*The Annals of Statistics*
1984, Vol. 12, No. 4, 1589–1596

# BOOK REVIEW

P. McCullagh and J. A. Nelder, *Generalized Linear Models*, 1983, xiii + 261 pages, $31.00.

Review by Daryl Pregibon

*AT&T Bell Laboratories*

**1. Introduction.** A generalized linear model is composed of three parts:

(i) a *random component* $f(y; \mu)$ specifying the stochastic behavior of a response variable $Y$;

(ii) a *systematic component* $\eta = x\beta$ specifying the variation in the response variable accounted for by known covariates $x$; and

(iii) a *link function* $g(\mu) = \eta$ specifying the relationship between the random and systematic components.

The random component $f(y; \mu)$ is typically an exponential family distribution with $E(Y) = \mu$. The link function $g$ is any strictly monotone differentiable function.

Particular instances of generalized linear models have appeared in the statistical literature over the past century. These include classical linear models, logit and probit models for proportions, loglinear models for counts, and regression models with constant coefficient of variation rather than constant variance.

Grizzle, Starmer, and Koch (1969) proposed the general class of models defined by (ii) and (iii) above but (implicitly) with $f(y; \mu) = \text{Poisson}(\mu)$ in (i). Dempster (1971) proposed the general class of models defined by (i) and (ii) above but (implicitly) with $g(\mu) = \theta$, the canonical link, in (iii). Nelder and Wedderburn (1972) unified the theory and coined the name "generalized linear model." Wedderburn (1974) extended the theory to the important class of quasi-likelihoods where the assumption of an exponential family distribution in (i) is relaxed by second-moment assumptions of the form $\text{var}(Y) \propto V(\mu)$. Numerous other papers have been written on various aspects of generalized linear models in the past decade. An international conference on generalized linear models was held in London (Gilchrist, 1982). An important software package, GLIM (Baker and Nelder, 1978), specifically designed to fit generalized linear models, is used widely.

The monograph by McCullagh and Nelder is the first extensive treatment of generalized linear models. It is important for at least two reasons:

- it makes the theory and application of generalized linear models accessible to a wide audience which until now has (largely) been restricted to the British school; and

Received April 1984; revised July 1984.

- it can significantly influence the way regression modeling is currently taught and practiced, emphasizing the similarities between various models rather than their differences.

This review demonstrates my overall enthusiasm for the monograph. It's major shortcoming is that the authors have fallen victim to the generality of their model. That is, the generalized linear model framework leaves much scope for testing hypotheses, estimating nuisance parameters, and assessing goodness-of-fit. The authors do a thorough job of describing the alternatives, but rarely come out strongly with a preference. Despite this criticism, I believe the monograph will be exceedingly useful for researchers faced with a nonstandard modeling problem.

The review is divided into two parts. The first contains an annotated description of the material contained in the monograph. The second contains important new directions which I think research in generalized linear models is likely to follow.

## 2. Review.

2.1 *What's there.* The monograph begins with a nice historical account of the origins of generalized linear models. The authors ease the reader into the material in a manner reminiscent of Cox (1970). Selected examples illustrate well the scope and flexibility of generalized linear models. Bibliographic notes at the end of the chapter provide a useful pointer to readers interested in pursuing details. Unfortunately, this practice is not consistently followed in later chapters.

Next follows the introduction of generalized linear models as a unified theory of regression modeling as summarized in my initial paragraph. Brief accounts of goodness of fit, residual analysis, and fitting algorithms are presented. This chapter would have been an ideal place to contrast the variety of ad hoc methods, e.g. response variable transformation, for analyzing data which do not follow the classical linear model assumptions with the unified generalized linear model framework. Instead this material is dispersed throughout the monograph with the main thrust appearing in Section 10.3.3 (page 198!).

The next chapter provides an overview of classical linear models. A number of topics are lucidly discussed including qualitative and continuous covariates, operators for model specification, aliasing (collinearity), geometry of least squares, and fitting algorithms. Selection of covariates is the final topic in this relatively long chapter. The authors justify this length by arguing that methods for the classical linear model are relevant to the entire class of generalized linear models. I would have preferred the general development rather than see it done (yet again!) for least squares.

The next four chapters deal with specific instances of generalized linear models for dichotomous (binary) data, polytomous data, count data, and continuous data with constant coefficient of variation. The examples tend to be somewhat complicated, but each is done with much clarity, and with emphasis correctly placed on interpretation of results. Asymptotic results are touched upon in each chapter though the details are relegated to appendices. Conditional methods are discussed in detail, another reminder of the Coxian style of this monograph. However, the connection between the neat, unified theory of generalized linear models and specific conditional models is a bit tenuous.

