Semi-parametric Generalized Linear Model Diagnostics

Brian S. Yandell and Peter J. Green, U. of Wisconsin–Madison and U. of Durham

Diagnostic tools are developed for generalized linear models in which the linear predictor is semi-parametric, linear in most of the explanatory variables but with an arbitrary functional dependence on the remaining extraneous variables. Estimation is by penalized maximum likelihood. Diagnostic tools are proposed with analogy to Pregibon (1981, 1982). Data on phone ownership in two states is analyzed in depth.

1. Introduction

We consider generalized linear models in which the linear predictor has an additive semi-parametric form, linear in most of the explanatory variables but with an arbitrary functional dependence on the remainder. Estimation of the parameters and the non-parametric curve in the model is approached by maximizing a penalized likelihood. Emphasis is placed on development of diagnostic tools along the lines of Pregibon (1981, 1982). We analyse data on phone ownership in two states kindly provided by Ed Fowlkes, AT&T Bell Laboratories.

The semi-parametric regression idea via penalty functions has been considered by several authors in varying degrees of generality; see for example Green, Jennison and Seheult (1983), Wahba (1984), Green (1985) and Engle et al. (1986). While our approach fits in a very general framework including iteratively reweighted least squares (Green, 1984) and quasi-likelihood models (Gay and Welsch, 1986; Nelder and Pregibon, 1986), we focus here upon the generalized linear model. We begin with the log-likelihood, which can be written as

$$L(\theta(\beta,\gamma)) = \sum_{i=1}^{n} z_i\theta_i - a(\theta_i) + b(z_i) \ , \tag{1}$$

with $\mathbf{z}=\{z_i\}_{i=1}^n$ the observed responses and $\theta=\{\theta_i\}_{i=1}^n$ related to the expected responses $\{\mu_i\}_{i=1}^n$ through a link function $\theta=g(\mu)$. We suppress explicit mention of $\mathbf{z}$ in notation. We replace the familiar linear predictors $\theta_i = \mathbf{x}_i^T\beta$ by the more general predictors

$$\theta_i = \mathbf{x}_i^T\beta + \gamma(\mathbf{t}_i) \ , \tag{2}$$

with β the p-vector of parameters of interest and $\mathbf{x}_i$ the corresponding explanatory variables for the ith observation. The scalar (or vector) $\mathbf{t}_i$ consists of extraneous variables, with $\gamma(\bullet)$ a function or curve whose form is not specified. For instance, we may want to model the probability of a household having a phone as a function of several socio-economic factors. While age and income affect this probability, we may not be interested in these, but need to allow for an arbitrary form for such a relation. We consider this in detail in section 4.

We cannot simply maximize the log-likelihood, as this would lead to interpolation by γ, producing an implausibly rough fit with β non-identifiable. However, if we introduce a suitable "roughness penalty", the problem of maximizing

$$L(\theta(\beta,\gamma)) - \tfrac{1}{2}\lambda J(\gamma) \tag{3}$$

is well defined. The scalar λ is a tuning constant, used to regulate the smoothness of the fitted curve γ. The penalty functional J is some numerical measure of the "roughness" of γ. This might be adopted on ad-hoc grounds, such as the integrated squared derivative which globally penalizes curvature, or it might follow from a Bayesian argument specifying a prior distribution for γ (in which λ is essentially a ratio of variances). Typically we try a range of values for λ in an exploratory fashion, as well as considering an automatic choice based on the data.

One may use this approach to discover the form of γ in the hope of modelling it parametrically in the future. However, we focus instead on inference for β in the presence of an unknown γ, and develop diagnostics for generalized residuals. The next section briefly presents the maximum penalized likelihood estimates. Section 3 introduces diagnostic tools culled from several related lines of investigation (*e.g.* Pregibon (1981, 1982) and Eubank (1984,1985)). Section 4 presents a summary of an analysis of the data on phones. For a more complete version see Yandell and Green (1986)

2. Maximum Penalized Likelihood Estimates

The maximization of (3) can be obtained by the method of iteratively reweighted least squares (Green, 1984; O'Sullivan, Yandell and Raynor, 1986). This scheme is based on the Newton-Raphson method with Fisher scoring. We present the algorithm for a chosen, fixed λ. For further algorithm details see Green and Yandell (1985) and Green (1985); see Yandell (1986) for an alternative scheme.

