$, $\{Z_{t,n}(\mathbf{x}_{it})\}$ has zero mean, is bounded by $b = C_1M_n$ where $|K(u)| \le C_1$ for all u. For $\eta > 0$, let $\epsilon = h^q\eta$, $q^* = M_n^3h^{-2q}$ and C denote a generic positive constant whose value may be changed. From Theorem 1.3 of Bosq (1998) and C2 and the geometric α-mixing,

$$\begin{aligned} &P\big[|I^-(\mathbf{x}_{it}) - E\{I^-(\mathbf{x}_{it})\}| > \eta\big] = P\left(|\sum Z_{t,n}(\mathbf{x}_{it})| > nh^q\eta\right) \\ \le\ & 4\exp[-\epsilon^2q^*/\{8b^2\}] + 22(1+4b/\epsilon)^{1/2}q^*\alpha\{[n/(2q^*)]\} \\ \le\ & 4\exp\{-\eta^2M_n/(8C_1^2)\} + Cb^{1/2}\epsilon^{-1/2}M_n^3h^{-2q}\rho_1^{nh^{2q}M_n^{-3}/2} \\ \le\ & 4n^{-Cb_0\eta^2} + CM_n^{7/2}n^{5q/(4r+2q)}\exp\{n^{\frac{(2r-q)}{(2r+q)}}M_n^{-3}log(\rho_1)/2\}. \end{aligned} \tag{26}$$

Note that the two terms on the RHS all converges to zero and are free of $\mathbf{x}_{it}$.

Let $\{B_{itk}\}_{k=1}^{v_{itn}}$ be a set of equal volume disjoint hypercubes with centers $\{s_{itk}\}_{k=1}^{v_{itn}}$ such that $S = \bigcup_{k=1}^{v_{itn}} B_{itk}$, $v_{itn} = [n^{t_0}]$ for some $t_0 > 0$ and $\sup_{\mathbf{x}_{it}\in B_{itk}}||\mathbf{x}_{it} - s_{itk}|| \le cv_{itn}^{-1}$. Based on this partition of S, and let $I^{-\star}(\mathbf{x}_{it}) = I^-(\mathbf{x}_{it}) - E\{I^-(\mathbf{x}_{it})\}$,

$$\sup_{\mathbf{x}_{it}\in S}|I^-(\mathbf{x}_{it}) - E\{I^-(\mathbf{x}_{it})\}| \le \max_{k=1,\ldots,v_{itn}}|I^{-\star}(s_{itk})| + \sup_{\mathbf{x}_{it}\in S}|I^{-\star}(\mathbf{x}_{it}) - I^{-\star}(s_{k(\mathbf{x}_{it})})|$$

where $k(\mathbf{x}_{it})$ being the index of the hypercube containing $\mathbf{x}_{it}$. Note that

$$P\{\max_{k=1,\ldots,v_{itn}} |I^{-\star}(s_{itk})| \geq \eta\} \leq n^{t_0} \sup_{\mathbf{x}_{it}\in S} P\{|I^{-}(\mathbf{x}_{it}) - E\{I^{-1}(\mathbf{x}_{it})\}| \geq \eta\},$$

By properly choosing b_0, (26) implies that

$$\max_{k=1,\ldots,v_{itn}} |I^{-\star}(s_{itk})| = o_p(1). \tag{27}$$

As K is Lipschitz continuous,

$$\sup_{\mathbf{x}_{it}\in S} |I^{-\star}(\mathbf{x}_{it}) - I^{-\star}(s_{k(\mathbf{x}_{it})})| \leq Ch^{-q-1}n^{-t_0}n^{-1}\left(\sum_{s=q+1}^{n} v^*_{i\mathbf{x}_s} + E(v^*_{i\mathbf{x}_s})\right).$$

Note that $n^{-1}\sum |v^*_{i\mathbf{x}_s}| \stackrel{w.s.}{\to} E|v^*_{i\mathbf{x}_s}|$, and $E|v^*_{i\mathbf{x}_s}| \leq C$. Then with probability one

$$\sup_{\mathbf{x}_{it}\in S} |I^{-\star}(\mathbf{x}_{it}) - I^{-\star}(s_{k(\mathbf{x}_{it})})| \leq Ch^{-q-1}n^{-t_0}.$$

By choosing $t_0 > (q+1)/(2r+q)$, we have

$$Pr\{\sup_{\mathbf{x}_{it}\in S} |I^{-\star}(\mathbf{x}_{it}) - I^{-\star}(s_{k(\mathbf{x}_{it})})| \geq \eta\} \to 0.$$

