Statistica Sinica **13**(2003), 69-81

AN EXTENDED EMPIRICAL LIKELIHOOD FOR GENERALIZED LINEAR MODELS

Song Xi Chen and Hengjian Cui

National University of Singapore and Beijing Normal University

Abstract: The paper considers improving the efficiency of parameter estimation of the quasi-likelihood in generalized linear models. The improvement is offered by employing the empirical likelihood and incorporating extra constraints which better utilize the provided variance structure of the models. We recommend a particular choice for the extra constraints to reduce the variance of the quasi-likelihood based parameter estimators.

Key words and phrases: Empirical likelihood, exponential family, extra constraints, generalized linear models, quasi-likelihood.

1. Introduction

Let a scalar random variable Y be the response of a random vector $X \in R^p$ such that

$$E(Y|X) = G(X^\tau \beta) \quad \text{and} \quad \mathrm{Var}\,(Y|X) = \sigma^2 V\{G(X^\tau \beta)\}, \tag{1}$$

where $\beta \in R^p$ is a vector of real parameters, G is a known smooth link function and V is a known variance function. This is the framework of generalized linear models under which the quasi-likelihood (Wedderburn (1974)) has been a popular tool for semiparametric inference, see McCullagh and Nelder (1989). Let $\mu(\beta) = G(X^\tau \beta)$. The log quasi-likelihood ratio of β is defined as

$$Q\{y; \mu(\beta)\} = \int_y^{\mu(\beta)} \frac{y-u}{V(u)} du.$$

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be an independent and identically distributed sample from (1), and $\mu_i(\beta) = G(X_i^\tau \beta)$. The joint quasi-likelihood ratio of the data is

$$\sum_{i=1}^n Q\{Y_i, \mu_i(\beta)\}. \tag{2}$$

The maximum quasi-likelihood estimator of β is the root of

$$\sum_{i=1}^n \frac{\{Y_i - \mu_i(\beta)\} G'(X_i^\tau \beta) X_i}{V\{\mu_i(\beta)\}} = 0, \tag{3}$$

which coincides with the optimal estimating function of Godambe (1960, 1976). Wedderburn (1974) and McCullagh (1983) have shown that the quasi-likelihood possesses some properties of a real likelihood. The maximum quasi-likelihood estimator is optimal (smallest asymptotic variance) within the family of all the linear estimating functions $\sum w_i\{Y_i - \mu_i(\beta)\} = 0$. Despite its good properties, the quasi-likelihood only fully utilizes information on the mean structure of the model assisted by the variance structure. Information on the variance has not been fully utilized when the conditional distribution of Y given X is outside the exponential family. Firth (1987) studied the efficiency of quasi-likelihood relative to the maximum likelihood estimator outside the exponential family and reported low efficiency in some cases. As a way to improve the efficiency, Firth proposed a "refinement" based on knowledge of the third moment. Along the same line Godambe and Thompson (1989) proposed to extend the estimating functions based on knowledge of the third and fourth moments of the conditional distribution of Y given X. To bring more information on the variance structure into the modeling process, Nelder and Pregibon (1987) added an extra term $\log\{V(y)\}$ in the quasi-score, while Smyth (1989) considered a sub-model for the dispersion parameter σ^2.

The approach taken in this paper does not assume knowledge of higher moments, nor a sub-model. To fully utilize information contained in V, we consider using the empirical likelihood due to its ability to conduct a nonparametric inference without knowledge of higher order moments of the distribution while implicitly taking them into consideration. Empirical likelihood is a computer intensive statistical method introduced by Owen (1988, 1990) as an alternative to the bootstrap. Instead of resampling with an equal probability weight for all data values like the bootstrap, the empirical likelihood chooses the weights by profiling a multinomial likelihood under a set of constraints. The constraints reflect the characteristics of the quantity of interest. An updated comprehensive overview of the empirical likelihood is available in Owen (2001). Empirical likelihood has been shown in a wide range of situations to have properties analogous to a real likelihood, see Hall and La Scala (1990), Chen (1994, 1996), Qin and Lawless (1994, 1995), Li (1995), Jing (1995) and others. Standard empirical likelihood for generalized linear models has been considered by Kolaczyk (1994), based on constraints derived from the score function of the quasi-likelihood. A general framework of empirical likelihood based on estimating function is discussed in Qin and Lawless (1994).

In this paper we propose an extended empirical likelihood for generalized linear models that incorporates extra constraints which explore the provided variance structure. What we propose is different from the approach taken to find the optimal estimating functions which satisfies $E\{Y_i - \mu_i(\beta)\} = 0$ and

$E[\{Y_i - \mu_i(\beta)\}^2 - \sigma^2 V(\mu_i)] = 0$. Following Godambe and Thompson (1989), the latter approach, which abandons the original estimating functions based on the quasi-score, leads to $p+1$ estimating functions which depend on the third and fourth order moments. Our approach is to keep the original estimating function based on the quasi-score and to add extra constraints from the variance structure. The total number of constraints will be larger than $p+1$. Adding these extra constraints is to (i) have a better usage of the provided information on the variance, and (ii) produce more efficient estimators for β and σ^2. A choice of $p+1$ extra constraints, which depends only on G, V and X, is recommended. It is shown that the extended empirical likelihood is more efficient than the quasi-likelihood when the conditional distribution of Y given X is outside the exponential family. When the distribution is within the exponential family, the empirical likelihood gives estimators which are as efficient as the quasi-likelihood estimators. So, the extended empirical likelihood acts like an insurance policy which provides protection against distributions outside the exponential family.