We first restrict attention to γ of the form $\gamma(\bullet)=\sum\xi_k\phi_k(\bullet)$, with ϕ_k, $k=1,\cdots,q$, prescribed basis functions. Thus $\gamma(\mathbf{t}_i)=\{\mathbf{E}\xi\}_i$ for an $n\times q$ matrix $\mathbf{E}$. This may limit γ to some smooth subspace, *e.g.* one spanned by B-splines, or may provide no restriction. We can write θ as

$$\theta = \mathbf{D}\beta + \mathbf{E}\xi \ ,$$

with $\mathbf{D}=(\mathbf{x}_1 \cdots \mathbf{x}_n)^T$. Based on an initial guess of θ we create pseudo-values $\mathbf{y}=\{y_i\}_{i=1}^n$,

$$\mathbf{y} = \mathbf{A}^{-1}\mathbf{u} + \theta \ , \tag{4}$$

in which

$$\mathbf{u} = \frac{\partial L}{\partial \theta} \quad \text{and} \quad \mathbf{A} = E\left[-\frac{\partial^2 L}{\partial \theta \partial \theta^T} \right] .$$

For binomial data of the kind considered here, we use $g(\mu) = log(\mu/(1-\mu)) = \theta$, with μ the probability of owning a phone. The partial matrices $\mathbf{u}$ and $\mathbf{A}$ take on simple forms, with $\mathbf{A}$ diagonal. For each case i, let $\{\mathbf{A}\}_{ii} = \mu_i(1-\mu_i)$ and $\{\mathbf{u}\}_i = z_i - \mu_i$, with $z_i = 1$ if the household has a phone, 0 otherwise, leading to pseudo-values $y_i = \theta_i + (z_i - \mu_i)/[\mu_i(1-\mu_i)]$.

The penalty $J(\gamma)$ can often be expressed in a quadratic form with some $q \times q$ symmetric $\mathbf{K}$ satisfying certain conditions (Green, 1985). Thus (3) can be approximated by a locally linearized problem which involves minimizing a quadratic form

$$(\mathbf{y} - \mathbf{D}\beta - \mathbf{E}\xi)^T\mathbf{A}(\mathbf{y} - \mathbf{D}\beta - \mathbf{E}\xi) + \lambda\, \xi^T\mathbf{K}\xi ,$$

leading to estimates for β and ξ, which are used to update θ, $\mathbf{u}$ and $\mathbf{A}$, and hence the pseudo-values (4), with iteration until convergence. The MPLE $\hat{\theta} = \theta(\hat{\beta}, \hat{\xi})$, has the form

$$\begin{aligned} \hat{\theta} &= \mathbf{D}\hat{\beta} + \mathbf{E}\hat{\xi} \\ &= [\mathbf{S} + (\mathbf{I}-\mathbf{S})\mathbf{D}\mathbf{M}_I^{-1}\mathbf{D}^T\mathbf{A}(\mathbf{I}-\mathbf{S})]\mathbf{y} \qquad (5) \\ &= \mathbf{A}^{-\frac{1}{2}}\mathbf{H}\mathbf{A}^{\frac{1}{2}}\mathbf{y} \end{aligned}$$

with $\mathbf{M}_k = \mathbf{D}^T\mathbf{A}(\mathbf{I}-\mathbf{S})^k\mathbf{D}$ and $\mathbf{S} = \mathbf{E}(\mathbf{E}^T\mathbf{A}\mathbf{E} + \lambda\mathbf{K})^{-1}\mathbf{E}^T\mathbf{A}$.