This together with (25) implies (23) and thus

$$\sup_{\mathbf{x}_{it},\mathbf{x}_{jt}\in S} |\hat{\phi}^{\star}_{n,h}(\mathbf{x}_{it},\mathbf{x}_{jt}) - \phi(\mathbf{x}_{it},\mathbf{x}_{jt})| \stackrel{p}{\to} 0, \quad i = j.$$

Now consider the case when $i \neq j$ and let $M_n = b_0 \log(n)$,

$$I^{+}(\mathbf{x}_{it},\mathbf{x}_{jt}) = (nh^{2q})^{-1} \sum_{s=q+1}^{n} K\left(\frac{\mathbf{x}_{it}-\mathbf{x}^{\star}_{is}}{h}\right) K\left(\frac{\mathbf{x}_{jt}-\mathbf{x}^{\star}_{js}}{h}\right) v^*_{ijs} I(v^*_{ijs} \geq M_n)$$

and

$$I^{-}(\mathbf{x}_{it},\mathbf{x}_{jt}) = (nh^{2q})^{-1} \sum_{s=q+1}^{n} K\left(\frac{\mathbf{x}_{it}-\mathbf{x}^{\star}_{is}}{h}\right) K\left(\frac{\mathbf{x}_{jt}-\mathbf{x}^{\star}_{js}}{h}\right) v^*_{ijs} I(v^*_{ijs} \leq M_n).$$

As $\sup_{\mathbf{x}_{it},\mathbf{x}_{jt}\in S} |I^{+}(\mathbf{x}_{it},\mathbf{x}_{jt})| \leq C(nh^{2q})^{-1}\sum_{s=q+1}^{n} |v^*_{ijs}| I(v^*_{ijs} \geq M_n)$ for some $C > 0$, the Cauchy-Schwartz inequality implies that

$$E\left[\sup_{\mathbf{x}_{it},\mathbf{x}_{jt}\in S} |I^{+}(\mathbf{x}_{it},\mathbf{x}_{jt}) - E\{I^{+}(\mathbf{x}_{it},\mathbf{x}_{jt})\}|\right]$$

$$\le 2C(nh^{2q})^{-1}\sum_{s=q+1}^{n}\{E(|v^*_{ijs}|^2)P(v^*_{ijs}\ge M_n)\}^{1/2}.$$

From the Markov inequality and C4, for a positive constant η_0,

$$P\left[M_n^{-1}(nh^{2q})^{1/2}\sup_{\mathbf{x}_{it},\mathbf{x}_{jt}\in S}|I^+(\mathbf{x}_{it},\mathbf{x}_{jt})-E\{I^+(\mathbf{x}_{it},\mathbf{x}_{jt})\}|\ge\eta_0)\right]$$

$$\le 2C\eta_0^{-1}n^{-1/2}h^{-q}M_n^{-1}(n-q-1)\exp\{-a_0b_0\log(n)/2\}.$$

By properly choosing b_0, the RHS converges to zero as $n\to\infty$. This means that

$$\sup_{\mathbf{x}_{it},\mathbf{x}_{jt}\in S}|I^+(\mathbf{x}_{it},\mathbf{x}_{jt})-E\{I^+(\mathbf{x}_{it},\mathbf{x}_{jt})\}|=o_p\{(nh^{2q})^{-1/2}\log(n)\}. \tag{28}$$

Let

$$\phi_{t,n}(\mathbf{x}_{it},\mathbf{x}_{jt})=K(\frac{\mathbf{x}_{it}-\mathbf{x}^\star_{is}}{h})K(\frac{\mathbf{x}_{jt}-\mathbf{x}^\star_{js}}{h})v^*_{ijs}I(v^*_{ijs}<M_n),$$

and

$$Z_{t,n}(\mathbf{x}_{it},\mathbf{x}_{jt})=\phi_{t,n}(\mathbf{x}_{it},\mathbf{x}_{jt})-E\{\phi_{t,n}(\mathbf{x}_{it},\mathbf{x}_{jt})\}.$$

Clearly, at each fixed $\mathbf{x}_{it},\mathbf{x}_{jt}$, $\{Z_{t,n}(\mathbf{x}_{it},\mathbf{x}_{jt})\}$ has zero mean, is bounded by $b=C_1M_n$ where $|K(u)|\le C_1$ for all u. For $\eta>0$, let $\epsilon=h^q\eta$, $q^*=M_n^3h^{-2q}$ and C denote a generic positive constant whose value may be changed. From Theorem 1.3 of Bosq (1998) and C2 and the geometric α-mixing,