Chapter 8 introduces quasi-likelihood models, which were hinted at in nearly every preceding chapter. The presentation of this particularly exciting area is very clear. Basically, optimal (in a generalized Gauss–Markov sense) estimates of $\beta$ can be obtained by making only second-moment assumptions about the random variation in the response $Y$. In particular, the assumption of an exponential family distribution $f(y; \mu)$ is relaxed to the functional relationship $\text{var}(Y) = \phi V(\mu)$, where $\phi$ is assumed constant across samples. This is extremely useful in applied work when data are limited and information on the distribution of $Y$ is lacking. Indeed, if Tukey's Rule of 5 (that is, don't even think of estimating the $k$th moment of a distribution unless you have $5^k$ observations) is generally applicable, one is seldom in the comfortable position of having enough information to estimate a distribution (say by the first four moments). However one may have enough data to reliably estimate a *relationship* between the first two moments as required by the quasi-likelihood model. My only criticism of the chapter is that the first example is introduced in some detail and then rather abruptly dismissed. For those unfamiliar with Wedderburn's (1974) paper, the exposition is likely to be somewhat terse.

The next chapter deals with models for survival data. This is my least favorite chapter. Essentially it demonstrates that censored survival data can be "bent" to fit into the generalized linear model framework. This is undesirable since I feel much of the motivation for generalized linear models comes from the desire to obviate the need to "bend" data (via transformations) to fit the classical linear model. In defense of the authors, when I taught a course on generalized linear models at the University of Washington in 1981, I too included the tricks necessary to analyze censored survival data within the generalized linear model framework. This proved confusing to the students and in retrospect should have been avoided.

Chapter 10 is concerned with formal parametric methods and Chapter 11 with informal graphical methods for assessing a fitted generalized linear model. Both of these chapters are essential and represent much of the recent research on generalized linear models. Topics include methods for assessing link function adequacy, residual analysis, and the detection of influential points. An extended quasi-likelihood model is introduced (see below) which relaxes the assumption that the variance function is known up to a multiplicative constant, to include nonlinear parameters, e.g. $V_\lambda(\mu) = \mu^\lambda$. An important implication of this extension is that the adequacy of an hypothesized variance function can be assessed without replication. An example which required much hand-waving to justify in Chapter 7 is used to illustrate many of the methods. This is indeed useful but I can't help wonder why this material couldn't have come *before* particular instances of generalized linear models rather than *after*.

The final chapter is a five page look at what the authors see as the important topics for future research. The section headings are "Second-moment assumptions," "Higher-order asymptotic approximations," "Composite link functions," "Missing values and the EM algorithm", and "Components of dispersion." The list is short and clearly demonstrates the authors' research interests. The section entitled "Components of Dispersion" is a superb start into the important and difficult extension of the theory of generalized linear models to random effects models.

2.2 *What's not there.* Page constraints for monographs in this series make it easy for a reviewer to identify areas which the authors failed to give adequate coverage. In order to be fair then, the reviewer must indicate areas of the monograph which should have been deemphasized to make room for the additional material. My suggestion would have been to exclude some of the material in Chapter 3 dealing with the classical linear model, e.g. the sections on estimation and algorithms for least squares, and the whole of Chapter 9 (Models for survival data), to make room for the following.

The unified theory of regression-type modeling encompassed by generalized linear models suggests that the monograph would be an ideal text for teaching a one-semester course on the subject. Although I feel this is indeed the case, more could have been done along these lines. One simple mechanism would be through a collection of exercises at the end of each chapter. These could stress both theoretical and practical aspects of the material. The availability of the computer program GLIM would allow instructors to concentrate their teaching efforts on interpretation, rather than on computation, in the applied exercises.

None of the examples in the monograph correspond to the large observational studies which are now the rule rather than the exception. In contrast to well-defined experimental studies, the problems of model specification (especially variable selection) and goodness of fit (especially influential data) are particularly acute here. One or two examples of this sort would have provided the authors with superb motivation for some of their proposals, while forcing them to address the difficult issues which they discussed in passing.

Finally the monograph is almost totally void of recent statistical methodology of general applicability. These include jackknifing and bootstrapping to reduce bias and assess variability, nonlinear smoothing to enhance scatter plots, and monte carlo as an alternative to asymptotics. Some of these areas appear below as "Future Research", though ample opportunity existed for their introduction in the monograph:

- reducing the bias in the estimate of the dispersion parameter $\phi$;
- interpreting plots with varying density of points along the abcissa; and
- the general applicability of monte carlo methods to study distributional properties for small samples where asymptotics are not applicable.

Limited discussion of these topics as they apply to generalized linear models would have made the monograph more current without detracting from the overall theme.

**3. Future research directions.** Much of the future research on generalized linear models will be based on refining and extending methods discussed in the monograph. Other research will be aimed at extensions of state-of-the-art methodology for linear models to the class of generalized linear models. Most of this methodology is a result of advances in computing hardware and software in the past ten years. The areas listed below emphasize the importance of computing in statistics research.