The paper is structured as follows. We review the empirical likelihood for generalized linear models in Section 2. In Section 3, we introduce the extended empirical likelihood based on extra constraints on the variance and obtain conditions under which the inference based on the empirical likelihood is more efficient than that based on the quasi-likelihood. The issue of how to select the extra constraints is discussed in Section 4. Results from a simulation study are reported in Section 5. Proofs are given in the appendix.

2. Empirical Likelihood in Generalised Linear Models

For $1 \le i \le n$, define $Z_i^{(1)}(\beta) = \{Y_i - G(X_i^\tau \beta)\}G'(X_i^\tau \beta)X_i/V\{G(X_i^\tau \beta)\}$ and $Z_i^{(2)}(\beta,\sigma^2) = \{Y_i - G(X_i^\tau \beta)\}^2/\sigma^4 V\{G(X_i^\tau \beta)\} - 1/\sigma^2$. Clearly $Z_i^{(1)}$ is the quasi-score associated with the i-th observation, and $Z_i^{(2)}$ describes the variance structure of the model. As $E\{Z_i^{(j)}(\beta)\} = 0$ for $j = 1$ and 2, an empirical likelihood for (β,σ^2) is $L(\beta,\sigma^2) = \max \prod_{i=1}^n p_i$ subject to constraints

$$\sum p_i Z_i^{(1)}(\beta) = 0 \quad \text{and} \quad \sum p_i Z_i^{(2)}(\beta,\sigma^2) = 0, \tag{4}$$

where $p_1, \ldots, p_n$ are nonnegative real numbers summing to unity.

As the number of constraints equals the number of parameters, it may be shown that the optimal $p_i = n^{-1}$, i.e., Owen (1988). Thus, the maximum empirical likelihood estimators for (β,σ^2) are the solutions of

$$\sum \frac{\{Y_i - G(X_i^\tau \beta)\}G'(X_i^\tau \beta)X_i}{V\{G(X_i^\tau \beta)\}} = 0, \tag{5}$$

$$\sum \frac{\{Y_i - G(X_i^\tau \beta)\}^2}{\sigma^4 V\{G(X_i^\tau \beta)\}} - \frac{1}{\sigma^2} = 0. \tag{6}$$

This implies that the maximum empirical likelihood estimator for β is the same as that of quasi-likelihood, and the estimator for σ^2 is of the conventional form

$$\tilde{\sigma}^2 = \sum \frac{\{Y_i - G(X_i^\tau \hat{\beta})\}^2}{V\{G(X_i^\tau \hat{\beta})\}},$$

where $\hat{\beta}$ is the quasi-likelihood estimator of β. The advantage of the empirical likelihood in this case, as pointed out by Kolaczyk (1994), is in its construction of confidence regions of natural shape and orientation, rather than in parameter estimation.

The second constraint in (4) does not fully use the information on V. When $V' =: V'(G[X^\tau \beta]) \neq 0$, the change of V as reflected in V' may contain useful information on the variance structure. To utilize this information, a general form of extra constraints is given by

$$\sum p_i \Big(\frac{[Y_i - G(X_i^\tau \beta)]^2}{\sigma^4 V\{G(X_i^\tau \beta)\}} - \frac{1}{\sigma^2} \Big) w(X_i^\tau \beta, X_i) = 0, \tag{7}$$

where w is a r-dimensional weight function and $1 \leq r \leq p$. We will show later that adding these extra constraints leads to variance reduction in parameter estimation.

Let $\epsilon = Y - G(X^\tau \beta)$, $V = V[G(X^\tau \beta)]$, $w = w(X^\tau \beta, X)$, $V' = V'(G[X^\tau \beta])$, $Z_i^{(3)}(\beta, \sigma^2) = ([Y_i - G(X_i^\tau \beta)]^2 / \sigma^4 V\{G(X_i^\tau \beta)\} - 1/\sigma^2) w(X_i^\tau \beta, X_i)$ and $Z_i(\beta, \sigma^2) = (Z_i^{(1)\tau}(\beta), Z_i^{(2)\tau}(\beta, \sigma^2), Z_i^{(3)\tau}(\beta, \sigma^2))^\tau$.

We assume the following conditions.

C1: $G(\cdot)$ is twice continuously differentiable and $V(\cdot)$ is continuously differentiable;

C2: $E\{Z_1(\beta, \sigma^2) Z_1^\tau(\beta, \sigma^2)\}$ is non-singular;

C3: for some $\delta > 0$, $E\{|\epsilon|^2 + \|X\|\}^{2+\delta} < \infty$, $E\{|G'(X^\tau \beta)| + V^{-1} + w\}^{2+\delta} < \infty$ and $E\{|G''(X^\tau \beta)| + |V'|\}^{1+\delta} < \infty$;

C4: The matrix $\Big(E[\partial Z_1(\beta, \sigma^2)/\partial \beta], E[\partial Z_1(\beta, \sigma^2)/\partial \sigma^2]\Big)$ has full rank.

These conditions are standard in the theory of quasi-likelihood.

3. An Extended Empirical Likelihood

The extended empirical likelihood for β and σ^2 is $L(\beta, \sigma^2) = \max \prod_{i=1}^n p_i$, where $p_1, \ldots, p_n$ are nonnegative real numbers summing to unity and are subject to structure constraints

$$\sum p_i Z_i(\beta, \sigma^2) = 0. \tag{8}$$

There are $q = p + 1 + r$ constraints for $p + 1$ parameters. As the total number of the constraints is larger than the number of parameters, the maximum empirical likelihood estimator for β is different from that based on the quasi-likelihood. This is an empirical likelihood within the framework of estimating functions considered in Qin and Lawless (1994). While Qin and Lawless showed that the variance of the maximum empirical likelihood estimators does not increase with extra constraints added, we are interested here in the construction of extra constraints to achieve best possible reduction in the variance.

