

Generalized Quasi-likelihood (GQL) Inference*

by Brajendra C. Sutradhar

Memorial University

Email address: bsutradh@mun.ca

QL Estimation for Independent Data. For $i = 1, \dots, K$, let Y_i denote the response variable for the i th individual, and $x_i = (x_{i1}, \dots, x_{iv}, \dots, x_{ip})'$ be the associated p -dimensional covariate vector. Also, let β be the p -dimensional vector of regression effects of x_i on y_i . Further suppose that the responses are collected from K independent individuals. It is understandable that if the probability distribution of Y_i is not known, then one can not use the well known likelihood approach to estimate the underlying regression parameter β . Next suppose that only two moments of the data, that is, the mean and the variance functions of the response variable Y_i for all $i = 1, \dots, K$, are known, and for a known functional form $a(\cdot)$, these moments are given by

$$E[Y_i] = a'(\theta_i) \text{ and } \text{var}[Y_i] = a''(\theta_i), \quad (1)$$

where for a link function $h(\cdot)$, $\theta_i = h(x_i'\beta)$, and $a'(\theta_i)$ and $a''(\theta_i)$ are the first and second order derivatives of $a(\theta_i)$, respectively, with respect to θ_i . For the estimation of the regression parameter vector β under this independence set up, Wedderburn (1974) (see also McCullagh (1983)) proposed to solve the so-called quasi-likelihood (QL) estimating equation given by

$$\sum_{i=1}^K \left[\frac{\partial a'(\theta_i)}{\partial \beta} \frac{(y_i - a'(\theta_i))}{a''(\theta_i)} \right] = 0. \quad (2)$$

Let $\hat{\beta}_{QL}$ be the QL estimator of β obtained from (2). It is known that this estimator is consistent and highly efficient. In fact, for Poisson and binary data, for example, $\hat{\beta}_{QL}$ is equivalent to the maximum likelihood (ML) estimator and hence it turns out to be an optimal estimator.

Illustration for the Poisson case: For the Poisson data, one uses

$$a(\theta_i) = \exp(\theta_i) \quad (3)$$

with identity link function $h(\cdot)$, that is, $\theta_i = x_i'\beta$. This gives the mean and the variance functions as

$$\text{var}(Y_i) = a''(\theta_i) = E(Y_i) = a'(\theta_i) = \mu_i \text{ (say)} = \exp(x_i'\beta),$$

yielding by (2), the QL estimating equation for β as

$$\sum_{i=1}^K x_i (y_i - \mu_i) = 0. \quad (4)$$

Note that as the Poisson density is given by $f(y_i|x_i) = \frac{1}{y_i!} \exp[y_i \log(\mu_i) - \mu_i]$, with $\mu_i = \exp(\theta_i) = \exp(x_i'\beta)$, it follows that the log likelihood function of β has

the form $\log L(\beta) = -\sum_{i=1}^K \log(y_i!) + \sum_{i=1}^K [y_i \theta_i - a(\theta_i)]$, yielding the likelihood equation for β as

$$\frac{\partial \log L}{\partial \beta} = \sum_{i=1}^K [y_i - a'(\theta_i)] \frac{\partial \theta_i}{\partial \beta} = \sum_{i=1}^K x_i (y_i - \mu_i) = 0, \quad (5)$$

which is the same as the QL estimating equation (4). Thus, if the likelihood function were known, then the ML estimate of β would be the same as the QL estimate $\hat{\beta}_{QL}$.

Illustration for the binary case: For the binary data, one uses

$$a'(\theta_i) = \frac{\exp(\theta_i)}{1 + \exp(\theta_i)} = \mu_i \text{ and } a''(\theta_i) = \mu_i(1 - \mu_i), \quad (6)$$

with $\theta_i = x_i' \beta$. The QL estimating equation (2) for the binary data, however, provides the same formula (4) as in the Poisson case, except that now for the binary case $\mu_i = \frac{\exp(\theta_i)}{1 + \exp(\theta_i)}$, whereas for the Poisson case $\mu_i = \exp(\theta_i)$.

As far as the ML estimation for the binary case is concerned, one first writes the binary density given by $f(y_i|x_i) = \mu_i^{y_i} (1 - \mu_i)^{1-y_i}$. Next by writing the log likelihood function as $\log L(\beta) = \sum_{i=1}^K y_i \mu_i + \sum_{i=1}^K (1 - y_i)(1 - \mu_i)$, one obtains the same likelihood estimating equation as in (5), except that here $\mu_i = \frac{\exp(x_i' \beta)}{1 + \exp(x_i' \beta)}$, under the binary model. Since the QL estimating equation (4) is the same as the ML estimating equation (5), it then follows that the ML and QL estimates for β would also be the same for the binary data.