The parameter estimates depend on the choice of λ. While λ can be picked ad-hoc for visual smoothness, we often use an automatic choice based on generalized cross-validation (Craven and Wahba, 1979). Choosing λ to minimize

$$GCV(\lambda) = n\, \nu_\lambda^{-2} \, \| (\mathbf{I} - \mathbf{H}_\lambda)\mathbf{A}_\lambda^{\frac{1}{2}}\mathbf{y} \|^2 , \qquad (6)$$

with $\nu_\lambda = tr(\mathbf{I} - \mathbf{H}_\lambda)$, comes close to minimizing the predictive mean square error for linear models, and appears to serve the same purpose for generalized linear models (O'Sullivan, 1983).

3. Diagnostics

We present a variety of diagnostics in the spirit of Pregibon (1981,1982). While we cannot formally justify these at this time, they appear to be on the right track and to follow the type of generalizations to non-parametric problems developed by Eubank (1984,1985); see Yandell and Green (1986). Throughout this section the dependence on λ is suppressed in the notation.

3.1. Tests of parameters

We can test the parameters β in the same manner as Pregibon (1982) and others, using the (conjectured) asymptotic normal distribution of β. At the MPLEs we have approximately $COV(\hat{\beta}) = \mathbf{M}_I^{-1}\mathbf{M}_2\mathbf{M}_I^{-T}$. Goodness-of-fit can be assessed globally, as in generalized linear models, by the deviance $D^2(\hat{\theta}) = 2\{L(g(\mathbf{z})) - L(\hat{\theta})\}$ or by the chi-square statistic $\chi^2(\hat{\theta}) = (\mathbf{y}-\hat{\theta})^T\mathbf{A}(\mathbf{y}-\hat{\theta})$ with degrees-of-freedom approximated by $\nu = tr(\mathbf{I} - \mathbf{H})$ (Green, 1985).

We now consider testing a full model against a reduced model,

$$\theta_F = \mathbf{D}_1\beta_1 + \mathbf{D}_2\beta_2 + \mathbf{E}\xi$$
$$\theta_R = \mathbf{D}_1\beta_1 + \mathbf{E}\xi ,$$

with $\mathbf{D}_1$ and $\mathbf{D}_2$ of full ranks p and r, respectively, and $\mathbf{D} = [\mathbf{D}_1 : \mathbf{D}_2]$ of full rank $p + r$. We could use a difference of deviance statistics, or similarly with chi-square statistics. However, in general the degrees-of-freedom $\nu_R - \nu_F$ may not be near r, with possibly differing degrees of smoothness in the two models. Fixing λ the same for both models does not rectify this since the degrees-of-freedom depend on $\mathbf{S}$ and the model design, either $\mathbf{D}_1$ or $[\mathbf{D}_1 : \mathbf{D}_2]$, in a complicated way.

One natural alternative is the score test as developed by Pregibon (1982) for the parametric generalized linear model. The information matrix for the parametric piece is $\mathbf{F}_0 = \mathbf{A}(\mathbf{I} - \mathbf{S})$, with $\mathbf{A}$ and $\mathbf{S}$ evaluated at the reduced model. The score vector for the reduced model is

$$\mathbf{u} = [\mathbf{F}_0 - \mathbf{F}_0\mathbf{D}_1(\mathbf{D}_1^T\mathbf{F}_0\mathbf{D}_1)^{-1}\mathbf{D}_1^T\mathbf{F}_0]\mathbf{y} = \mathbf{F}_1\mathbf{y} .$$

Following Pregibon (1982) the score statistic can be written as

$$S^2(\hat{\theta}_R, \hat{\theta}_F^1) = \mathbf{u}^T\mathbf{D}_2(\mathbf{D}_2^T\mathbf{F}_1\mathbf{D}_2)^{-1}\mathbf{D}_2^T\mathbf{u} .$$

The score vector for the one-step (linearized) approximation to the full model is $\mathbf{F}\mathbf{y}$, with

$$\begin{aligned} \mathbf{F} &= \mathbf{F}_0 - \mathbf{F}_0\mathbf{D}(\mathbf{D}^T\mathbf{F}_0\mathbf{D})^{-1}\mathbf{D}^T\mathbf{F}_0 . \\ &= \mathbf{F}_1 - \mathbf{F}_1\mathbf{D}_2(\mathbf{D}_2^T\mathbf{F}_1\mathbf{D}_2)^{-1}\mathbf{D}_2^T\mathbf{F}_1 . \end{aligned}$$

The latter form of $\mathbf{F}$ allows us to express the score statistic as

$$S^2(\hat{\theta}_R, \hat{\theta}_F^1) = \mathbf{y}^T\mathbf{F}_1\mathbf{y} - \mathbf{y}^T\mathbf{F}\mathbf{y} ,$$

which is easily computed. The score test for the semi-parametric generalized linear model does not in general reduce to a difference in chi-square statistics as in Pregibon (1982) since $(\mathbf{I} - \mathbf{S})$ is not a projection matrix.