$$\begin{aligned}
&P\left[|I^-(\mathbf{x}_{it},\mathbf{x}_{jt})-E\{I^-(\mathbf{x}_{it},\mathbf{x}_{jt})\}|>\eta\right]=P\left(|\sum Z_{t,n}(\mathbf{x}_{it},\mathbf{x}_{jt})|>nh^{2q}\eta\right)\\
\le\;&4\exp[-\epsilon^2q^*/\{8b^2\}]+22(1+4b/\epsilon)^{1/2}q^*\alpha\{[n/(2q^*)]\}\\
\le\;&4\exp\{-\eta^2M_n/(8C_1^2)\}+Cb^{1/2}\epsilon^{-1/2}M_n^3h^{-2q}\rho_1^{nh^{2q}M_n^{-3}/2}\\
\le\;&4n^{-Cb_0\eta^2}+CM_n^{7/2}n^{5q/(4r+2q)}\exp\{n^{\frac{(2r-q)}{(2r+q)}}M_n^{-3}log(\rho_1)/2\}.
\end{aligned}\tag{29}$$

Note that the two terms on the RHS all converges to zero and are free of $\mathbf{x}_{it},\mathbf{x}_{jt}$.

Let $\{B_{itk}\}_{k=1}^{v_{itn}}$ be a set of equal volume disjoint hypercubes with centers $\{s_{itk}\}_{k=1}^{v_{itn}}$ such that $S=\bigcup_{k=1}^{v_{itn}}B_{itk}$, $v_{itn}=[n^{t_0}]$ for some $t_0>0$ and

$\sup_{\mathbf{x}_{it}\in B_{itk}} ||\mathbf{x}_{it}-s_{itk}|| \leq cv_{itn}^{-1}$. Based on this partition of S, and let $I^{-\star}(\mathbf{x}_{it}, \mathbf{x}_{jt}) = I^{-}(\mathbf{x}_{it}, \mathbf{x}_{jt}) - E\{I^{-}(\mathbf{x}_{it}, \mathbf{x}_{jt})\}$

$$\sup_{\mathbf{x}_{it},\mathbf{x}_{jt}\in S} |I^{-}(\mathbf{x}_{it}, \mathbf{x}_{jt}) - E\{I^{-}(\mathbf{x}_{it}, \mathbf{x}_{jt})\}| \leq \max_{k=1,\ldots,v_{itn},k=1,\ldots,v_{jtn}} |I^{-\star}(s_{itk}, s_{jtk})|$$

$$+ \sup_{\mathbf{x}_{it},\mathbf{x}_{jt}\in S} |I^{-\star}(\mathbf{x}_{it}, \mathbf{x}_{jt}) - I^{-\star}(s_{k(\mathbf{x}_{it})}, s_{k(\mathbf{x}_{jt})})|$$

where $k(\mathbf{x}_{it})$ and $k(\mathbf{x}_{jt})$ being the index of the hypercube containing $\mathbf{x}_{it}$ and $\mathbf{x}_{jt}$ respectively. Note that

$$P\{\max_{k=1,\ldots,v_{itn},k=1,\ldots,v_{jtn}} |I^{-\star}(s_{itk}, s_{jtk})| \geq \eta\}$$

$$\leq 2n^{t_0} \sup_{\mathbf{x}_{it},\mathbf{x}_{jt}\in S} P\{|I^{-}(\mathbf{x}_{it}, \mathbf{x}_{jt}) - E\{I^{-1}(\mathbf{x}_{it}, \mathbf{x}_{jt})\}| \geq \eta\}.$$

By properly choosing b_0, (29) implies that

$$\max_{k=1,\ldots,v_{itn},k=1,\ldots,v_{jtn}} |I^{-\star}(s_{itk}, (s_{jtk})| = o_p(1). \tag{30}$$

As K is Lipschitz continuous,

$$\sup_{\mathbf{x}_{it},\mathbf{x}_{jt}\in S} |I^{-\star}(\mathbf{x}_{it}, \mathbf{x}_{jt}) - I^{-\star}(s_{k(\mathbf{x}_{it})}, s_{k(\mathbf{x}_{jt})})|$$

$$\leq Ch^{-2q-1}n^{-t_0}n^{-1}\left(\sum_{s=q+1}^{n} v_{ijs}^* + E(v_{ijs}^*)\right).$$

Note that $n^{-1}\sum |v_{ijs}^*| \overset{w.s.}{\to} E|v_{ijs}^*|$, and $E|v_{ijs}^*| \leq C$. Then with probability one

$$\sup_{\mathbf{x}_{it}\in S} |I^{-\star}(\mathbf{x}_{it}, \mathbf{x}_{jt}) - I^{-\star}(s_{k(\mathbf{x}_{it})}, s_{k(\mathbf{x}_{jt})})| \leq Ch^{-2q-1}n^{-t_0}.$$