GQL Estimation: A Generalization of the QL Estimation to the Correlated Data.

As opposed to the independence set up, we now consider y_i as a vector of T repeated binary or count responses, collected from the i -th individual, for all $i = 1, \dots, K$. Let $y_i = (y_{i1}, \dots, y_{it}, \dots, y_{iT})'$, where y_{it} represents the response recorded at time t for the i th individual. Also, let $x_{it} = (x_{it1}, \dots, x_{itv}, \dots, x_{itp})'$ be the p -dimensional covariate vector corresponding to the scalar y_{it} , and β be the p -dimensional regression effects of x_{it} on y_{it} for all $i = 1, \dots, K$, and all $t = 1, \dots, T$. Suppose that μ_{it} and σ_{itt} be the mean and the variance of Y_{it} , that is $\mu_{it} = E[Y_{it}]$ and $\text{var}[Y_{it}] = \sigma_{itt}$. Note that both μ_{it} and σ_{itt} are functions of β . But, when the variance is a function of mean, it is sufficient to estimate β involved in the mean function only, by treating β involved in the variance function to be known. Further note that since the T repeated responses of an individual are likely to be correlated, the estimate of β to be obtained by ignoring the correlations, that is, the solution of the independence assumption based QL estimating equation

$$\sum_{i=1}^K \sum_{t=1}^T \left[\frac{\partial \mu_{it}}{\partial \beta} \frac{(y_i - \mu_{it})}{\sigma_{itt}} \right] = 0, \quad (7)$$

for β , will be consistent but inefficient. As a remedy to this inefficient estimation problem, Sutradhar (2003) has proposed a generalization of the QL estimation approach, where β is now obtained by solving the GQL estimating equation given by

$$\sum_{i=1}^K \frac{\partial \mu'_i}{\partial \beta} \Sigma_i^{-1}(\rho)(y_i - \mu_i) = 0, \quad (8)$$

where $\mu_i = (\mu_{i1}, \dots, \mu_{it}, \dots, \mu_{iT})'$ is the mean vector of Y_i , and $\Sigma_i(\rho)$ is the covariance matrix of Y_i that can be expressed as $\Sigma_i(\rho) = A_i^{\frac{1}{2}} C_i(\rho) A_i^{\frac{1}{2}}$, with $A_i = \text{diag}[\sigma_{i11}, \dots, \sigma_{itt}, \dots, \sigma_{iT T}]$ and $C_i(\rho)$ as the correlation matrix of Y_i , ρ being a correlation index parameter.

Note that the use of the GQL estimating equation (8) requires the structure of the correlation matrix $C_i(\rho)$ to be known, which is, however, unknown in practice. To overcome this difficulty, Sutradhar (2003) has suggested a general stationary auto-correlation structure given by

$$C_i(\rho) = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{T-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{T-2} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho_{T-1} & \rho_{T-2} & \rho_{T-3} & \cdots & 1 \end{bmatrix}, \quad (9)$$

(see also Sutradhar and Das (1999, Section 3)), for all $i = 1, \dots, K$, where for $\ell = 1, \dots, T-1$, ρ_ℓ represents the lag ℓ auto-correlation. As far as the estimation of the lag correlations is concerned, they may be consistently estimated by using the well known method of moments. For $\ell = |u - t|$, $u \neq t$, $u, t = 1, \dots, T$, the moment estimator for the autocorrelation of lag ℓ , ρ_ℓ , has the formula

$$\hat{\rho}_\ell = \frac{\sum_{i=1}^K \sum_{t=1}^{T-\ell} \tilde{y}_{it} \tilde{y}_{i,t+\ell} / K(T-\ell)}{\sum_{i=1}^K \sum_{t=1}^T \tilde{y}_{it}^2 / KT}, \quad (10)$$

(Sutradhar and Kovacevic (2000, eqn. (2.18), Sutradhar (2003)), where \tilde{y}_{it} is the standardized residual, defined as $\tilde{y}_{it} = (y_{it} - \mu_{it}) / \{\sigma_{itt}\}^{\frac{1}{2}}$.

The GQL estimating equation (8) for β and the moment estimate of ρ_ℓ by (10) are solved iteratively until convergence. The final estimate of β obtained from this iterative process is referred to as the GQL estimate of β , and may be denoted by $\hat{\beta}_{GQL}$. This estimator $\hat{\beta}_{GQL}$ is consistent for β and also highly efficient, the ML estimator being fully efficient which is however impossible or extremely complex to obtain in the correlated data set up.

With regard to the generality of the stationary auto-correlation matrix $C_i(\rho)$ in (9), one may show that this matrix, in fact, represents the correlations of many stationary dynamic such as stationary auto-regressive order 1 (AR(1)), stationary moving average order 1 (MA(1)), and stationary equi-correlations (EQC) models. For example, consider the stationary AR(1) model given by

$$y_{it} = \rho * y_{i,t-1} + d_{it}, \quad (11)$$

Table 1. A class of stationary correlation models for longitudinal count data and basic properties.