3.2. Generalized Residuals

Following Pregibon (1981) and Eubank (1984, 1985) we propose examining constructs of generalized residuals. The chi-square and deviance residuals are, respectively,

$$r_i = a_{ii}^{\frac{1}{2}}(y_i - \hat{\theta}_i) \quad \text{and} \quad d_i = \pm 2^{\frac{1}{2}}(L(g(z_i)) - L(\hat{\theta}_i)) ,$$

in which $L(\theta_i)$ is the ith term of the log-likelihood (1).

The matrix $\mathbf{H}$ of (5) is the "hat matrix", with diagonal elements h_{ii} being the leverage values. Eubank (1984) showed that for the non-parametric linear (normal) model, $\mathbf{H}$ shares many of the properties of the least-squares "hat matrix". Pregibon (1981) identified this matrix for the parametric generalized linear model. The chi-square residuals $\mathbf{r}$ have covariance $\phi(\mathbf{I}-\mathbf{H})$, suggesting the use of standardized residuals

$$\tau_i = r_i \phi^{-1/2}(1-h_{ii})^{-1/2}$$

for plotting purposes. The dispersion ϕ is commonly estimated by $D^2(\hat{\theta})/\nu$. If we set $\phi=1$ we have the square root of the decrease in chi-square due to deleting the i th observation. τ_i^2 is a first-order approximation to the chi-square "goodness-of-fit sensitivity" (Pregibon, 1981). A slightly more complicated expression arises for the deviance goodness-of-fit sensitivity, as $d_i^2 + h_{ii}\tau_i^2$.

Cross-validated residuals arise from fitting the model using all points but the point of interest, and may provide a more accurate measure of the fit at each point (Craven and Wahba, 1979). The cross-validated (CV) estimate of θ_i with the i th observation removed is, to first order,

$$\hat{\theta}_{(i)} = (\hat{\theta}_i + h_{ii} y_i) / (1-h_{ii}) \ . \qquad (7)$$

This can be used to define CV residuals (Pregibon's "coefficient sensitivity") and standardized CV residuals.

Influence measures can be developed to compare the fit with and without a particular point. One may also be interested in the effect of an influential point on fits at neighboring points. Yandell and Green (1986) develop these using the ideas of Pregibon (1981) and Eubank (1985).

4. Data Analysis of Household Phone Ownership

We present an exploratory analysis of data on household phone use kindly provided by Edward Fowlkes, AT&T Bell Laboratories. The data comes from 2134 households, 1810 in Texas and 324 in Missouri, gathered from the 1980 census by the National Economic Research Association (NERA). Of these, 1605 households in Texas and 300 in Missouri had phones. Data were collected on several socio-economic and family factors, as well as phone cost. Table 1 contains the principal factors which were ultimately selected. Other factors include use and installation rates, urban/rural, line density in the area, and some other family and language factors. It was suggested that we investigate the probability of having a phone as it relates to *income*, *age* and the other factors. Initial examination of histograms suggested that only *income* needed transformation for symmetry. Taking the *4th* root made *income* fairly normal, and we called the resultant new variable "*income4*".