By choosing $t_0 > (2q+1)/(2r+q)$, we have

$$Pr\{\sup_{\mathbf{x}_{it},\mathbf{x}_{jt}\in S} |I^{-\star}(\mathbf{x}_{it}, \mathbf{x}_{jt}) - I^{-\star}(s_{k(\mathbf{x}_{it})}, s_{k(\mathbf{x}_{jt})})| \geq \eta\} \to 0.$$

This together with (28) implies (23) and thus

$$\sup_{\mathbf{x}_{it},\mathbf{x}_{jt}\in S} |\hat{\phi}_{n,h}^{\star}(\mathbf{x}_{it}, \mathbf{x}_{jt}) - \phi(\mathbf{x}_{it}, \mathbf{x}_{jt})| \overset{p}{\to} 0.$$

The proof of $\sup_{\mathbf{x}_{it},\mathbf{x}_{jt}\in S} |\hat{f}_{n,h}^{\star}(\mathbf{x}_{it}, \mathbf{x}_{jt}) - f(\mathbf{x}_{it}, \mathbf{x}_{jt})| \overset{p}{\to} 0$ for $i \neq j$ is similar and omitted.

Lemma A.2 *Under conditions C1-C7,*

$$\sup_{\mathbf{x}_{it},\mathbf{x}_{jt}\in S} |\hat{\phi}^{\star}_{n,h}(\mathbf{x}_{it},\mathbf{x}_{jt}) - \hat{\phi}_{n,h}(\mathbf{x}_{it},\mathbf{x}_{jt})| \xrightarrow{p} 0$$

and

$$\sup_{\mathbf{x}_{it},\mathbf{x}_{jt}\in S} |\hat{f}^{\star}_{n,h}(\mathbf{x}_{it},\mathbf{x}_{jt}) - \hat{f}_{n,h}(\mathbf{x}_{it},\mathbf{x}_{jt})| \xrightarrow{p} 0,$$

as $n \to \infty$.

Proof: When $i = j$, let $M_n = b_0 \log(n)$, $\xi^{+}_{is} = K\left(\frac{\mathbf{x}_{it}-\mathbf{x}_{is}}{h}\right) v^{*}_{i\mathbf{x}_s} I(v^{*}_{i\mathbf{x}_s} \geq M_n)$, $\xi^{-}_{is} = K\left(\frac{\mathbf{x}_{it}-\mathbf{x}_{is}}{h}\right) v^{*}_{i\mathbf{x}_s} I(v^{*}_{i\mathbf{x}_s} < M_n)$, $\xi^{\star+}_{is} = K\left(\frac{\mathbf{x}_{it}-\mathbf{x}^{\star}_{is}}{h}\right) v^{*}_{i\mathbf{x}_s} I(v^{*}_{i\mathbf{x}_s} \geq M_n)$ and $\xi^{\star+}_{is} = K\left(\frac{\mathbf{x}_{it}-\mathbf{x}^{\star}_{is}}{h}\right) v^{*}_{i\mathbf{x}_s} I(v^{*}_{i\mathbf{x}_s} < M_n)$. Let

$$J^{+}(\mathbf{x}_{it}) = (nh^{q})^{-1} \sum_{s=q+1}^{n} J^{+}_{s}(\mathbf{x}_{it})$$

and

$$J^{-}(\mathbf{x}_{it}) = (nh^{q})^{-1} \sum_{s=q+1}^{n} J^{-}_{s}(\mathbf{x}_{it})$$

where

$$J^{+}_{s}(\mathbf{x}_{it}) = |\xi^{\star+}_{is} - \xi^{+}_{is}|$$

and

$$J^{-}_{s}(\mathbf{x}_{it}) = |\xi^{\star-}_{is} - \xi^{-}_{is}|$$

Clearly,

$$\begin{aligned} &\sup_{\mathbf{x}_{it},\mathbf{x}_{jt}\in S} |\hat{\phi}^{\star}_{n,h}(\mathbf{x}_{it},\mathbf{x}_{jt}) - \hat{\phi}_{n,h}(\mathbf{x}_{it},\mathbf{x}_{jt})| \\ &\leq \sup_{\mathbf{x}_{it},\mathbf{x}_{jt}\in S} |J^{+}(\mathbf{x}_{it})| + \sup_{\mathbf{x}_{it},\mathbf{x}_{jt}\in S} |J^{-}(\mathbf{x}_{it})|. \end{aligned}$$