*Graphical methods.* Informal graphical methods enable the modeler to detect unexpected deviations from the fitted model. Many of the author's suggestions in Chapter 11 are extensions of standard plots used for linear models. The extent to which these will be useful for generalized linear models is uncertain. Recent work by Landwehr et al. (1984) for logistic regression is encouraging but not the final word.

*Estimation.* In the past two decades, a number of alternatives to least squares have been proposed for linear models. These include ridge/shrinkage/empirical-Bayes estimators for sparse or ill-conditioned data, and robust/resistant estimators to accommodate outliers and influential observations. The problems these methods address are likely to be present in generalized linear models as well as the classical linear model, and I expect this will also be an active research area. Some progress has been made on generalizing $M$-estimators to generalized linear models (Pregibon, 1982) though more work is needed here. The quasi-likelihood formulation should prove useful in this respect as it provides a direct analogy to the normal theory $\varepsilon$-contamination model. In particular, let $E(Y) = \mu$, and

$$\operatorname{var}(Y) = \begin{cases} \phi V(\mu) & \text{with probability } 1 - \varepsilon \\ k\phi V(\mu) & \text{with probability } \varepsilon. \end{cases}$$

Robust estimates, $\tilde{\mu}$, of $\mu$ can now be considered along the lines of Huber (1981). Note in particular that there is no ambiguity in what $\tilde{\mu}$ is estimating, even when the distribution of $Y$ is asymmetric.

*Goodness of fit.* Much of the theory of generalized linear models is concerned with estimation and testing of the regression coefficients $\beta$. Formal methods of assessing goodness of fit have received less attention. This situation is unlikely to change in the future but the latter will still be a lively and important research area. McCullagh (1984a, b) has already made an important contribution by arguing that the use of the marginal distributions of the standard goodness-of-fit statistics is misguided. Instead he considers asymptotic approximations to their conditional distributions, where the conditioning is on the complete sufficient statistic for $\beta$. The theoretical derivations are tedious but result in relatively easy computing formulae. Alternatively, the conditional distributions can be approximated by Monte Carlo sampling methods. In either case, some relief from the inadequacies of current procedures is in sight.

*Further generalizations.* The class of generalized linear models encompasses a large number of useful models. The class is overly restrictive in that

(1) it presupposes a *linear* predictor $\eta = x\beta$;
(2) the class of distributions proper is restricted to the exponential family;
(3) it presupposes independent observations $\{y_i: i = 1, \cdots, n\}$; and
(4) the dispersion parameter $\phi$ is assumed constant across observations.

Further research will be concerned with eliminating these restrictions. And this research is well underway! Jorgensen (1983) presents the "extended class of generalized linear models" which allows arbitrary predictors $\eta = \eta(x, \beta)$, a wider class of distributions for $Y$: $f(y, \mu, \phi) = c(y, \phi)\exp\{a(\phi)t(y, \mu)\}$, and correlated observations ($Y$ can be vector valued). He shows that the extended class has properties similar to those of generalized linear models, including computing considerations. An (as yet) unresearched approach to relax (4) is to generalize the quasi-likelihood model according to

$$g(\mu) = x\beta, \quad \text{var}(Y) = \phi V(\mu), \quad h(\phi) = z\gamma$$

where $h$ is yet another link function and $z$ is a vector of explanatory and/or stratification variables, possibly, though not necessarily equal to $x$. For the particular case $\text{var}(Y) = \phi$, i.e. normal theory models, the above generalization was independently suggested by John Henstridge (University of Western Australia) and William H. Rogers (The Rand Corporation), both by personal communication. Otherwise the generalization is apparently new. Tests of homogeneity of $\phi$ ($H$: $\gamma = 0$) are analogous to classical homogeneity of variance tests. Fitting this model would require the notion of the "extended quasi-likelihood function" (Nelder and Pregibon, 1984)

$$Q^+(y; \mu, \phi) = -\tfrac{1}{2} \log \phi V(y) - (1/2\phi)d(y; \mu)$$

where $d(y; \mu)$ is the generalized linear model deviance function, and $V(y)$ is the variance function applied to the datum $y$.

*Bootstrapping and other cross-validation methods.* The possibility of using current data to obtain "honest" estimates of model adequacy is attractive. Computers make it a reality. Extending bootstrapping methods to generalized linear models is one area challenging both computing resources and statistical ingenuity. Since I assume an *interactive* computing environment, bootstrapping nonlinear models with quick response time will challenge our desktop computers. Since no one has yet defined identically distributed residuals for generalized linear models, a statistical challenge is to determine exactly how the bootstrap samples are to be constructed. More headway is likely to be made using the related technique of jackknifing. In particular, Wu (1984) discusses a weighted jackknife and briefly indicates how it applies to generalized linear models.