Table 1. Data Description and Score Statistics

name	range	description	score
phone	0,1	1=household has phone	
age	17-90	in years (household head)	-
income	.145-70	in thousands of dollars	119.96
educ	0-20	education in years	57.36
black	0,1	1=black, 0=non-black	21.21
nonf	0,1	1=non-family household	14.11
nonff	0,1	1=nonf with female head	25.74
olang	0,1	1=English is not primary	11.19
pch1	0-6	number of young children	8.48
spff	0,1	single person female family	2.89
...		(10 variables not selected)	(2.18)

We considered semi-parametric logistic regressions of phone use, with *age* and/or *income4* being non-parametric and other socio-economic factors entering in a parametric fashion. These were chosen primarily because they were the most "continuous" factors, most others taking on only a few possible values. We present details only on the *age* model here. We took a step-wise forward selection approach to adding variables one at a time. We considered as criteria the largest reduction in GCV, the most significant parameter and the largest score statistic for adding a single variable, leading to essentially the same sequence. The semi-parametric "age model" entered 8 variables at a 90% confidence level, or 7 at a 95% or 99% level, based on the score statistics with an (assumed) asymptotic χ_1^2 distribution (see Table 1). The semi-parametric "income4 model" allowed 10 variables at a 90% level, 9 at a 95% level, and only 6 at a 99% level. These variabels were the same as for the *age* model, with the addition of counts of young and mid-age children. It is reassuring that the same variables emerged in roughly the same signed order. *Age* was the first variable to be added into the "income4 model", while *income4* entered after *educ* in the "age model". This is not surprising when one views the non-parametric curves (Figure 1). The non-parametric *income4* curve γ_{inc} is very smooth, but "J" shaped, while the non-parametric γ_{age} is very rough but tracks a near-straight line except at the extremes. The roughness of γ_{age} may be due to undersmoothing by the linearized GCV criterion.

Figure 1a. nonparametric age

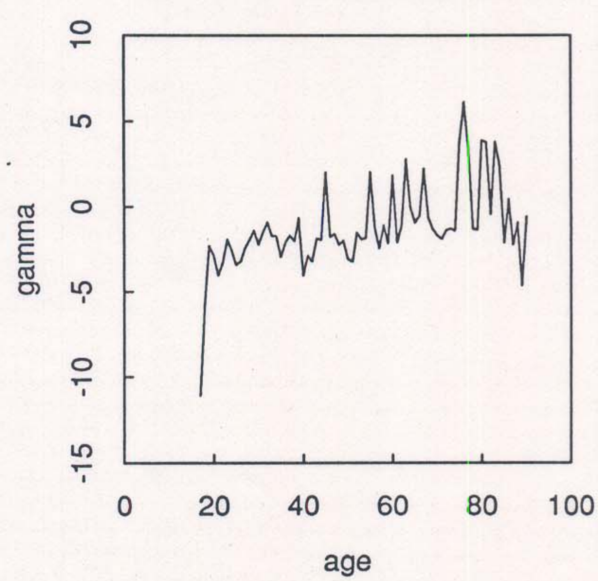
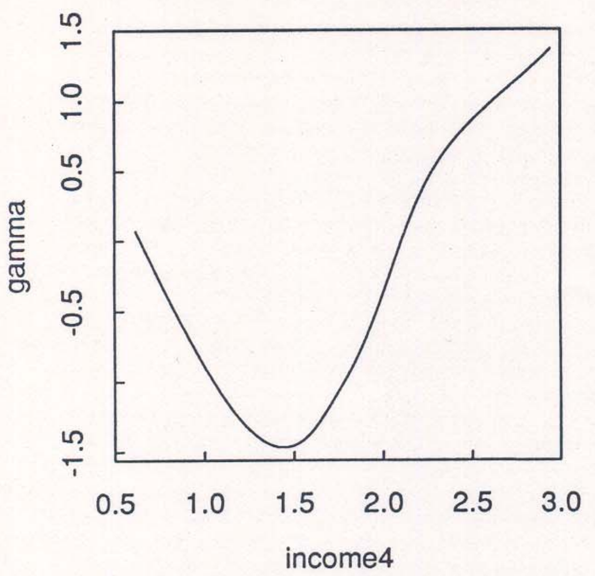