Let γ_n be a sequence of integers tending to ∞ and $\gamma_n = o\{(nh^{q})M_n\}$. Then, from the geometric ergodicity of the Markov chain,

$$\begin{aligned} E\{J^{-}(\mathbf{x}_{it})\} &\leq (nh^{q})^{-1} \sum_{s=q+1}^{\gamma_n} 2||K||_{\infty} M_n \\ &\quad + (nh^{q})^{-1} ||K||_{\infty} \sum_{s=\gamma_n}^{n} ||\pi_{is} - \pi_i||_{tv} \\ &\leq C(nh^{q})^{-1} M_n(\gamma_n + \rho_1^{\gamma_n}), \end{aligned} \tag{31}$$

where $\pi_{is}(\cdot)$ is the probability measure of y_{is} whose density is $p_i(\cdot)$. From the Markov inequality, we immediately have

$$|J^-(\mathbf{x}_{it})| = o_p\{(nh^q)^{-1}M_n^2\gamma_n\}. \tag{32}$$

We also note that the left hand side of (31) is free of $\mathbf{x}_{it}$, and thus the order described in (32) is free of $\mathbf{x}_{it}$.

Then, choose a covering of S as in the proof of Lemma 1 for (23) and take the same route by properly choosing the order of γ_n, we have

$$\sup_{\mathbf{x}_{it}\in S} |J^-(\mathbf{x}_{it})| = o_p(1). \tag{33}$$

Using almost the same derivation that leading to (25) in the proof of Lemma 1, we can establish

$$\sup_{\mathbf{x}_{it},\mathbf{x}_{jt}\in S} |J^+(\mathbf{x}_{it})| = o_p\{(nh^q)^{-1/2}\gamma_n \log^2(n)\}.$$

This together with (33) implies $\sup_{\mathbf{x}_{it},\mathbf{x}_{jt}\in S} |\hat{\phi}^\star_{n,h}(\mathbf{x}_{it},\mathbf{x}_{jt}) - \hat{\phi}_{n,h}(\mathbf{x}_{it},\mathbf{x}_{jt})| \xrightarrow{p} 0$.

When $i \neq j$, let $M_n = b_0 \log(n)$,

$$\xi^+_{ijs} = K\left(\frac{\mathbf{x}_{it}-\mathbf{x}_{is}}{h}\right) K\left(\frac{\mathbf{x}_{jt}-\mathbf{x}_{js}}{h}\right) v^*_{ij\mathbf{x}_s} I(v^*_{ij\mathbf{x}_s} \geq M_n),$$

$$\xi^-_{ijs} = K\left(\frac{\mathbf{x}_{it}-\mathbf{x}_{is}}{h}\right) K\left(\frac{\mathbf{x}_{jt}-\mathbf{x}_{js}}{h}\right) v^*_{ij\mathbf{x}_s} I(v^*_{ij\mathbf{x}_s} < M_n),$$

$$\xi^{\star+}_{ijs} = K\left(\frac{\mathbf{x}_{it}-\mathbf{x}^\star_{is}}{h}\right) K\left(\frac{\mathbf{x}_{jt}-\mathbf{x}^\star_{js}}{h}\right) v^*_{ij\mathbf{x}_s} I(v^*_{ij\mathbf{x}_s} \geq M_n),$$

and

$$\xi^{\star-}_{ijs} = K\left(\frac{\mathbf{x}_{it}-\mathbf{x}^\star_{is}}{h}\right) K\left(\frac{\mathbf{x}_{jt}-\mathbf{x}^\star_{js}}{h}\right) v^*_{ij\mathbf{x}_s} I(v^*_{ij\mathbf{x}_s} < M_n).$$

Let

$$J^+(\mathbf{x}_{it},\mathbf{x}_{jt}) = (nh^{2q})^{-1} \sum_{s=q+1}^{n} J^+_s(\mathbf{x}_{it},\mathbf{x}_{jt})$$

and

$$J^-(\mathbf{x}_{it},\mathbf{x}_{jt}) = (nh^q)^{-1} \sum_{s=q+1}^{n} J^-_s(\mathbf{x}_{it},\mathbf{x}_{jt})$$

where

$$J_s^+(\mathbf{x}_{it}, \mathbf{x}_{jt}) = |\xi_{ijs}^{\star +} - \xi_{ijs}^+|$$

and

$$J_s^-(\mathbf{x}_{it}, \mathbf{x}_{jt}) = |\xi_{ijs}^{\star -} - \xi_{ijs}^-|$$