Let $\ell(\beta,\sigma^2) = -2\log\{n^n L(\beta,\sigma^2)\}$ be the log empirical likelihood ratio. Then, according to a standard procedure of optimization using the Lagrange multipliers (Owen (1988)), $\ell(\beta,\sigma^2) = 2\sum_{i=1}^n \log\{1+\lambda^\tau Z_i(\beta,\sigma^2)\}$, where $\lambda \in R^q$ satisfies

$$Q_{1n}(\beta,\sigma^2,\lambda) := \sum_{i=1}^n \frac{Z_i(\beta,\sigma^2)}{1+\lambda^\tau Z_i(\beta,\sigma^2)} = 0. \tag{9}$$

Differentiating $\ell(\beta,\sigma^2)$ with respect to (β,σ^2), we have from (9) that

$$\begin{aligned} \frac{\partial \ell(\beta,\sigma^2)}{\partial\beta} &= \lambda^\tau \sum_{i=1}^n \frac{\partial Z_i(\beta,\sigma^2)/\partial\beta}{1+\lambda^\tau Z_i(\beta,\sigma^2)} := Q_{2n}(\beta,\sigma^2,\lambda), \\ \frac{\partial l(\beta,\sigma^2)}{\partial\sigma^2} &= \lambda^\tau \sum_{i=1}^n \frac{\partial Z_i(\beta,\sigma^2)/\partial\sigma^2}{1+\lambda^\tau Z_i(\beta,\sigma^2)} := Q_{3n}(\beta,\sigma^2,\lambda). \end{aligned} \tag{10}$$

Let $A = E[Z_1(\beta,\sigma^2)Z_1^\tau(\beta,\sigma^2)]$ and $B = \Big(E\Big(\frac{\partial Z_1(\beta,\sigma^2)}{\partial\beta}\Big), E\Big(\frac{\partial Z_1(\beta,\sigma^2)}{\partial\sigma^2}\Big)\Big)$. According to Qin and Lawless (1994), under the conditions C1$-$C4, the empirical likelihood ratio $\ell(\beta,\sigma^2)$ attains its minimum at $(\hat\beta,\hat\sigma^2)$ and $\hat\lambda = \lambda(\hat\beta,\hat\sigma^2)$ such that $Q_{1n}(\hat\beta,\hat\sigma^2,\hat\lambda) = 0$, $Q_{2n}(\hat\beta,\hat\sigma^2,\hat\lambda) = 0$, $Q_{3n}(\hat\beta,\hat\sigma^2,\hat\lambda) = 0$; and

$$\begin{pmatrix} \sqrt{n}(\hat\beta-\beta) \\ \sqrt{n}(\hat\sigma^2-\sigma^2) \end{pmatrix} \xrightarrow{d} N(0,\Sigma),$$

where $\Sigma = (B^\tau A^{-1} B)^{-1}$. So, the asymptotic variances of the empirical likelihood estimators $\sqrt{n}\hat\beta$ and $\sqrt{n}\hat\sigma^2$ are respectively $\Sigma_{\hat\beta} = (I_p,0)\Sigma(I_p,0)^\tau$ and $\Sigma_{\hat\sigma^2} = (0,1)\Sigma(0,1)^\tau$. The corresponding asymptotic variance of the quasi-likelihood estimator for β, denoted by $\tilde\beta_{ql}$, is $\Sigma_{\tilde\beta_{ql}} = \sigma^2\{E[G'(X^\tau\beta)^2 XX^\tau/V]\}^{-1}$. A standard estimator for σ^2 is $\tilde\sigma^2 = n^{-1}\sum\{Y_i - G(X_i^\tau\tilde\beta_{ql})\}^2/V\{G(X_i^\tau\tilde\beta_{ql})\}$ whose asymptotic variance is denoted by $\Sigma_{\tilde\sigma^2}$.

Put $\mu_3(X) = E[(\epsilon/(\sigma\sqrt{V}))^3 \mid X]$ and $\mu_4(X) = E[(\epsilon/(\sigma\sqrt{V}))^4 \mid X]$. To simplify notation, we write them as μ_3 and μ_4. Define

$$\begin{aligned} A_{11} &= \sigma^4 \Sigma_{\tilde\beta_{ql}}^{-1} = \sigma^2 E\Big[\frac{G'(X^\tau\beta)^2}{V} XX^\tau\Big], \\ A_{22} &= \begin{pmatrix} E\Big[\frac{\mu_4-1}{\sigma^4}\Big] & E\Big[\frac{\mu_4-1}{\sigma^4} w^\tau\Big] \\ E\Big[\frac{\mu_4-1}{\sigma^4} w\Big] & E\Big[\frac{\mu_4-1}{\sigma^4} ww^\tau\Big] \end{pmatrix}, \end{aligned}$$

$$A_{12} = \left(E\Big[\frac{\mu_3 G'(X^\tau\beta)}{\sigma\sqrt{V}} X\Big], E\Big[\frac{\mu_3 G'(X^\tau\beta)}{\sigma\sqrt{V}} X w^\tau\Big] \right),$$

$$B_1 = \left(E\Big[\frac{V'G'(X^\tau\beta)}{\sigma^2 V} X\Big], E\Big[\frac{V'G'(X^\tau\beta)}{\sigma^2 V} X w^\tau\Big] \right) \quad \text{and} \quad B_2 = \left(\frac{1}{\sigma^4}, E\Big[\frac{w^\tau}{\sigma^4}\Big] \right).$$

Then,

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{12}^\tau & A_{22} \end{pmatrix} \quad \text{and} \quad B^\tau = -\begin{pmatrix} \sigma^{-2}A_{11} & B_1 \\ 0 & B_2 \end{pmatrix}.$$

For a $d \times d$ non-negative definite symmetric matrix M we write $M \geq_l 0$ if M has exactly $l (\leq d)$ positive eigenvalues. Moreover, define $\Delta = (A_{12}^\tau - B_1^\tau, -B_2^\tau)^\tau$.