Model	Dynamic relationship	Mean-variance & Correlations
AR(1)	$y_{it} = \rho * y_{i,t-1} + d_{it}, t = 2, \dots$ $y_{i1} \sim Poi(\mu_{i.})$ $d_{it} \sim P(\mu_{i.}(1 - \rho)), t = 2, \dots$	$E[Y_{it}] = \mu_{i.}$ $\text{var}[Y_{it}] = \mu_{i.}$ $\text{corr}[Y_{it}, Y_{i,t+\ell}] = \rho^\ell$ $= \rho^\ell$
MA(1)	$y_{it} = \rho * d_{i,t-1} + d_{it}, t = 2, \dots$ $y_{i1} = d_{i1} \sim Poi(\mu_{i.}/(1 + \rho))$ $d_{it} \sim P(\mu_{i.}/(1 + \rho)), t = 2, \dots$	$E[Y_{it}] = \mu_{i.}$ $\text{var}[Y_{it}] = \mu_{i.}$ $\text{corr}[Y_{it}, Y_{i,t+\ell}] = \rho^\ell$ $= \begin{cases} \frac{\rho}{1+\rho} & \text{for } \ell = 1 \\ 0 & \text{otherwise,} \end{cases}$
EQC	$y_{it} = \rho * y_{i1} + d_{it}, t = 2, \dots$ $y_{i1} \sim Poi(\mu_{i.})$ $d_{it} \sim P(\mu_{i.}(1 - \rho)), t = 2, \dots$	$E[Y_{it}] = \mu_{i.}$ $\text{var}[Y_{it}] = \mu_{i.}$ $\text{corr}[Y_{it}, Y_{i,t+\ell}] = \rho^\ell$ $= \rho$

(McKenzie (1988), Sutradhar (2003)) where it is assumed that for given $y_{i,t-1}$, $\rho * y_{i,t-1}$ denotes the so-called binomial thinning operation (McKenzie, 1988). That is,

$$\rho * y_{i,t-1} = \sum_{j=1}^{y_{i,t-1}} b_j(\rho) = z_{i,t-1}, \text{ say,} \quad (12)$$

with $\Pr[b_j(\rho) = 1] = \rho$ and $\Pr[b_j(\rho) = 0] = 1 - \rho$. Furthermore, it is assumed in (11) that y_{i1} follows the Poisson distribution with mean parameter $\mu_{i.}$, that is, $y_{i1} \sim Poi(\mu_{i.})$, where $\mu_{i.} = \exp(x'_{i.}\beta)$ with stationary covariate vector x_i , such that $x_{it} = x_i$ for all $t = 1, \dots, T$. Further, in (11), $d_{it} \sim P(\mu_{i.}(1 - \rho))$ and is independent of $z_{i,t-1}$. This model in (11) yields the mean, variance and auto-correlations of the data as shown in Table 1. The Table 1 also contains the MA(1) and EQC models and their basic properties including the correlation structures.

It is clear from Table 1 that the correlation structures for all three processes can be represented by $C_i(\rho)$ in (9). By following Qaqish (2003), one may write similar but different dynamic models for the repeated binary data, with their correlation structures represented by $C_i(\rho)$. Thus, if the count or binary data follow this type of auto-correlations model, one may then certainly estimate the regression vector consistently and efficiently by solving the general auto-correlations matrix based GQL estimating equation (8), where the lag correlations are estimated by (10) consistently.

[* Reprinted with permission from Lovric, Miodrag (2011), International Encyclopedia of Statistical Science. Heidelberg: Springer Science & Business Media, LLC]

References

- McCullagh, P. (1983). Quasilikelihood functions. *Ann. Statist.* 11, 59-67.
- McKenzie, E. (1988). Some ARMA models for dependent sequences of Poisson counts. *Advances in Applied Probability* 20, 822-35.
- Qaqish, B. F. (2003). A family of multivariate binary distributions for simulating correlated binary variables with specified marginal means and correlations. *Biometrika* 90, 455-463.
- Sutradhar, B. C. (2003). An overview on regression models for discrete longitudinal responses. *Statistical Science* 18, 377-93.
- Sutradhar, B. C. & Das, K. (1999). On the efficiency of regression estimators in generalized linear models for longitudinal data. *Biometrika* 86, 459-65.
- Sutradhar, B. C. & Kovacevic, M. (2000). Analyzing ordinal longitudinal survey data: Generalized estimating equations approach. *Biometrika* 87, 837-848.
- Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalised linear models, and the Gauss-Newton method. *Biometrika* 61, 439-447.