Figure 1b. nonparametric income4



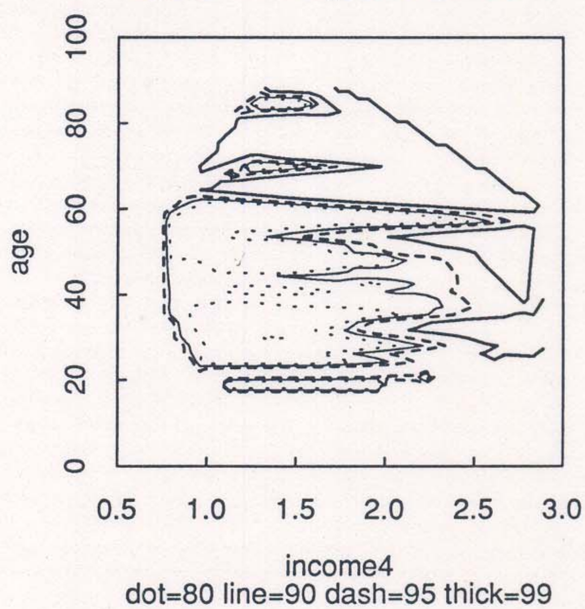
The model fits were achieved to minimize the GCV, with deviance used instead of the residual sum of squares (chi-square) in (6). The GCV decreased with increasing number of model parameters, as expected (see Table 2, *age* model only). However, while the deviance decreased in a similar manner, the chi-square did not. This may be due in part to using the deviance instead of the chi-square in (6), and requires further study. However, there is no guarantee that either the deviance or chi-square should decrease, as optimization, *i.e. the choice of* λ, is based on the GCV. The degrees-of-freedom difference between models is rarely close to 1; in fact it increased for the "income4 model" when *age* was added, which could possibly be explained by the increase in λ. This argues against a naive comparison of model fits without controling the tuning constant.

Table 2. Model Fit Statistics for AGE model

model	dev	chi	df	$lambda$	GCV
null	1454.8	2134.0	2133	-	0.682
0	1272.0	1858.5	2065.8	0.00238	0.636
1 +educ	1155.4	1909.6	2065.1	0.00251	0.578
2 +income4	1098.0	1697.2	2063.7	0.00177	0.550
3 +black	1079.1	1762.2	2062.6	0.00162	0.541
4 +nonf	1065.6	1658.3	2061.9	0.00181	0.535
5 +nonff	1038.9	1722.2	2060.5	0.00146	0.522
6 +olang	1028.1	1651.0	2059.5	0.00137	0.517
7 +pch1	1020.0	1632.0	2058.5	0.00142	0.514
8 +spff	1017.0	1655.3	2057.5	0.00139	0.513

The interesting patterns with *age* and *income4* prompted us to perform a 2-D non-parametric regression of logits on *income4* and *age*. Due to limitations of available computer hardware and software, we were only able to perform a 2-D non-parametric regression on a randomly chosen subset of 200 cases. For this subset, the null deviance and chi-square were 84.54 and 200.0, respectively, while the fitted deviance and chi-square were 64.79 and 47.13, respectively, with $\nu = 154.85$ and $\lambda = .009111$. Figure 2 suggests an interaction between *income4* and *age*. The γ_{inc} of Figure 1b appears to be valid for ages 21-55 or so, with a more dramatic relation appropriate for people over 65. In fact, 93% of the people 65 and over have phones, regardless of income, while only 88% between 21 and 59 do, with income being an important factor. It would be interesting to allow a different degree of smoothing in *age* and *income4*, or to analyse subsets by age in addition to running a full 2-D semi-parametric model.

Figure 2. nonparametric 2-D



4.1. Diagnostics

We focus on diagnostics for the "full" semi-parametric age model to illustrate some graphical diagnostics. Several leverage values were over .3 (Figure 3). Most of these were for household heads over age 80, with a few under age 20. This could reflect the *age* and *income4* pattern of phone use discussed earlier.

The different types of residuals gave essentially the same scatter plots. We present only the standardized CV deviance residuals $\tau_{(i)}$ in Figure 4. Goodness-of-fit sensitivity appears to be a useful tool to ferret out points which have a large influence on the deviance (Figure 5). None of the cases seem to have a great influence on the fit, and are not presented here.