Theorem 1. (i) *Under conditions* C1$-$C4, $\Sigma_{\tilde{\beta}_{ql}} - \Sigma_{\hat{\beta}} \geq_{l-1} 0$ *where* $l = rank(\Delta)$. (ii) *The asymptotic variance of* $\hat{\sigma}^2$ *is less than the asymptotic variance of* $\tilde{\sigma}^2$ *if and only if* $l > 1$.

Remark 1. Given a choice of weight w, the rank l depends on $\mu_3 = \mu_3(X)$ but is free of the fourth moment $\mu_4(X)$. If the rank $l = p+1$, then there will be reduction in variance of $\tilde{\beta}_{ql}$ in all directions.

Remark 2. If $l = 1$ (note that $l \geq 1$), there will be no reduction in the variance, which is also implied by Corollary 3 of Qin and Lawless (1994). This happens when the conditional distribution of Y given X is within the exponential family. This means that the extended empirical likelihood does not offer any improvement over (but has the same efficiency as) the quasi-likelihood. This is not surprising as, within the exponential family, the quasi-likelihood estimator is optimal and the information on the variance has been fully represented by the mean structure. The extended empirical likelihood function acts like an insurance policy that provides protection against departure from the exponential family.

4. Choosing the Weight

We discuss the choice of the weight function w in order to achieve variance reduction. Ideally we are tempted to find a w such that

$$\Sigma_{\tilde{\beta}_{ql}} - \Sigma_{\hat{\beta}(w)} \text{ is maximized among all choices of } w, \tag{11}$$

where $\hat{\beta}(w)$ is the extended empirical likelihood estimator of β based on the weight w. (Here, for two non-negative definite symmetric matrices T and S of the same order, T is said to be larger than S (written $T > S$) if $T - S \geq_l 0$ for some positive integer l). However, an optimal w free of unknown quantities is difficult, if not impossible, to obtain. The following theorem suggests a particular weight function.

Theorem 2. *If* $E(\epsilon^3|X)=0$, $E(\epsilon^4|X)=\kappa\sigma^4V^2$ *for some* $\kappa>1$ *and* $\mathrm{Cov}(V'GX/V)>0$, *then the optimal weight function under* (11) *is*

$$w^\star(X^\tau\beta, X) = V'\{G(X^\tau\beta)\}G'(X^\tau\beta)X/V\{G(X^\tau\beta)\}. \tag{12}$$

Note that under the conditions of Theorem 2, $\Delta = (A_{12}^\tau - B_1^\tau,\ -B_2^\tau)^\tau$ has full rank $p+1$ if $w^\star$ is used as the weight function.

Now we consider the case when $\mu_3(X) \neq 0$. Let $\gamma =: \Big(\mu_3(X) - \frac{V'}{\sigma\sqrt{V}}\Big)\frac{G'X}{\sigma\sqrt{V}}$ and $P = I_{r+1} - A_{22.1}^{-1/2}B_2^\tau(B_2A_{22.1}^{-1}B_2^\tau)^{-1}B_2A_{22.1}^{-1/2}$ be the projection matrix onto the column space of $A_{22.1}^{-1/2}B_2^\tau$. It may be shown that

$$\Sigma_{\hat\beta(w)} = \Big(\sigma^{-4}A_{11} + (A_{12}-B_1)A_{22.1}^{-1/2}PA_{22.1}^{-1/2}(A_{12}-B_1)^\tau\Big)^{-1}.$$

According to (11), we want to find a w such that $(A_{12}-B_1)A_{22.1}^{-1/2}PA_{22.1}^{-1/2}(A_{12}-B_1)^\tau$ is as large as possible. As $A_{12}-B_1 = \Big(E(\gamma), E(\gamma w^\tau)\Big)$ and $PA_{22.1}^{-1/2}B_2^\tau = 0$,

$$\begin{aligned}
&(A_{12}-B_1)A_{22.1}^{-1/2}PA_{22.1}^{-1/2}(A_{12}-B_1)^\tau\\
&=[(E(\gamma), E(\gamma w^\tau)) - \sigma^4E(\gamma)B_2]A_{22.1}^{-1/2}PA_{22.1}^{-1/2}[(E(\gamma), E(\gamma w^\tau))^\tau - \sigma^4B_2^\tau E(\gamma)^\tau]\\
&=[0, E(\gamma w^\tau) - E(\gamma)E(w^\tau)]A_{22.1}^{-1/2}PA_{22.1}^{-1/2}[0, E(\gamma w^\tau) - E(\gamma)E(w^\tau)]^\tau\\
&=\mathrm{Cov}\,(\gamma, w)\Gamma(w)\,\mathrm{Cov}\,(\gamma, w)^\tau,
\end{aligned}$$

where $\Gamma(w) = (0, I_p)A_{22.1}^{-1/2}PA_{22.1}^{-1/2}(0, I_p)^\tau > 0$. The task of finding w to maximize $\mathrm{Cov}\,(\gamma, w)\Gamma(w)\,\mathrm{Cov}\,(\gamma, w)^\tau$ is not trivial due to the complexity of $\Gamma(w)$. If some additional assumptions on $\mu_3(X)$ and $\mu_4(X)$ are given, it may be possible to find the optimal w within a confined family of distributions.

