J. Statistics
1995(1), 01-10

QUASI-LIKELIHOOD AND RELATED METHODS

BY

Dr. Mujahid Rasul*

SUMMARY

It may be difficult to decide what distribution one's observations follow, but the form of the mean-variance relationship is often much easier to postulate; this is what makes quasi-likelihood useful. In some of the cases where quasi-likelihood method is unable to provide reasonable results, a more general class of estimating functions (Crowder, 1987) can be applied for better results. In this paper we describe the performance of these methods and explained it with some examples.

Keywords: Quasi-Likelihood; Generalized linear models; Exponential family; Optimal estimating function.

1-INTRODUCTION

Many of the ideas about, and procedures for fitting, generalized, linear models can be extended without difficulty when likelihoods are replaced by quasi-likelihoods. Wedderburn's (1974) introduction of quasi-likelihood greatly widened the scope of generalized linear models by allowing the full distributional assumption about the random component in the model to be replaced by a much weaker assumption in which only the first and second moments were defined. In some cases maximum quasi-likelihood estimation, which is at the core of GLIM, can fail to give reasonable results. A more general class of estimating functions has been investigated by Crowder (1987) which avoids such

* *Department of Statistics, Islamia University, Bahawalpur*

Where primes denote differentiation with respect to ϕ and $b''(\theta)$ is called the variance function. Let $l(\mu,\phi;z)$ be the log-likelihood and $l(z,\phi;z)$ be the maximum likelihood achievable for an exact fit in which fitted values equal the data.

Then minimizing

$$D(z;\mu) = -2[l(\mu,\phi;z) - l(z,\phi;z)]$$

which for generalized linear models is called as deviance is equivalent to maximizing $l(\mu,\phi;z)$. Since the variance function determines the units of measurement for $D(z,\mu)$, differencing these discrepancy measures across variance functions is not possible. To assess variance functions it is necessary to apply extended quasi-likelihood (Nelder and Pregibon, 1987).

2-QUASI-LIKELIHOOD FUNCTIONS

To define a likelihood we have to specify the form of distribution of the observation, but to define a quasi-likelihood function we need only specify a relation between the mean and variance of the observations and the quasi-likelihood can then be used for estimation. The least-squares estimates for the parameters β_i are obtained by solving the estimating equation.

$$\sum_{t=i}^{n}(z_t - \mu_t)\mu_t' = 0 \qquad (2)$$

Where $\mu_t' = \partial\mu_t/\partial\beta_i$. Wedderburn (1974) has investigated that the least-squares equations (2) may be generalized to the quasi-likelihood equations

$$\sum_{t=i}^{n}\sigma_t^{-2}(z_t - \mu_t)\mu_t' = 0 \qquad (3)$$

failure and also is not restricted to the particular forms of mean and variance function of GLIM.

Suppose observations are taken on independent random variables $Z_1, \ldots\ldots, Z_n$ with

$$E(Z_t) = \mu_t \,,\; g(\mu_t) = \sum_{j=1}^{k} \beta_j X_{tj}$$

$$Var(Z_t) = \phi V(\mu_t)\,,\; t = 1,2,\ldots\ldots,n \qquad (1)$$

Where the constant X_{tj}'s the link function g(.) and the variance function V(.) are known. The generalized linear models which is an extension of classical linear models allow that the distribution of Z may come from an exponential family and the link function may become any monotonic differentiable function (McCullagh and Nelder, 1989).

