Quasi-Likelihood Models in R

Jeff Gill
University of Florida
jgill@polisci.ufl.edu

Model Specification R

One of the best features of the R implementation of the S statistical language is that it has a uniform expression for parametric model specification. Model formulas defining variable relationships are essentially the same whether the researcher is specifying a linear model, generalized linear model, generalized additive model, or even some nonparametric forms. This general form is given by: `Y ~ X1 + X2` within the specific model being called. There is more flexibility here than it initially appears. Placing `-1` on the right-hand-side excludes the intercept. The plus sign here is just the most basic specification: changing `+` to `*` gives an interaction term *and* the main effects for `X1` and `X2`, or using `:` in its place just gives the interaction effect, terms can be nested with `%in%` (or equivalently `/`), and `-` to include all of `X1` not in `X2`. It is also sometimes convenient to embed functions into the specification, such something like: `sqrt(Y) ~ cos(X1) + I(X2 / X3)`, where the `I()` function protects the division operation from being interpreted as a nesting specification.

There are more ways that model formulas can be specified (described nicely in Chambers and Hastie (1993) Ch.2, and Fox (2002) Ch.4). Wrapped around this construction is the particular model being run: `lm`, `glm`, `gam`, `nls`, `nlme`, etc. Within that function the user can specify: the data used, factor contrasts, the link function for glms, offsets for fixing constant certain coefficients, random effects terms, weights, starting values for the algorithm, and in some cases the form of the optimization method used. So for example, a gamma glm might look like:

```
copper.stage1 <- glm(COPPERPRICE ~ INCOMEINDEX
    + ALUMPRICE + INVENTORYINDEX + log(TIME),
    family=Gamma,data=copper.dat,
    control=glm.control(epsilon=0.0001,
    maxit=10, trace=F))
```

where the `glm.control` sub-function stipulates: the convergence threshold, the maximum number of *iteratively weighted least squares* (the workhorse glm numerical estimation method) iterations, and whether or not to print steps to the screen. The specification `family` is the means by which the glm link function is identified, and other common forms are obtained by simply replacing `Gamma` here with: `poisson`, `gaussian`, `inverse.gaussian`, `binomial`, and others. Chapter 7 of what is certainly the best *statistical* reference to the S language, Venables and Ripley (1999), contains an explanation of these and other forms.

The purpose of this article is to demonstrate the power and flexibility of the R environment with a specific example of an under-appreciated model form. Even though the theory for quasi-likelihood models can be a little involved (but given below), the means of specifying a quasi-likelihood glm in R are very nearly trivial. For instance, the gamma glm given above can be made into a quasi-likelihood model with the change:

```
copper.stage1 <- glm(COPPERPRICE ~ INCOMEINDEX
    + ALUMPRICE + INVENTORYINDEX + log(TIME),
    family=quasi(link="inverse",var="mu^2"),
    data=copper.dat, control=glm.control(
    epsilon=0.0001, maxit=10, trace=F))
```

where to preserve the gamma link parameter we stipulate that the link is “inverse,” and the general `quasi` function allows three other types of variance terms: `mu(1-mu)`, `mu`, and `mu^3`.

Quasi-Likelihood

Wedderburn (1974) introduced the concept of “quasi-likelihood” estimation to extend the standard generalized linear model of Nelder and Wedderburn 1972 to the circumstance when the parametric form of the likelihood is known to be misspecified, or only the first two moments are definable. The goal is to create a more flexible form that retains desirable GLM properties (i.e. those described in Fahrmeir and Kaufmann 1985 and Wedderburn 1976).

Suppose that we know something about the parametric form of the distribution generating the data, but not in complete detail. Obviously this precludes the standard maximum likelihood estimation of unknown parameters since we cannot specify a full likelihood equation. Wedderburn's idea was to develop an estimation procedure that only requires specification for the mean function of the data and a stipulated relationship between this mean function and the variance function. This is also useful in a Bayesian context when we have prior information readily at hand but only a vague idea of the form of the likelihood.

Instead of taking the first derivative of log likelihood with respect to the parameter vector, $\boldsymbol{\theta}$, suppose we take this derivative with respect to the mean function in a generalized linear model, μ, with the analogous properties:

- $E\left[\frac{\partial \ell(\boldsymbol{\theta})}{\partial \mu_i}\right] = 0.$
- $Var\left[\frac{\partial \ell(\boldsymbol{\theta})}{\partial \mu_i}\right] = \frac{1}{\phi v(\mu_i)}.$
- $-E\left[\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \mu_i^2}\right] = \frac{1}{\phi v(\mu_i)}.$

Therefore what we have here is a linkage between the mean function and the variance function that does not depend on the form of the likelihood function, and we have a replacement for the unknown specific form of the score function that still provides the desired properties of maximum likelihood estimation as described. Thus we imitate these three criteria of the score function with a function that contains significantly less parametric information: only the mean and variance.