When $\mu_3(X)$ and $\mu_4(X)$ are unknown, the choice of $w^\star$ is supported by the following heuristic arguments. Let Z_i be a random variable that has the same mean and variance structure as $\{Y_i - G(X_i^\tau\beta)\}^2$, that is $E(Z_i|X_i) = \sigma^2V\{G(X_i^\tau\beta)\}$ and $\mathrm{Var}\,(Z_i|X_i) = \{\mu_4(X_i)-1\}\sigma^4V^2\{G(X_i^\tau\beta)\}$. The optimal estimating function within the family of $\sum[Z_i - \sigma^2V\{G(X_i^\tau\beta)\}]w_i = 0$ is

$$\sum\frac{\{Y_i - G(X_i^\tau\beta)\}^2 - \sigma^2V\{G(X_i^\tau\beta)\}}{(\mu_4-1)\sigma^2V^2\{G(X_i^\tau\beta)\}}V'\{G(X_i^\tau\beta)\}G'(X_i^\tau\beta)X_i = 0,$$

which is equivalent to

$$\sum\Big[\frac{\{Y_i - G(X_i^\tau\beta)\}^2}{V\{G(X_i^\tau\beta)\}} - \sigma^4\Big]\frac{V'\{G(X_i^\tau\beta)\}G'(X_i^\tau\beta)X_i}{\{\mu_4(X)-1\}V\{G(X_i^\tau\beta)\}} = 0.$$

This indicates $w^\star/\{\mu_4(X)-1\}$ is the optimal weight and would be close to $w^\star$ if $\mu_4(X)$ is close to a constant. So, the extended empirical likelihood is based on $2p+1$ estimating functions with the first $p+1$ being the standard quasi-likelihood estimating functions for the mean, and the last p being "close" to the quasi-likelihood estimation functions for the variance.

When $w^\star$ is used as the weight function,

$$\Delta = \begin{pmatrix} A_{12} - B_1 \\ -B_2 \end{pmatrix}$$
$$= \begin{pmatrix} E\Big\{\Big(\mu_3(X) - \dfrac{V'}{\sigma\sqrt{V}}\Big)\dfrac{G'(X^\tau\beta)X}{\sqrt{V}}\Big\} & E\Big\{\Big(\mu_3(X) - \dfrac{V'}{\sigma\sqrt{V}}\Big)\dfrac{G'(X^\tau\beta)X}{\sqrt{V}}w^{\star\tau}\Big\} \\ \dfrac{1}{\sigma^4} & E\Big(\dfrac{w^{\star\tau}}{\sigma^4}\Big) \end{pmatrix}.$$

This matrix has rank $p+1$ if and only if $E[\{\mu_3(X) - V'/(\sigma\sqrt{V})\}G'(X^\tau\beta)\times Xw^{\star\tau}/\sqrt{V}]$ has rank p. When this happens, there will be reduction in variance in all directions, according to Theorem 1.

5. Simulation Results

In this section we report some simulation results to show the performance of the extended empirical likelihood estimators. The following generalized linear models were considered in the simulation: $Y_i = G(X_i^\tau\beta) + \sigma V^{1/2}\{G(X_i^\tau\beta)\}\epsilon_i$, where $G(t) = \exp(t)$, $V(t) = t^2$, $X_i = (X_{i1}, X_{i_2})^\tau$ were independent and identically uniformly distributed in $[0,2]\times[0,2]$ and ϵ_i were i.i.d. random variables independent of X_i, with zero expectation. We chose two distributions for ϵ_i: the standard normal $N(0,1)$ and the uniform$[-1,1]$. The parameter values were $\beta_1 = 0.1$, $\beta_2 = 0.2$ and $\sigma = 0.5$.

The recommended weight $w^\star$ given in (12) was used to set up the extended empirical likelihood for (β, σ^2). The conjugate gradient method was used to search for the optimal empirical likelihood parameter estimates by modifying a routine given in Press, Teukolsky, Vetterling and Flannery (1992). The sample size used in the simulation ranged from 60 to 200.

The simulation results are summarized in Figure 1 for normal errors, and in Figure 2 for uniform errors, based on 500 simulations. We observe from these figures that the variances of the extended empirical likelihood estimates for β_1 and β_2 were substantially smaller than their quasi-likelihood counterparts for both types of error distributions. The results on the mean square errors of the estimates show that the reduction in variance by the extended empirical likelihood came at no cost in bias. At the same time, the simulation shows that the extended empirical likelihood offered only small improvement on the estimation of σ^2. This is not unexpected within our theory.

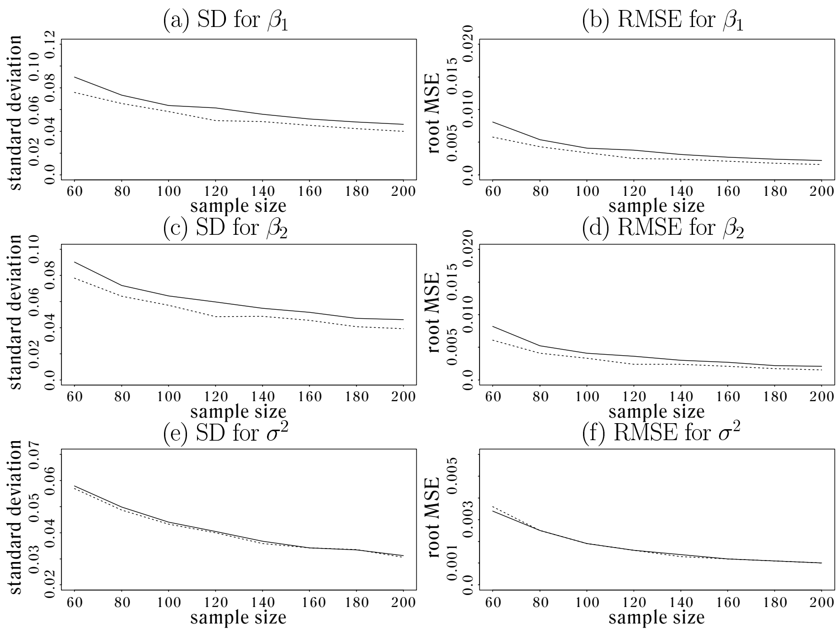

Figure 1. Standard deviations and root mean square errors of quasi-likelihood (in solid lines) and extended empirical likelihood likelihood (in dashed lines) estimates for $(\beta_1, \beta_2, \sigma^2)$, with normally distributed errors.