For the exponential family of densities, consider the distribution of Z in the form

$$f(z,\theta,\phi) = \frac{\exp\{z\theta - b(\theta)\}}{\{d(\phi) + c(z,\phi)\}}$$

for some functions d( ), b( ) and c( ). If ϕ is known, this is an exponential family with canonical parameter ϕ. Log-likelihood function of the above equation is

$$l(\theta,\phi,z) = \ln f(z;\theta,\phi) = \frac{z(\theta) - b(\theta)}{d(\phi) + c(z,\phi)}$$

with

$$E(Z) = \mu = b'(\theta) \quad \text{and}$$

$$Var(Z) = b''(\theta)\, d(\phi)$$

relaxed in the extended quasi-likelihood. Consider embedding the variance function in a family of functions indexed by an unknown parameter θ, so that $Var(z) = \phi\ V_\theta(\mu)$. If $K(z;\mu)$ is viewed as an approximation to a discrete distribution such as the binomial, poisson, or negative binomial, then the problem lies with the stirling's approximation used for the factorials, which approaches zero for $z \to 0$ instead of unity. Then Nelder and Pregibon (1987) suggested a modified form $m! = \{2\pi\ (m+k)\}^{1/2} m^m e^{-m}$ which can be used to get the better results. For the discrete distributions mentioned above, the use of the modified stirling's approximation yields the following results,[14] with V(z;k) replacing V(z) in (4)

	Binomial	Poisson	Negative Binomial
$V(z;0)$	$n^{-1}z(n-z)$	z	$v^{-1}z(z+v)$
$V(z;k)$	$(n+k)^{-1}(z+k)(n-z+k)$	$(z+k)$	$v^{-2}(z+v)^2(z+k)(v+k)(z+v+k)^{-1}$

The use of $\hat{V}(z;k)$ allows $K(z;\mu)$ to be defined for all sample sets and will be important if V(.) itself contains unknown parameters.

3-OPTIMAL ESTIMATING FUNCTIONS

In some cases quasi-likelihood method is not suitable to give reasonable results. Given below are the examples in which quasi-likelihood breaks down because it concentrates solely on μ_t for the information about β.

Widely-used GLIM system has main components:

(i) A known link function g relating μ_t to parameter β and explanatory variables x via $x^T\beta = g(\mu_t)$,

(ii) a known variance function V such that $\sigma_t^2 = \phi\ V(\mu_t)$ and

(iii) Maximum quasi-likelihood estimation.

which can be solved to get estimates for the parameters β. The estimator β_n is consistent and asymptotically normal with variance matrix $S_{no}^{-1}Q_{no}(S_{no}^{-1})^T$, where $S_{no} = Q_{no} = -\sum \sigma_i^{-2} \mu_i' \mu_i'^T$ (Crowder, 1987). McCullagh (1983) has also shown that under some conditions quasi-likelihood estimates are consistent and asymptotically normal.

For generalized linear models with distributions in the exponential family, likelihood ratio and score tests are used for testing hypotheses concerning nested subsets of covariates in the linear predictor and for assessing hypothesized link function. These methods are also applicable with Wedderburn's from of quasi-likelihood. However neither of these methods is suitable for the comparison of different variance functions. Nelder and Pregibon (1987) has introduced an extended quasi-likelihood function which allows for the comparison of various forms of all the components of a generalized linear model. For a single observation z with mean μ and variance $\phi V(\mu)$ this function is defined as

$$K(z;\mu) = -0.5\log\{2\pi\phi V(z)\} - 0.5D(z;\mu)\phi^{-1} \qquad (4)$$

where $D(z;\mu)$ is the deviance and ϕ is the dispersion parameter. The estimates of β obtained by maximizing $K(z;\mu)$ are the same as the maximum quasi-likelihood estimates. The estimates of ϕ obtained from maximizing $K(z;\mu)$ is $\hat{\phi} = D(z;\hat{\mu})/n$, the mean deviance. For the special cases where $K(z;\mu)$ corresponds to the normal and inverse Gaussian distributions, $\hat{\phi}$ is the MLE of ϕ. For the gamma distribution $K(z;\mu)$ differs from the log-likelihood by a factor depending only an ϕ. For discrete distributions, $K(z;\mu)$ is obtainable from the respective log-likelihood function by replacing any factorial m! by stirling's approximation $m! = (2\pi m)^{1/2} m^m e^{-m}$

Wedderburn's original quasi-likelihood model required knowing the variance function upto a multiplicative constant. This requirement is