A function that satisfies these three conditions is:

$$q = \frac{y_i - \mu_i}{\phi v(\mu_i)} \qquad (1)$$

(reference: McCullagh and Nelder 1989, p.325; Shao 1999, p.314). The associated contribution to the log likelihood function from the i^{th} point is defined by:

$$Q_i = \int_{y_i}^{\mu_i} \frac{y_i - t}{\phi v(\mu_i)},$$

so finding the maximum likelihood estimator for this setup, $\hat{\boldsymbol{\theta}}$ is equivalent to solving:

$$\begin{aligned} \frac{\partial}{\partial \theta} \sum_{i=1}^{n} Q_i &= \sum_{i=1}^{n} \frac{y_i - \mu_i}{\phi) v(\mu_i)} \frac{\partial \mu_i}{\partial \theta} \\ &= \sum_{i=1}^{n} \frac{y_i - \mu_i}{\phi) v(\mu_i)} \frac{\mathbf{x}_i}{g(\mu_i)} = \mathbf{0}, \end{aligned}$$

where $g(\mu)$ is the canonical link function for a generalized linear model specification. In other words we can use the usual maximum likelihood engine for inference with complete asymptotic properties such as consistency and normality (McCullagh 1983), by only specifying the relationship between the mean and variance functions as well as the link function (which actually comes directly from the form of the outcome variable data).

As an example suppose we assume that the mean and variance function are related by stipulating that $\phi = \sigma^2 = 1$, and $b(\theta(\mu_i)) = \frac{\theta(\mu_i)^2}{2}$, so $v(\mu) = \frac{\partial^2 b(\theta(\mu_i))}{\partial \theta(\mu_i)^2} = 1$. Then it follows that:

$$Q_i = \int_{y_i}^{\mu_i} \frac{y_i - t}{\phi) v(\mu)} = -\frac{(y_i - \mu_i)^2}{2}.$$

The quasi-likelihood solution for $\hat{\boldsymbol{\theta}}$ comes from solving the quasi-likelihood equation:

$$\frac{\partial}{\partial \theta} \sum_{i=1}^{n} Q_i = \frac{\partial}{\partial \theta} \sum_{i=1}^{n} \frac{y_i - \theta}{2} = -\sum_{i=1}^{n} y_i + n\theta = \mathbf{0}.$$

In other words, $\hat{\boldsymbol{\theta}} = \bar{y}$, because this example was setup with the same assumptions as a normal maximum likelihood problem but without specifying a normal likelihood function.

In this way quasi-likelihood models drop the requirement that the true underlying density of the outcome variable belong to a particular exponential family form. Instead, all that is required is the identification of the first and second moments and an expression for their relationship up to a proportionality constant. It is assumed that the observations are independent and that mean function describes the mean effect of interest. Even given this generalization of the likelihood assumptions, it can be shown that quasi-likelihood estimators are consistent asymptotically equal to the true estimand (Fahrmeir and Tutz 2001, p.55-60, Firth 1987; McCullagh 1983). However, a quasi-likelihood estimator is often less efficient than a corresponding maximum likelihood estimator and can never be more efficient: $V_{quasi}(\theta) \geq [I(\theta)]^{-1}$, where $I(\theta)$ is the Fisher information from the maximum likelihood estimation (McCullagh and Nelder 1987, p.347-8; Shao 1999, p.248-57).

Despite this drawback with regard to variance, there are often times when it is convenient or necessary to specify a quasi-likelihood model. A number of authors have extended the quasi-likelihood framework to: *extended quasi-likelihood* models to compare different variance functions for the same data (Nelder and Pregibon 1987), *pseudo-likelihood* models which build upon extended quasi-likelihood models by substituting a χ^2 component instead of a deviant component in dispersion analysis (Breslow 1990; Carroll and Ruppert 1982; Davidian and Carroll 1987), and models where the dispersion parameter is dependent on specified covariates (Smyth 1989). Nelder and Lee (1992) provide an informative overview of these variations. It is also the case that quasi-likelihood models are not more difficult to compute (Nelder 1985), and the R package has pre-programmed functions that make the process routine.

A Detailed Example of Quasi-Likelihood Estimation in R

A relatively well-known example of a glm specification that is improved by a quasi-likelihood specification is the ship damage dataset from McCullagh and Nelder (1989, p. 204). The data are provided as 19 rows corresponding to the observed combinations of type of ship and year built and 4 columns, as follows: ship type, coded 1-5 for A, B, C, D and E, year built (1=1960-64, 2=1965-70, 3=1970-74, 4=1975-79), months of service, ranging from 63 to 20,370, and damage incidents, ranging from 0 to 53. Note that there no ships of type E built in 1960-64. The outcome variable is the number of damage incidents (`Accidents`).

Here I will replicate McCullagh and Nelder's results using a quasi-likelihood approach which they justify by noting (p.206) that "For the random variation in the model, the Poisson distribution might be thought appropriate as a first approximation, but there is undoubtedly some inter-ship variability in accident-proneness." The primary modeling difference, as stated before, is simply the inclusion of the "quasi" statement. This has some different forms depending on the link:

```
quasi(link="identity",variance="constant")
quasibinomial(link="logit")
quasipoisson(link="log")
```

There is some flexibility with regard to the link specification. The most general and flexible form, `quasi`, requires the link and variance functions only, with the defaults given above. The `quasibinomial` form allows the link to asserted as `logit`, `log`, `probit`, and `cloglog`. The `quasipoisson` form allows `identity`, `log`, and `sqrt`. The purpose of separating out the last two types is that these forms do not fix the dispersion parameter at one, and thus are able to capture over-dispersion.