We now highlight several of the points which stand out via the diagnostics. Case 1106 had the highest leverage, while cases 1305 and 1679 had large negative deviances and high goodness-of-fit sensitivity. These are near the extremes of education level, have low incomes, and are single-persons. The high leverage case is male with no education, while the other two are female with 17 years

Figure 3. age leverage values

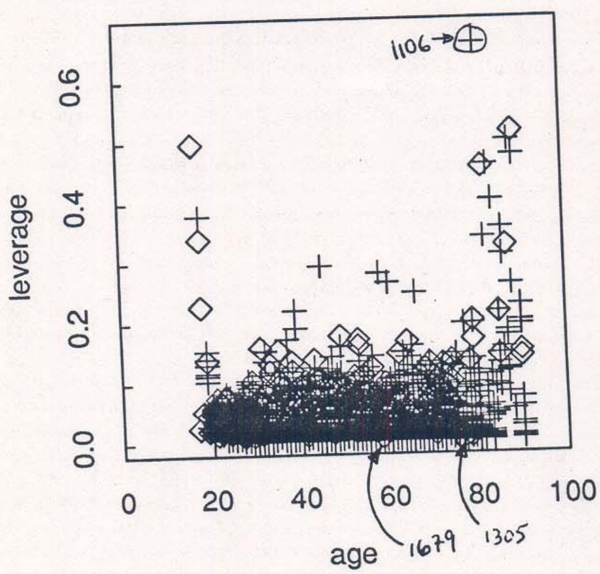


Figure 4. age CV deviance residuals

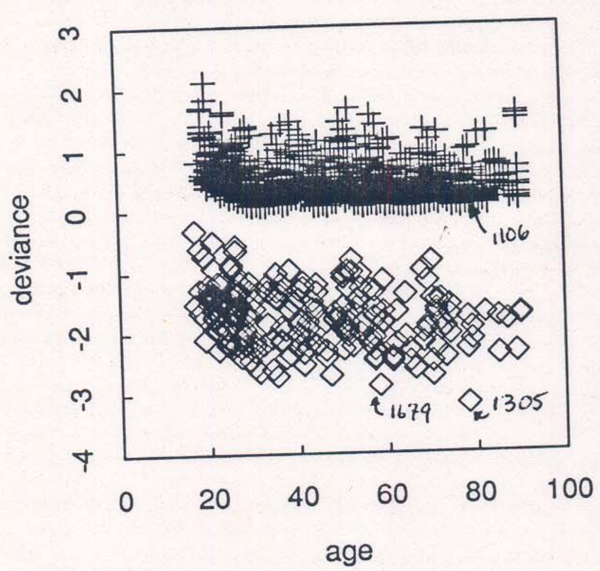
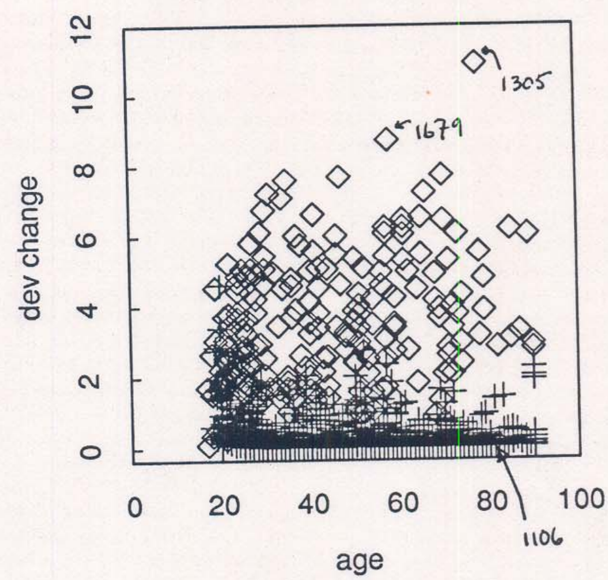


Figure 5. age fit sensitivity



education. For the age model, the point with high leverage has a small deviance, and the points with large deviance and sensitivity have small leverage.

5. Conclusions and Future Work

We have presented a collection of tools for semi-parametric generalized linear models and have demonstrated their value in understanding the preference for household phones. The picture that emerges is not simple and illustrates some of the flexibility of these tools. It appears that the diagnostics can pick out interesting cases, although the statistical properties of these diagnostic tools remain to be determined.