$S_{no}^{-1}Q_{no}(S_{no}^{-1})^T$, where

$$S_{no} = Q_{no} = \Sigma\sigma_i^{-2}\{\mu_i^{'}\mu_i^{'T} + \gamma_i^{-1}(\gamma_{1i}\mu_i^{'} - 2\sigma_i^{'})(\gamma_{1i}\mu_i^{'} - 2\sigma_i^{'})^T\}.$$

As among the subclass of linear estimating functions, MQL has the optimal property. Likewise, Crowder's optimal method (OPT) has the corresponding property among fully quadratic estimating functions, but requires specification of γ_{1i} and γ_{2i} for its use.

4-EXAMPLES

Few examples are presented in this section to illustrate the performance of the methods discussed in previous sections.

Example: 4.1 (Negative Binomial Distribution) Suppose Z/θ is Poisson (θ), and θ is Gamma (λ, ν). then unconditional

$$P(Z=r) = \{\Gamma(r+\nu)/(r!\Gamma\nu)\}\{(\lambda)^\nu/(\lambda+1)^{r+\nu}\}$$

is negative binomial distribution with $p = \frac{1}{\lambda+1}$ (Doss, 1979). Thus $\ln\theta$ has mean $\psi(\nu) - \ln\lambda$ and variance $\psi'(\nu)$ over the population of individual counts; here ψ is the digamma function. To relate the parameters to the explanatory variables let us now take $\psi(\nu) - \ln\lambda = x^T\beta$ and $\psi'(\nu)$ homogenous, i.e., independent of x. In the model thus constructed Z has negative binomial distribution with parameters (ν, p). The moments of Z are

$\mu_i = \nu p/1-p$, $\sigma_i^2 = \nu p/(1-p)^2$, $\gamma_{1i} = (1+p)/(\nu p)^{1/2}$, $\gamma_{2i} = (1+4p+p^2)/\nu p$ and the derivatives are

$$p_i^{'} = \begin{pmatrix} p(1-p)x \\ 0 \end{pmatrix},\ \mu_i^{'} = \mu\begin{pmatrix} x \\ 1/\nu \end{pmatrix}, \sigma_i^{'} = \sigma/2\begin{pmatrix} 1+p \\ 1/\nu \end{pmatrix}$$

Suppose that $\mu_t = \beta/t$ and $\sigma_t^2 = 1/\mu_t$ with $\beta > 0$. If the variance function is correctly specified as $V(\mu) = 1/\mu$, then the maximum quasi-likelihood estimating equation is $\beta \sum t^{-2}\left(z_t - \beta/t\right) = 0$, $\hat{\beta}_n = \sum t^{-2}$ and $Var(\hat{\beta}_n) = 1/\beta\sum t^{-2}$. Now as $n \to \infty$ the $Var(\hat{\beta}_n)$ does not tend to zero and hence $\hat{\beta}_n$ is inconsistent. In this MQLE fails to use the information on in the second moment of z_t. To avoid this problem one might apply the method of z_t^2 rather than z_t. But the variance function for z_t^2 needs the skewness and Kurtosis of z_t as known functions of μ_t.

Suppose that $\mu_t = \lambda\beta_1 x_t + (1-\lambda)\beta_2 x_t$, corresponding to a mixture of two populations with linear regression of z on x in each, the slopes being β_1 and β_2. If the variance is s^2 in each population then $\sigma_t^2 = s^2 x_t^2 + \lambda(1-\lambda)(\beta_1 - \beta_2)^2 x_t^2$. Thus $\sigma_t^2 = \phi\, V(\mu_t)$, with $V(\mu) = \mu^2$ and ϕ unknown. The MQL estimating function is $\Sigma(z_t - \mu_t)\mu_t^{-2} x_t (\lambda, 1-\lambda)^T = 0$ reduces to single equation. Thus estimates β_1 and β_2 are not obtained. The problem here is that β_1 and β_2 are confounded in μ_t. MQL does yield a consistent estimator for the parameteric function $\lambda\beta_1 + (1-\lambda)\beta_2$, but not for β_1 and β_2 separately.