Start by setting up the data structure:

```
ships.df <- data.frame(read.table(
    "http://web.clas.ufl.edu/~jgill/
        GLM.Data/ship.data",header=T))
ships.df$Type <- factor(ships.df$Type)
ships.df$Construction
    <- factor(ships.df$Construction)
ships.df$Start <- factor(ships.df$Start)
attach(ships.df)
```

where we specifically indicate the factors in the dataframe. The default unordered contrast in R is "treatment" which is what McCullagh and Nelder use. This can be verified (or changed) with `options()$contrasts`. They also stipulate that the coefficient on years of service (`Period`) is known to be 1, thus requiring the use an offset to replicate their model. The model is run with the simple command:

```
ships.out <- glm(Accidents ~ Type+Construction
    + Start + offset(log1p(Period)),
    family=quasipoisson, control=glm.control(
    epsilon=0.0001,maxit=100))
```

where the results are given by:

```
summary(ships.out)

Deviance Residuals:
     Min       1Q   Median       3Q      Max
-1.67595 -0.65993 -0.09363 0.37454 2.79094

Coefficients:
                Estimate Std. Error t value
(Intercept)     -6.40691    0.25347 -25.277
TypeB           -0.54261    0.20702  -2.621
TypeC           -0.68788    0.38248  -1.798
TypeD           -0.07675    0.33870  -0.227
TypeE            0.32499    0.27499   1.182
Construction65   0.69721    0.17444   3.997
Construction70   0.81880    0.19787   4.138
Construction75   0.45335    0.27176   1.668
Start75          0.38463    0.13785   2.790
---
(Dispersion parameter for quasipoisson family
    taken to be 1.359193)
```

These are exactly McCullagh and Nelder's results (except that they round). For details about interpreting glm results I can (not surprisingly) recommend my favorite book on the topic (Gill 2000), or some of the standards listed in the reference sections.

Redux

The point here has been to show the ease with which model specifications can be as complex as necessary, but still easy to implement in `R`. While one leaves behind the "point and click" world of more primitive statistical software, the gain is quite obviously increased power and flexibility.

Bibliography

Breslow, N. (1990). Tests of Hypotheses in Overdispersed Poisson Regression and Other Quasi-Likelihood Models. *Journal of the American Statistical Association* **85**, 565-571.

Carroll, R. J. and Ruppert, D. (1982). Robust Estimation in Heteroscedastic Linear Models. *Annals of Statistics* **10**, 429-441.

Chambers, John M. and Trevor J. Hastie. (1993). *Statistical Models in S.* New York: Chapman and Hall.

Davidian, M. and Carroll, R. J. (1987). Variance Function Estimation. *Journal of the American Statistical Association* **82**, 1079-1091.

Fahrmeir, L., and Kaufmann, H. (1985). Consistency and Asymptotic Normality of the Maximum Likelihood Estimator in Generalized Linear Models. *The Annals of Statistics* **13**, 342-368.

Fahrmeir, L. and Tutz, G. (2001). *Multivariate Statistical Modelling Based on Generalized Linear Models.* Second Edition. New York: Springer.

Firth, David. (1987). On the Efficiency of Quasi-Likelihood Estimation. *Biometrika* **74**, 233-245.

Fox, John. (2002). *An R and S-Plus Companion to Applied Regression.* Thousand Oaks, CA: Sage Publications.

McCullagh, Peter. (1983). Quasi-Likelihood Functions. *Annals of Statistics* **11**, 59-67.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models.* Second Edition. New York: Chapman & Hall.

Nelder, J. A. (1985). Quasi-likelihood and GLIM. *Lecture Notes in Statistics* **32**, 120-127.

Nelder, J. A. and Lee, Y. (1992). Likelihood, Quasi-likelihood and Pseudo-likelihood: Some Comparisons. *Journal of the Royal Statistical Society, Series B* **54**, 273-284.

Nelder, J. A. and Pregibon, D. (1987). An Extended Quasi-Likelihood Function. *Biometrika* **74**, 221-232.

Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society, Series A* **135**, 370-385.

Shao, Jun. (1999). *Mathematical Statistics.* New York: Springer-Verlag.

Smyth, Gordon K. (1989). Generalized Linear Models with Varying Dispersion. *Journal of the Royal Statistical Society, Series B* **51**, 47-60.

Venables, W. N. and B. D. Ripley. (1999). *Modern Applied Statistics with S-PLUS.* New York: Springer-Verlag.

Wedderburn, R. W. M. (1976). On the Existence and Uniqueness of the Maximum Likelihood Estimates for Certain Generalized Linear Models. *Biometrika* **63**, 27-32.

Wedderburn, R. W. M. (1974). Quasi-Likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method. *Biometrika* **61**, 439-447.