Statistics and Probability Letters 103 (2015) 37–45

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Strong consistency of the maximum quasi-likelihood estimator in quasi-likelihood nonlinear models with stochastic regression

Tian Xia[a,*], Xuejun Jiang[b], Xueren Wang[c]

[a] School of Mathematics and Statistics, Guizhou University of Finance and Economics, Guiyang, 550025, China
[b] Department of Financial Mathematics and Financial Engineering, South University of Science and Technology of China, Shenzhen 518055, China
[c] Department of Statistics, Yunnan University, Kunming 650091, China

A R T I C L E I N F O

Article history:
Received 1 November 2014
Received in revised form 11 April 2015
Accepted 12 April 2015
Available online 21 April 2015

MSC:
62 F 12
62 J 02
62 J 12

Keywords:
Quasi-likelihood nonlinear models with stochastic regression
Strong consistency
Convergence rate
Maximum quasi-likelihood estimator

A B S T R A C T

This paper proposes some mild regularity conditions analogous to those given by Wu (1981) and Chang (1999). On the basis of the proposed regularity conditions, the strong consistency as well as convergence rate for maximum quasi-likelihood estimator (MQLE) is obtained in quasi-likelihood nonlinear models (QLNMs) with stochastic regression.

1. Introduction

Since the generalized linear models (GLMs) were introduced by Nelder and Wedderburn (1972), statisticians have been trying to extend its scope of application. An important extension proposed by Wedderburn (1974) is the quasi-likelihood function, which requires the correct specification of a relationship between the mean and variance function only, rather than the entire distribution of the response variable. The quasi-likelihood approach is useful because in many situations the exact distribution of the observations is unknown. Moreover, a quasi-likelihood function has statistical properties similar to those of a log-likelihood function. For more comprehensive explanation about GLMs and quasi-likelihood functions, we refer to McCullagh and Nelder (1989).

Suppose that $(\boldsymbol{x}_i, y_i),\ i = 1, 2, \ldots, n$ are n pairs of design vectors and responses. In many practical applications, the design variables at a particular stage are chosen according to the information from previous observations. In other words, the choice of design vector $\boldsymbol{x}_n$ at the stage n depends on the previous observations of $\boldsymbol{x}_1, y_1, \ldots, \boldsymbol{x}_{n-1}, y_{n-1}$. Formally, let

* Corresponding author.
E-mail address: xiatian718@hotmail.com (T. Xia).

http://dx.doi.org/10.1016/j.spl.2015.04.016
0167-7152/

$\mathcal{F}_n = \sigma(\boldsymbol{x}_1, y_1, \ldots, \boldsymbol{x}_n, y_n)$ for $n \in N$, where $\sigma(\boldsymbol{x}_1, y_1, \ldots, \boldsymbol{x}_n, y_n)$ denotes a σ-algebra generated by $(\boldsymbol{x}_1, y_1, \ldots, \boldsymbol{x}_n, y_n)$, then the design vector at n-stage, $\boldsymbol{x}_n$, is $\mathcal{F}_{n-1}$-measurable. For each i, assume that $(\boldsymbol{x}_i, y_i)$ satisfies

$$E(y_i|\mathcal{F}_{i-1}) = \mu_i = h(\boldsymbol{x}_i, \boldsymbol{\beta}), \tag{1.1}$$

$$\mathrm{Var}(y_i|\mathcal{F}_{i-1}) = \sigma_i^2 = \sigma^2 v(\mu_i), \tag{1.2}$$

where $\boldsymbol{x}_i \in R^q$, $y_i \in R$, $h(\cdot,\cdot)$ is a known differentiable function, $\boldsymbol{\beta}$ is an unknown parameter vector to be estimated, σ^2 is a scale parameter which is usually treated as a nuisance parameter, and $v(\cdot)$ is called a variance function, which may or may not be known.

If the variance function $v(\cdot)$ is known, following McCullagh and Nelder (1989), Wedderburn (1974), Xia and Kong (2008), and Xia et al. (2008, 2010), the log quasi-likelihood is defined as

$$\sum_{i=1}^{n} \int_{y_i}^{\mu_i} \frac{y_i - t}{\sigma^2 v(t)} dt, \quad \mu_i = h(\boldsymbol{x}_i, \boldsymbol{\beta}) \triangleq \mu_i(\boldsymbol{\beta}). \tag{1.3}$$

When $v(\cdot)$ is unknown, we replace $v(\cdot)$ in (1.3) by a suitably-chosen known function $\Lambda(\cdot)$ which is called a "working" variance function (see, for example, Liang and Zeger, 1986, Zeger and Liang, 1986, Fahrmeir, 1990, Xia et al., 2014 and references therein), thus the log quasi-likelihood is changed into

$$Q(\boldsymbol{\beta}; Y) = \sum_{i=1}^{n} \int_{y_i}^{\mu_i} \frac{y_i - t}{\sigma^2 \Lambda(t)} dt, \quad \mu_i = h(\boldsymbol{x}_i, \boldsymbol{\beta}) \triangleq \mu_i(\boldsymbol{\beta}). \tag{1.4}$$