A more general class of estimating functions has been investigated by Crowder(1987) which avoids such failure. The following optimal estimating equations:

$$\sum_{t=1}^{n}\left[\frac{\dfrac{\left\{-(\gamma_{2t}+2)\mu_t' + (2\gamma_{1t} + \sigma_t'\right\}}{\sigma_t^2(\gamma_{2t}+2-\gamma_{1t}^2)}}{\left\{\dfrac{(z_t i\mu_t) + (\gamma_{1t}\mu_t' - 2\sigma_t')}{(\sigma_t^3 + 2 - \gamma_{1t}^2)}\right\}\left\{(z_t - \mu_t)^2 - \sigma_t^2\right\}}\right] = 0$$

is solved to get the estimate of the parameters β. The resulting estimator β_n is consistent and asymptotically normal with variance matrix this is that the estimating equations for β_1 and η both are equivalent to $\Sigma[\mu_t(z_t - \mu_t)] = 0$. So we have only one equation for two unknowns. Hence we can not estimate the parameters β_1 and η separately by the method of quasi-likelihood. We may apply Crowder's optimal method to estimate the parameters.

To estimate the parameters β_1 and v, we have asymptotic variance matrix for MQL:

$$S_{no}^{-1}Q_{no}(S_{no}^{-1})^T = nvp\begin{pmatrix} 1 & 1/v \\ 1/v & 1/v^2 \end{pmatrix}^{-1}$$

which is unobtainable because the above matrix is singular. The reason for this is that we have effectively only one estimating equation for two unknown parameters. Thus estimates for v and β_1 are not possible by this method. The estimates can be obtained by using Crowder's optimal method.

Example 4.2: (Weibull Distribution) The observations y_t independent with distribution function

$$F(t) = 1 - e^{-(\lambda z_i)^\eta}, (0,\infty)$$

The parameter η is assumed to be homogeneous over individuals and λ_i obeys the log-linear form $\log\lambda_i = x^T\beta$. The moments are: $\mu_i = \lambda_i^{-1}C_1, \sigma_i^2 = \lambda_i^2 \cdot C_2^2, \gamma_{1i} = C_3/C_2^3, \gamma_{2i} = C_4/C_2^4$

where, $C_1 = \Gamma_1, C_2 = \Gamma_2 - \Gamma_1^2, C_3 = \Gamma_3 - 3\Gamma_1\Gamma_2 + 2\Gamma_1^3,$

$$C_4 = \Gamma_4 + 12\Gamma_1^2\Gamma_2 - 4\Gamma_1\Gamma_3 - 3\Gamma_2^2 - 6\Gamma_1^4 \text{ and } \Gamma_j = \Gamma\left(1 + j/\eta\right)$$

for j=1,2,3,4. To estimate the parameters (β, η) we have asymptotic variance matrix for MQL

$$S_{no}^{-1}Q_{no}(S_{no}^{-1})^T = \Sigma\mu_i^2\sigma_i^{-2}\begin{pmatrix} 1/\eta^2 & -S/\eta^3 \\ -S/\eta^3 & S^2/\eta^4 \end{pmatrix}^{-1}$$

Where $S = \beta_1 - \psi(1 + 1/\eta)$. The above matrix is singular so it is not possible to obtain the estimates of the parameters (β_1, η) the reason for

REFERENCES

Crowder, M.J. (1987) *On linear and quadratic estimating functions.* Biometrika, 74, 591-597

McCullagh, P. (1983) *Quasi-likelihood functions.* Ann. Statist.11, 59-67

McCullagh, P. and Nelder, J.A (1989) *Generalized linear models, 2nd edn.* New York: Chapman and Hall.

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

Wedderburn, R.W.M. (1974). *Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method.* Biometrika, 61, 439-47