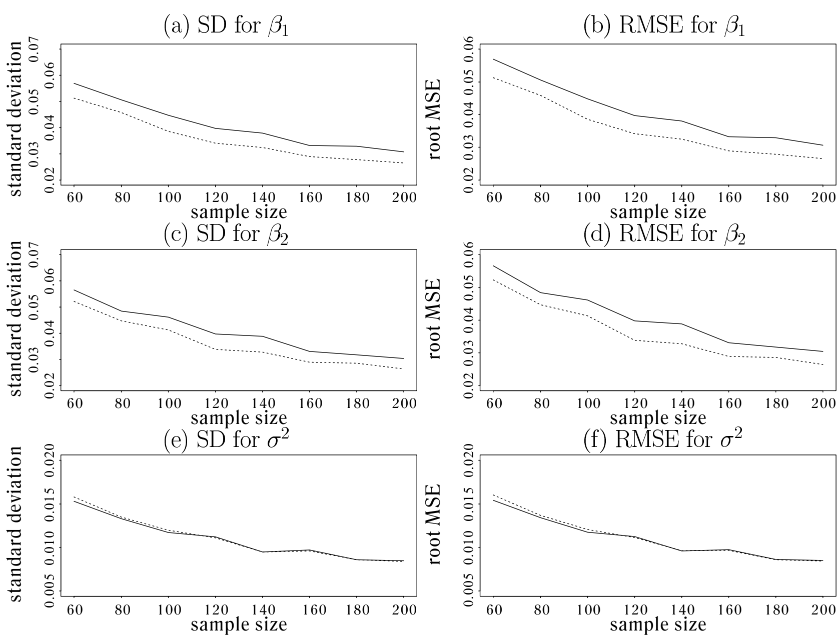

Figure 2. Standard deviations and root mean square errors of quasi-likelihood (in solid lines) and extended empirical likelihood likelihood (in dashed lines) estimates for $(\beta_1, \beta_2, \sigma^2)$, with uniformly distributed errors.

Acknowledgements

We thanks the Co-Editor, an associate editor and two referees for helpful comments that improved the presentation of the paper. The research was supported by a National University of Singapore Academic Research grant (R-150-000-018-112) and a grant from the NSFC and EYTP of China.

Appendix A. Proof of Theorem 1

(i) Let $A_{22.1} = A_{22} - A_{21}A_{11}^{-1}A_{12}$, non-singular when A is which is non-singular. Then,

$$\Sigma^{-1} = B^\tau A^{-1} B = \begin{pmatrix} \sigma^{-4}A_{11} & 0 \\ 0 & 0 \end{pmatrix} + B^\tau \begin{pmatrix} A_{11}^{-1}A_{12} \\ -I_{r+1} \end{pmatrix} A_{22.1}^{-1}\Big(A_{21}A_{11}^{-1}, -I_{r+1}\Big)B$$
$$= \begin{pmatrix} \sigma^{-4}A_{11} & 0 \\ 0 & 0 \end{pmatrix} + D, \tag{13}$$

where $D = \Delta A_{22.1}^{-1}\Delta^\tau$ and $\Delta = \begin{pmatrix} A_{12} - B_1 \\ -B_2 \end{pmatrix}$. Clearly $\text{rank}(D) = \text{rank}(\Delta) = l$. Let $D_1 = I_p - (I_p, 0)\{\begin{pmatrix} I_p & 0 \\ 0 & 0 \end{pmatrix} + C\}^{-1}(I_p, 0)^\tau$ and

$$C = \sigma^4 \begin{pmatrix} A_{11}^{-1/2} & 0 \\ 0 & 1 \end{pmatrix} D_1 \begin{pmatrix} A_{11}^{-1/2} & 0 \\ 0 & 1 \end{pmatrix}.$$

Then, $\Sigma_{\tilde\beta_{ql}} - \Sigma_{\hat\beta} = \sigma^4 A_{11}^{-1} - (I_p, 0)\Big(B^\tau A^{-1}B\Big)^{-1}(I_p, 0)^\tau = \sigma^4 A_{11}^{-1/2} D_1^{-1} A_{11}^{-1/2}$. If $l = 1$, C can be written as $C = (b_1^\tau, b_2)^\tau(b_1^\tau, b_2)$, where $b_1 \in R^{p+1}, b_2 = \sqrt{d} > 0$ for some $d > 0$. Then

$$D_1 = I_p - (I_p, 0)\begin{pmatrix} I_p + b_1b_1^\tau & b_2b_1 \\ b_2b_1^\tau & b_2^2 \end{pmatrix}^{-1}(I_p, 0)^\tau = I_p - (I_p + b_1b_1^\tau - b_2b_1b_2^{-2}b_2b_1^\tau)^{-1} = 0,$$

which means that

$$\Sigma_{\tilde\beta_{ql}} - \Sigma_{\hat\beta} = 0 \quad \text{if} \ \ l = 1. \tag{14}$$