Clearly,

$$\begin{aligned} &\sup_{\mathbf{x}_{it}, \mathbf{x}_{jt} \in S} |\hat{\phi}_{n,h}^\star(\mathbf{x}_{it}, \mathbf{x}_{jt}) - \hat{\phi}_{n,h}(\mathbf{x}_{it}, \mathbf{x}_{jt})| \\ \leq &\sup_{\mathbf{x}_{it}, \mathbf{x}_{jt} \in S} |J^+(\mathbf{x}_{it}, \mathbf{x}_{jt})| + \sup_{\mathbf{x}_{it}, \mathbf{x}_{jt} \in S} |J^-(\mathbf{x}_{it}, \mathbf{x}_{jt})|. \end{aligned}$$

Let γ_n be a sequence of integers tending to ∞ and $\gamma_n = o\{(nh^q)M_n\}$. Then, from the geometric ergodicity of the Markov chain,

$$\begin{aligned} E\{J^-(\mathbf{x}_{it}, \mathbf{x}_{jt})\} &\leq (nh^{2q})^{-1} \sum_{s=q+1}^{\gamma_n} 2||K_{it}||_\infty ||K_{jt}||_\infty M_n \\ &\quad + (nh^{2q})^{-1} ||K_{it}||_\infty ||K_{jt}||_\infty \sum_{s=\gamma_n}^{n} ||\pi_{ijs} - \pi_{ij}||_{tv} \\ &\leq C(nh^{2q})^{-1} M_n(\gamma_n + \rho_1^{\gamma_n}), \end{aligned} \tag{34}$$

where $\pi_s(\cdot)$ is the probability measure of $\mathbf{y_s}$. From the Markov inequality, we immediately have

$$|J^-(\mathbf{x}_{it}, \mathbf{x}_{jt})| = o_p\{(nh^{2q})^{-1} M_n^2 \gamma_n\}. \tag{35}$$

We also note that the left hand side of (34) is free of $\mathbf{x}_{it}, \mathbf{x}_{jt}$, and thus the order described in (35) is free of $\mathbf{x}_{it}, \mathbf{x}_{jt}$.

Then, choose a covering of S as in the proof of Lemma 1 for (23) and take the same route by properly choosing the order of γ_n, we have

$$\sup_{\mathbf{x}_{it}, \mathbf{x}_{jt} \in S} |J^-(\mathbf{x}_{it}, \mathbf{x}_{jt})| = o_p(1). \tag{36}$$

Using almost the same derivation that leading to (28) in the proof of Lemma 1, we can establish

$$\sup_{\mathbf{x}_{it}, \mathbf{x}_{jt} \in S} |J^+(\mathbf{x}_{it}, \mathbf{x}_{jt})| = o_p\{(nh^{2q})^{-1/2} \gamma_n \log^2(n)\}.$$

This together with (36) implies $\sup_{\mathbf{x}_{it},\mathbf{x}_{jt}\in S}|\hat{\phi}^{\star}_{n,h}(\mathbf{x}_{it},\mathbf{x}_{jt})-\hat{\phi}_{n,h}(\mathbf{x}_{it},\mathbf{x}_{jt})|\xrightarrow{p}0$. The proof for the rest of the lemma is almost the same.

Proof of Theorem3.1: Lemmas 1 and 2 imply that when $n\to\infty$,

$$\sup_{\mathbf{x}_{it},\mathbf{x}_{jt}\in S}|\hat{\phi}_{n,h}(\mathbf{x}_{it},\mathbf{x}_{jt})-\phi(\mathbf{x}_{it},\mathbf{x}_{jt})|\xrightarrow{p}0 \tag{37}$$

and

$$\sup_{\mathbf{x}_{it},\mathbf{x}_{jt}\in S}|\hat{f}_{n,h}(\mathbf{x}_{it},\mathbf{x}_{jt})-f(\mathbf{x}_{it},\mathbf{x}_{jt})|\xrightarrow{p}0. \tag{38}$$

Since

$$\hat{g}_{n,h}(\mathbf{x}_{it},\mathbf{x}_{jt})-g(\mathbf{x}_{it},\mathbf{x}_{jt})=\frac{\hat{g}_{n,h}(\mathbf{x}_{it},\mathbf{x}_{jt})\{f(\mathbf{x}_{it},\mathbf{x}_{jt})-\hat{f}_{n,h}(\mathbf{x}_{it},\mathbf{x}_{jt})\}}{f(\mathbf{x}_{it},\mathbf{x}_{jt})}+\frac{\hat{\phi}_{n,h}(\mathbf{x}_{it},\mathbf{x}_{jt})-\phi(\mathbf{x}_{it},\mathbf{x}_{jt})}{f(\mathbf{x}_{it},\mathbf{x}_{jt})},$$