Then models defined by (1.1) and (1.4) are called quasi-likelihood nonlinear models (QLNMs) with stochastic regression. It is easily seen from (1.4) that quasi-likelihood equation is

$$\sum_{i=1}^{n} \frac{\partial \mu_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} (\Lambda(\mu_i(\boldsymbol{\beta})))^{-1} (y_i - \mu_i(\boldsymbol{\beta})) = 0, \quad \mu_i(\boldsymbol{\beta}) = h(\boldsymbol{x}_i, \boldsymbol{\beta}). \tag{1.5}$$

Clearly, QLNMs with stochastic regression encompasses some important special cases. For example, if $\mu_i(\boldsymbol{\beta}) = \boldsymbol{x}_i^T\boldsymbol{\beta}$ and $\Lambda = 1$, it reduces to linear stochastic regression models (see Nelson, 1980, Lai and Wei, 1982, and Wei, 1985); if $\mu_i(\boldsymbol{\beta}) = h(\boldsymbol{x}_i^T\boldsymbol{\beta})$, $\Lambda(\cdot) = v(\cdot)$, and y_i is drawn independently from a one-parameter exponential family of distributions, with density function

$$\exp\{\theta_i y_i - b(\theta_i)\} d\gamma(y_i), \quad i = 1, \ldots, n,$$

where $b(.)$ is a known function, $\gamma(.)$ is a measure, then (1.5) can be rewritten as

$$\sum_{i=1}^{n} \boldsymbol{x}_i \left. \frac{\partial h(\eta)}{\partial \eta} \right|_{\eta = \boldsymbol{x}_i^T\boldsymbol{\beta}} (b''(\theta_i))^{-1} (y_i - \mu_i(\boldsymbol{\beta})) = 0, \quad \mu_i(\boldsymbol{\beta}) = h(\boldsymbol{x}_i^T\boldsymbol{\beta}). \tag{1.6}$$

Eq. (1.6) is just the well-known likelihood equation of generalized linear models with adaptive designs (see Chang, 1999, 2001). Therefore, QLNMs with stochastic regression is an extension of the linear stochastic regression models and the generalized linear models with adaptive designs. The root of Eq. (1.5), denoted by $\hat{\boldsymbol{\beta}}_n$, is called the maximum quasi-likelihood estimator (MQLE) of $\boldsymbol{\beta}_0$ (here and in the sequel $\boldsymbol{\beta}_0$ denotes the true value of parameter $\boldsymbol{\beta}$).

In the past three decades, a number of authors have been concerned about the strong consistency of estimator of a regression parameter in stochastic regression models. For example, write $e_i = y_i - E(y_i|\mathcal{F}_{i-1}) = y_i - h(\boldsymbol{x}_i, \boldsymbol{\beta}_0)$, when the system is linear, Lai and Wei (1982) obtained the strong consistency and convergence rate of least-squares estimate in stochastic regression models under $\sup_{i\ge 1} E(|e_i|^\alpha|\mathcal{F}_{i-1}) < \infty$ a.s. for some $\alpha > 2$. Ding and Chen (2006) proved the strong consistency of the maximum likelihood estimate in generalized linear models with stochastic regressors under some mild conditions. For generalized linear models with canonical link functions, and under similar conditions of Lai and Wei (1982), Chen et al. (1999) showed the strong consistency of the MQLE for both adaptive and fixed design cases. But, none of the results in the literature genuinely covers the strong consistency of the maximum quasi-likelihood estimator in QLNMs with stochastic regression in full generality. In this paper, we study the strong consistency of MQLE in QLNMs with stochastic regression.

This paper is organized as follows. Section 2 introduces some regularity conditions and lemmas. In Section 3, based on Lemma A of Chen et al. (1999) and the regularity conditions given in Section 2, the strong consistency as well as the convergence rate for MQLE is derived in QLNMs with stochastic regression, subsequently, an example is provided to illustrate that the proposed conditions are reasonable. The conclusion of the paper is given in Section 4.

2. Conditions and lemmas

Before formulating the assumptions, we introduce some notation. Let $\lambda_{\min}(A)(\lambda_{\max}(A))$ denote the smallest (largest) eigenvalue of a symmetric matrix A; let $\|B\| = (\sum_{i=1}^{p}\sum_{j=1}^{q}|b_{ij}|^2)^{1/2} = (\mathrm{tr}(B^TB))^{1/2}$ for any matrix $B = (b_{ij}) \in R^{p\times q}$;