If $l > 1$, C can be written as

$$C = \begin{pmatrix} A^{-1/2} & 0 \\ 0 & 1 \end{pmatrix} B \begin{pmatrix} A^{-1/2} & 0 \\ 0 & 1 \end{pmatrix} =: \begin{pmatrix} C_{11} & C_2 \\ C_2^\tau & b_{p+1,p+1} \end{pmatrix}$$

for some $b_{p+1,p+1} \neq 0$. Then

$$D_1 = I_p - \Big(I_p, 0\Big)\begin{pmatrix} I_p + C_{11} & C_2 \\ C_2^\tau & b_{p+1,p+1} \end{pmatrix}^{-1}\Big(I_p, 0\Big)^\tau$$
$$= I_p - \Big(I_p + C_{11} - b_{p+1,p+1}^{-1}C_2C_2^\tau\Big)^{-1}.$$

As $C_{11}-b_{p+1,p+1}^{-1}C_2C_2^\tau$ is non-negative definite and has rank $l-1$, there exists an orthogonal matrix U such that $U^\tau\left(C_{11}-b_{p+1,p+1}^{-1}C_2C_2^\tau\right)U=\mathrm{diag}(\eta_1,\ldots,\eta_{l-1},0,\ldots,0)$, where $\eta_1,\ldots,\eta_{l-1}$ are the positive eigenvalues of $C_{11}-b_{p+1,p+1}^{-1}C_2C_2^\tau$. So, $D_1=U\,\mathrm{diag}(1-(1+\eta_1)^{-1},\ldots,1-(1+\eta_{l-1})^{-1},0,\ldots,0)U^\tau$ and $\Sigma_{\tilde\beta_{ql}}-\Sigma_{\hat\beta}=\sigma^4A_{11}^{-1/2}U\,\mathrm{diag}(1-(1+\eta_1)^{-1},\ldots,1-(1+\eta_{l-1})^{-1},0,\ldots,0)U^\tau A_{11}^{-1/2}$, which complete the proof of (i).

(ii) The asymptotic variance of $\hat\sigma^2$ is $(0,1)\Sigma^{-1}(0,1)^\tau$. Denote

$$A_{22.1}=A_{22}-A_{21}A_{11}^{-1}A_{12}=\begin{pmatrix} a & A_1\\ A_1^\tau & A_2\end{pmatrix}\quad\text{and}\quad\begin{pmatrix} A_{12}-B_1\\ -B_2\end{pmatrix}=(b,B_3),$$

where a is a non-zero scalar, A_2 is a non-singular square matrix and b a $r+1$ dimensional vector. Then

$$\begin{aligned} B^\tau A^{-1}B &= \begin{pmatrix}\sigma^{-4}A_{11} & 0\\ 0 & 0\end{pmatrix}+(b,B_3)A_{22.1}^{*-1}(b,B_3)^\tau\\ &=\begin{pmatrix}\sigma^{-4}A_{11} & 0\\ 0 & 0\end{pmatrix}+(b,B_3)\left(\begin{pmatrix}a^{-1} & 0\\ 0 & 0\end{pmatrix}+A_3A_3^\tau\right)(b,B_3)^\tau\\ &=: A_1^*+(b,B_3)A_3A_3^\tau(b,B_3)^\tau,\end{aligned}$$

where

$$A_1^*=\begin{pmatrix}\sigma^{-4}A_{11} & 0\\ 0 & 0\end{pmatrix}+\frac{bb^\tau}{a},\quad A_3^\tau=A_{22.1}^{*-1/2}\left(A_1^\tau/a\ -I_r\right)\quad\text{and}\quad A_{22.1}^*=A_2-A_1A_1^\tau/a.$$

Note that A_1^* is non-singular as the $(r+1)$-th element of b, $b_{r+1}=1/\sigma^4\neq 0$. Hence, $(0,1)\Sigma(0,1)^\tau=(0,1)[B^\tau A^{-1}B]^{-1}(0,1)^\tau=(0,1)[A_1^*+(b,B_3)A_3A_3^\tau(b,B_3)^\tau]^{-1}(0,1)^\tau$. Notice that the rank of $(b,B_3)A_3A_3^\tau(b,B_3)^\tau=l=$ rank (Δ), and $\Sigma_{\tilde\sigma^2}=(0,1)\times A_1^{*-1}(0,1)^\tau$. Using the same argument as in proving (i), (ii) can be established.

Appendix B. Proof of Theorem 2

Lemma 1. *Let ξ be a q-variate and η be a p-variate random vector where $q\le p$, $E(\|\xi\|^2+\|\eta\|^2)<+\infty$ and $E(\xi\xi^\tau)>0$, then $E(\eta\xi^\tau)[E(\xi\xi^\tau)]^{-1}E(\xi\eta^\tau)\le E(\eta\eta^\tau)$. Furthermore, equality holds if and only if $\eta=C\xi$ a.s., where $C=E(\eta\xi^\tau)[E(\xi\xi^\tau)]^{-1}$.*

Proof. Let $C=E(\eta\xi^\tau)[E(\xi\xi^\tau)]^{-1}$. As $E(\|\xi\|^2+\|\eta\|^2)<+\infty$, we have $E[(C\xi-\eta)(C\xi-\eta)^\tau]\ge 0$, which implies that $CE(\xi\xi^\tau)C^\tau-C\xi\eta^\tau-\eta\xi^\tau C^\tau+\eta\eta^\tau\ge 0$. Substituting $C=E(\eta\xi^\tau)[E(\xi\xi^\tau)]^{-1}$ into the above inequality, we have $E(\eta\xi^\tau)[E(\xi\xi^\tau)]^{-1}E(\xi\eta^\tau)\le E(\eta\eta^\tau)$. If the equality holds, it means that $E[(C\xi-\eta)(C\xi-\eta)^\tau]=0$, and then $\eta=C\xi$ *a.s.*