$$\begin{aligned}&\sup_{\mathbf{x}_{it},\mathbf{x}_{jt}\in S}|\hat{g}_{n,h}(\mathbf{x}_{it},\mathbf{x}_{jt})-g(\mathbf{x}_{it},\mathbf{x}_{jt})|\\ \le&\frac{\sup_{\mathbf{x}_{it},\mathbf{x}_{jt}\in S}|\hat{g}_{n,h}(\mathbf{x}_{it},\mathbf{x}_{jt})|\sup_{\mathbf{x}_{it},\mathbf{x}_{jt}\in S}|\hat{f}_{n,h}(\mathbf{x}_{it},\mathbf{x}_{jt})-f(\mathbf{x}_{it},\mathbf{x}_{jt})|}{\inf_{\mathbf{x}_{it},\mathbf{x}_{jt}\in S}f(\mathbf{x}_{it},\mathbf{x}_{jt})}\\ &+\frac{\sup_{\mathbf{x}_{it},\mathbf{x}_{jt}\in S}|\hat{\phi}_{n,h}(\mathbf{x}_{it},\mathbf{x}_{jt})-\phi(\mathbf{x}_{it},\mathbf{x}_{jt})|}{\inf_{\mathbf{x}_{it},\mathbf{x}_{jt}\in S}f(\mathbf{x}_{it},\mathbf{x}_{jt})}.\end{aligned} \tag{39}$$

Let $M_n=b_0\log(n)$. Then, from C7,

$$\begin{aligned}P(\sup_{\mathbf{x}_{it},\mathbf{x}_{it}\in S}|\hat{g}_{n,h}(\mathbf{x}_{it},\mathbf{x}_{it})|>M_n) &\le P(sup_{1\le t\le n}v_{i\mathbf{x}_t}>M_n)\\ &\le \exp(-a_0M_n)\sum_{t=q+1}^{n}E\{\exp(a_0v_{ij\mathbf{x}_t})\}\\ &\le C_3n^{-a_0b_0+1}.\end{aligned} \tag{40}$$

Choosing $b_0>a_0/2$, $\sum_{n=1}^{\infty}P(\sup_{\mathbf{x}_{it},\mathbf{x}_{jt}\in S}|\hat{g}_{n,h}(\mathbf{x}_{it},\mathbf{x}_{jt})|>M_n)<\infty$. From the Borel Cantelli Lemma, the probability of $\sup_{\mathbf{x}_{it},\mathbf{x}_{jt}\in S}|\hat{g}_{n,h}(\mathbf{x}_{it},\mathbf{x}_{jt})|>M_n$ being infinitely often is zero. This together with (37), (38), (39), and (40) finishes the proof of the theorem.

TABLE 1: Comparison of AQL and OLS estimates based on 1000 replication with c=0.01. Root mean square error of estimates are reported below each estimate.

	$\sigma_\eta = 0.675$		$\sigma_\eta = 0.484$		$\sigma_\eta = 0.308$	
	β	ϕ	β	ϕ	β	ϕ
true	-0.612	0.90	-0.612	0.95	-0.612	0.98
AQL	-0.6186	0.8996	-0.6183	0.9496	-0.6176	0.9795
	0.0952	0.0082	0.0934	0.0054	0.0866	0.0034
OLS	-0.6183	0.8996	-0.6178	0.9495	-0.6171	0.9795
	0.0953	0.0072	0.0937	0.0053	0.0870	0.0035
	$\sigma_\eta = 0.312$		$\sigma_\eta = 0.223$		$\sigma_\eta = 0.142$	
true	0.15	0.90	0.15	0.95	0.15	0.98
AQL	0.1470	0.8996	0.1468	0.9496	0.1481	0.9795
	0.0639	0.0082	0.0628	0.0054	0.0572	0.0032
OLS	0.1473	0.8996	0.1471	0.9495	0.1485	0.9795
	0.0640	0.0073	0.0630	0.0053	0.0575	0.0035
	$\sigma_\eta = 0.111$		$\sigma_\eta = 0.079$		$\sigma_\eta = 0.051$	
true	0.373	0.90	0.373	0.95	0.373	0.98
AQL	0.3707	0.8996	0.3707	0.9497	0.3716	0.9795
	0.0568	0.0082	0.0559	0.0054	0.0511	0.0034
OLS	0.3710	0.8996	0.3710	0.9495	0.3720	0.9795
	0.0568	0.0073	0.0561	0.0053	0.0514	0.0035

TABLE 2: Comparison of AQL and OLS estimates based on 1000 replication with c=0.1. Root mean square error of estimates are reported below each estimate.