Future work will follow several lines. First, we wish to ascertain the properties of these diagnostic tools. Second, we would like to further investigate this data set, fitting the 2-D, and possibly 3-D (with education) semi-parametric models. We also would like to explore fitting of subsets of the data to substantiate the patterns observed. It appears that a separate analyses may be fruitful for the lower income group, the lower income 20-60 age group, and for those over 65. It became clear during analysis that the computational tools are quite handy, but are slow for large problems. We plan to address this by trying to improve some algorithms, and by obtaining access to a supercomputer for larger analyses.

Acknowledgements

This research has been supported in part for Yandell by United States Department of Agriculture CSRS grant 511-100 and National Sciences Foundation grant DMS-8404970, and for Green while visiting the Mathematics Research Center under United States Army contract DAAG29-80-C-0041. Computing was performed on the UW-Madison Statistics VAX 11/750 Research Computer and the Durham Amdahl 470/V8.

References

Craven, P. and Wahba, G. (1979) Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation. *Numer. Math.*, **31**, 377-403.

Engle, R. F., Granger, C. W. J., Rice, J. and Weiss, A. (1986) Semiparametric estimates of the relation between weather and electricity sales. *J. Amer. Statist. Assoc.*, **81**, 310-320.

Eubank, R. L. (1984) The hat matrix for smoothing splines. *Statist. and Prob. Letters*, **2**, 9-14.

Eubank, R. L. (1985) Diagnostics for smoothing splines. *J. Roy. Statist. Soc. Ser. B*, **47**, 332-341.

Gay, D. M. and Welsch, R. E. (1986) Maximum quasi-likelihood and maximum likelihood for exponential family models: algorithms and diagnostics. In *Proceedings of the 18th Symposium on the Interface, 19-21 March 1986* (T. Boardman, ed.), pp.277-284. Fort Collins.

Green, P. J. (1984) Iteratively reweighted least squares for maximum likelihood estimation and some robust and resistant alternatives (with discussion). *J. Roy. Statist. Soc. Ser. B*, **46**, 149-192.

Green, P. J. (1985) Penalized likelihood for general semi-parametric regression models. Technical Report#2819, Math. Research Center, U. of Wisconsin.

Green, P. J., Jennison, C. and Seheult, A. H. (1983) Contribution to the discussion of the paper by Wilkinson et al. *J. Roy. Statist. Soc. Ser. B*, **45**, 193-195.

Green, P. J. and Yandell, B. S. (1985) Semi-parametric generalized linear models. In *GLIM85: Proceedings of the International Conference on Generalized Linear Models, September 1985* (R. Gilchrist, ed.) Lecture Notes in Statistics, Springer-Verlag. (Technical Report#2847, Math. Res. Cen., U. of Wisconsin)

Nelder, J. A. and Pregibon, D. (1986) An extended quasi-likelihood function. Unpublished manuscript.

O'Sullivan, F. (1983) The analysis of some penalized likelihood schemes. Technical Report#726, Dept. of Statistics, U. of Wisconsin.

Pregibon, D. (1981) Logistic regression diagnostics. *Ann. Statist.*, **9**, 705-724.

Pregibon, D. (1982) Score tests in GLIM. In *Proc. GLIM82 Conf.* (R. Gilchrist, ed.) New York: Springer-Verlag.

Wahba, G. (1984) Cross validated spline methods for the estimation of multivariate functions from data on functionals. In *Statistics: An Appraisal, Proceedings 50th Anniversary Conference Iowa State Statistical Laboratory* (H. A. David,, ed.) Iowa State U. Press.

Yandell, B. S. and Green, P. J. (1986) Semi-parametric generalized linear model diagnostics. Technical Report, Dept. of Statistics, U. of Wisconsin.

Yandell, B. S. (1986) Algorithms for nonlinear generalized cross-validation. In *Proceedings of the 18th Symposium on the Interface, 19-21 March 1986* (T. Boardman, ed.), pp.450-455. Fort Collins.