Proof of Theorem 2. Under the assumptions, with $D_0 = \mathrm{Cov}\,(w)$,

$$B^\tau A^{-1} B = \begin{pmatrix} \sigma^{-4} A_{11} & 0 \\ 0 & 0 \end{pmatrix} + B_1 \Big(A_{22.1}^{-1} - \frac{A_{22.1}^{-1} B_2^\tau B_2 A_{22.1}^{-1}}{B_2 A_{22.1}^{-1} B_2^\tau} \Big) B_1^\tau,$$

$$A_{22.1}^{-1} = \frac{\sigma^4}{\kappa - 1} \begin{pmatrix} 1 + E(w^\tau) D_0^{-1} E(w) & -E(w^\tau) D_0^{-1} \\ -D_0^{-1} E(w) & D_0^{-1} \end{pmatrix},$$

$$B_2^\tau B_2 = \frac{1}{\sigma^8} \begin{pmatrix} 1 & E(w) \\ E(w^\tau) & E(w)E(w^\tau) \end{pmatrix} \quad \text{and} \quad \frac{A_{22.1}^{-1} B_2^\tau B_2 A_{22.1}^{-1}}{B_2 A_{22.1}^{-1} B_2^\tau} = \frac{\sigma^4}{\kappa - 1} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

Therefore,

$$\begin{aligned} B_1 \Big(A_{22.1}^{-1} - \frac{A_{22.1}^{-1} B_2^\tau B_2 A_{22.1}^{-1}}{B_2 A_{22.1}^{-1} B_2^\tau} \Big) B_1^\tau &= \frac{\sigma^4}{\kappa - 1} B_1 \Big(E(w^\tau), -1 \Big) D_0^{-1} \Big(E(w^\tau), -1 \Big)^\tau B_1^\tau \\ &=: g(w) g^\tau(w) / (\kappa - 1), \end{aligned}$$

where $g(w) = (E[V'G'(X^\tau\beta) X w^\tau / V] - E[V'G'(X^\tau\beta)X/V] E(w^\tau)) D_0^{-1/2} = E((V'G'(X^\tau\beta)X/V)[w^\tau - E(w^\tau)] D_0^{-1/2})$.

Now $\Sigma_{\hat\beta(w)} = (I_p, 0)\Sigma^{-1}(I_p, 0)^\tau = ((A_{11}/\sigma^4) + (g(w)g(w)^\tau/\kappa - 1))^{-1}$. For two $k \times k$ positive definite symmetric matrices T and S, $T - S > 0$ iff $T^{-1} - S^{-1} < 0$. So finding a w to maximize $\Sigma_{\tilde\beta_{ql}} - \Sigma_{\hat\beta(w)}$ is equivalent to finding w to maximize $g(w)g(w)^\tau$.

As $\mathrm{Cov}\,[V'G'(X^\tau\beta)X/V] > 0$, applying Lemma 1 we have $g(w)g(w)^\tau \le E\{[V'G'(X^\tau\beta)]^2 X X^\tau / V^2\}$, and the equality holds if and only if $r = p$ and $w = [\mathrm{Cov}\,\{V'G'(X^\tau\beta)X/V\}]^{-1/2} V'G'(X^\tau\beta)X/V$. The standardization by $[\mathrm{Cov}\,\{V'G'(X^\tau\beta)X/V\}]^{-1/2}$ is unnecessary, and the optimal weight is $w = V'G'(X^\tau\beta)X/V$. This completes the proof.

References

Chen, S. X. (1994). Comparing empirical likelihood and bootstrap hypothesis tests. *J. Multivariate Anal.* **51**, 277-293.

Chen, S. X. (1996). Empirical likelihood for nonparametric density estimation. *Biometrika* **83**, 329-341.

Firth, D. (1987). On the efficiency of quasi-likelihood. *Biometrika* **74**, 233-245.

Hall, P. and La Scala, B. (1990). Methodology and algorithms of empirical likelihood. *Internat. Statist. Rev.* **58**, 109-127.

Li, G. (1995). Nonparametric likelihood ratio estimation of probabilities for truncated data. *J. Amer. Statist. Assoc.* **90**, 997-1003.

Jing, B. Y. (1995). Two-sample empirical likelihood method. *Statist. Probab. Lett.* **24**, 315-319.

Kolaczyk, E. D. (1994). Empirical likelihood for generalized linear model. *Statist. Sinica* **4**, 199-218.

McCullagh, P. and Nelder, J. A. (1990). *Generalized Linear Models*. Chapman and Hall, London.

Nelder, J. A. and Pregibon, D. (1987). An extended quasi-likelihood function. *Biometrika* **74**, 221-232.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75**, 237-249.

Owen, A. B. (1990). Empirical likelihood confidence regions. *Ann. Statist.* **18**, 90-120.

Owen, A. B. (2001). *Empirical Likelihood*. Chapman and Hall, London.

Press, W. H., Teukolsky, S. A., Vetterling and Flannery, B. F. W. T. (1992). *Numerical Recipes in C: the Art of Scientific Computing*. Cambridge University Press, Cambridge.

Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.* **22**, 300-325.

Qin, J. and Lawless, J. (1995). Estimating equations, empirical likelihood and constraints on parameters. *Canad. J. Statist.* **23**, 145-159.

Smyth, G. K. (1989). Generalized linear model with varying dispersion. *J. Roy. Statist. Soc. Ser. B* **51**, 46-60.

Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalized linear models and the Gauss-Newton method. *Biometrika* **61**, 439-447.

Department of Statistics and Applied Probability, National University of Singapore, 117543, Singapore.

E-mail: stacsx@nus.edu.sg

Department of Mathematics, Beijing Normal University, Beijing 100875, China.

E-mail: hjcui@bnu.edu.cn

(Received April 2001; accepted July 2002)