	$\sigma_\eta = 0.675$		$\sigma_\eta = 0.484$		$\sigma_\eta = 0.308$	
	β	ϕ	β	ϕ	β	ϕ
true	-0.612	0.90	-0.612	0.95	-0.612	0.98
AQL	-0.6203	0.8927	-0.6195	0.9429	-0.6202	0.9736
	0.0958	0.0281	0.0940	0.0203	0.0905	0.0148
OLS	-0.6191	0.8889	-0.6173	0.9384	-0.6165	0.9685
	0.0958	0.0381	0.0946	0.0287	0.0912	0.0219
	$\sigma_\eta = 0.312$		$\sigma_\eta = 0.223$		$\sigma_\eta = 0.142$	
true	0.15	0.90	0.15	0.95	0.15	0.98
AQL	0.1458	0.8926	0.1453	0.9429	0.1425	0.9722
	0.0649	0.0281	0.0634	0.0203	0.0634	0.0158
OLS	0.1469	0.8889	0.1472	0.9385	0.1461	0.9672
	0.0649	0.0381	0.0633	0.0288	0.0637	0.0229
	$\sigma_\eta = 0.111$		$\sigma_\eta = 0.079$		$\sigma_\eta = 0.051$	
true	0.373	0.90	0.373	0.95	0.373	0.98
AQL	0.3699	0.8927	0.3690	0.9429	0.3659	0.9725
	0.0572	0.0281	0.0564	0.0203	0.0559	0.0159
OLS	0.3710	0.8889	0.3709	0.9385	0.3694	0.9681
	0.0572	0.0381	0.0563	0.0288	0.0558	0.0224

TABLE 3: Comparison of AQL1,AQL2, OLS1 and OLS2 estimates based on 1000 replication. Root mean square error of estimates are reported below each estimate.

	$\sigma_\eta = 0.079$		$\sigma_\eta = 0.079$		$\sigma_\eta = 0.079$		$\sigma_\eta = 0.079$	
	$n = 50$		$n = 100$		$n = 150$		$n = 200$	
	β	ϕ	β	ϕ	β	ϕ	β	ϕ
true	0.373	0.95	0.373	0.95	0.373	0.95	0.373	0.95
c=0.01								
AQL	0.3706	0.9496	0.3698	0.9496	0.3709	0.9494	0.3707	0.9497
	0.1004	0.0063	0.0770	0.0059	0.0646	0.0057	0.0559	0.0054
OLS	0.3711	0.9495	0.3702	0.9494	0.3714	0.9494	0.3710	0.9495
	0.1006	0.0062	0.0773	0.0058	0.0649	0.0056	0.0561	0.0053
c=0.1								
AQL	0.3653	0.9410	0.3658	0.9416	0.3681	0.9414	0.369	0.9429
	0.1059	0.0251	0.0785	0.0225	0.0660	0.0214	0.0564	0.0203
OLS	0.3697	0.9366	0.3688	0.9373	0.3703	0.9356	0.3709	0.9385
	0.1054	0.0311	0.0783	0.0303	0.0661	0.0318	0.0563	0.0288

TABLE 4: Comparison of AQL, OLS estimates based on the daily returns of (Australian Dollar/US Dollar) and (Australian Dollar/British Pound) for the period from 2003 to 2006.

		$\hat{\theta}_{i,0}$	$\hat{\theta}_{i,1}$	$\frac{\bar{\hat{\epsilon}}_{it}}{S.d(\hat{\epsilon}_{it})}$
	$OLS_{i=1}$	-0.000341	0.003471	0.099
	$OLS_{i=2}$	-0.000130	-0.111786	0.056
$h_1 = 2.0000$	$AQL_{i=1}$	-0.000341	0.003471	0.099
$h_2 = 2.0000$	$AQL_{i=2}$	-0.000130	-0.111786	0.056
$h_1 = 0.5000$	$AQL_{i=1}$	-0.000341	0.003471	0.099
$h_2 = 0.5000$	$AQL_{i=2}$	-0.000131	-0.111748	0.056
$h_1 = 0.0100$	$AQL_{i=1}$	-0.000336	0.010797	0.097
$h_2 = 0.0100$	$AQL_{i=2}$	-0.0001626	-0.101903	0.063
$h_{os1} = 0.003$	$AQL_{i=1}$	-0.000240	-0.004451	0.085
$h_{os2} = 0.002$	$AQL_{i=2}$	-0.000155	-0.020647	0.059
$h_1 = 0.0010$	$AQL_{i=1}$	-0.000252	0.003278	0.085
$h_2 = 0.0010$	$AQL_{i=2}$	-0.000121	-0.031063	0.053
$h_1 = 0.0005$	$AQL_{i=1}$	-0.000044	-0.024291	0.057
$h_2 = 0.0005$	$AQL_{i=2}$	-0.000121	-0.039834	0.053