*Nonparametric smoothing.* Smoothing is a flexible mechanism for fitting linear models. Hastie (1984) has suggested a way of generalizing logistic regression so that the systematic component of the model $\text{logit}(p) = \sum x_j\beta_j$ is replaced by $\text{logit}(p) = \sum \psi_j(x_j)$ where the $\psi$'s are smooth nonparametric (and not necessarily monotone!) functions of the original covariates. This is accomplished by using the concept of "local likelihood" to smoothly estimate the form of $\psi$ for a particular $x_j$. The extension of Hastie's method to the class of generalized linear models is, in effect, a relaxing of the linearity assumption in the systematic component from $\eta = \sum x_j\beta_j$ to $\eta = \sum \psi_j(x_j)$. A further generalization would relax the link function specification in (iii) from $g(\mu) = \eta$ to $\psi_0(\mu) = \eta$ where $\psi_0$ is a smooth nonparametric function of $\mu$ perhaps restricted to be monotone. A final generalization would be to estimate the variance function in a similar fashion. The resulting generalized–generalized linear model would appear to be a very flexible class of models. Future research will tell if this generalization is *too* flexible.

*Modeling strategies.* Knowledge-based "expert" systems are computer programs designed to capture, utilize, and explicitly describe how a trained individual solves problems. Such systems have been developed in a number of domains, including chemistry, medicine, mineral exploration, and even statistics (Gale and Pregibon, 1982). The problems in building such systems in the data analysis domain are many, but an important fact of life has emerged: statisticians just haven't paid enough attention to *how* data analysis is actually done. Much attention is given to details concerning individual tests, so that under *ideal* conditions (i.e., all assumptions are satisfied except perhaps the one(s) relevant to the test in question), a "good" statistician can describe the operating characteristics of the test in more detail than anyone would care to know. The problem is that although we understand the limitations of our theories and accordingly take them with a grain of salt when actually analyzing data, we have not made this expertise and heuristic judgement available to the less statistically sophisticated. The introduction of generalized linear models onto the scene seems to accentuate the problem, as the additional leeway in model specification will tend to bewilder the nonexperienced modeler. What is needed here is not a flood of papers on specific aspects of generalized linear models, but rather careful thought on *strategies* for analyzing data in the generalized framework.

## REFERENCES

BAKER, R. J. and NELDER, J. A. (1978). *The GLIM System*, Release 3. Numerical Algorithms Group, Oxford.

COX, D. R. (1970). *The Analysis of Binary Data.* Chapman and Hall, London.

DEMPSTER, A. P. (1971). An overview of multivariate data analysis. *J. Multivariate Analysis* **1** 316–346.

GALE, W. A. and PREGIBON, D. (1982). An expert system for regression analysis. *Proc. 14th Symp. on the Interface* 110–117. Springer-Verlag, New York.

GILCHRIST, R. (1982). *GLIM82: Proceedings of the International Conference on Generalized Linear Models.* Springer-Verlag, New York.

GRIZZLE, J. E., STARMER, C. F., and KOCH, G. G. (1969). Analysis of categorical data by linear models. *Biometrics* **25** 489–504.

HASTIE, T. (1984). Comment on "Graphical methods for assessing logistic regression models" by Landwehr, Pregibon, and Shoemaker. *J. Amer. Statist. Assoc.* **79** 77–78.

HUBER, P. J. (1981). *Robust Statistics*. Wiley, New York.

JORGENSEN, B. (1983). Maximum likelihood estimation and large-sample inference for generalized linear and nonlinear models. *Biometrika* **70** 19–28.

LANDWEHR, J. M., PREGIBON, D. and SHOEMAKER, A. S. (1984). Graphical methods for assessing logistic regression models. *J. Amer. Statist. Assoc.* **79** 61–83.

MCCULLAGH, P. (1984a). On the conditional distribution of goodness-of-fit statistics for discrete data. Unpublished manuscript.

MCCULLAGH, P. (1984b). On the asymptotic distribution of Pearson's statistic in linear exponential-family models. Unpublished manuscript.

NELDER, J. A. and PREGIBON, D. (1984). Quasi-likelihood models and data analysis. *AT&T Bell Laboratories Technical Memorandum.*

NELDER, J. A. and WEDDERBURN, R. W. M. (1972). Generalized linear models. *J. Royal Statist. Soc. A* **135** 370–384.

PREGIBON, D. (1982). Resistant fits for some commonly used logistic models with medical applications. *Biometrics* **38** 485–498.

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

WU, C. F. (1984). Jackknife and bootstrap inference in regression and a class of representations for the lse. *MRC Tech. Report* **2675**. University of Wisconsin, Madison.

2C-272
AT&T BELL LABORATORIES
600 MOUNTAIN AVENUE
MURRAY HILL, NEW JERSEY 07974