Who owns Phones: Diagnosis using Semi-parametric GLMs

by

Brian S. Yandell and Peter J. Green

University of Wisconsin-Madison and University 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 the problem of determining what factors are associated with the chance of owning a phone. As in many problems, we may have a partial idea about a logistic model describing associations, but some factors may not fit into a nice parametric form such as a quadratic or sine curve. We allow for this by examining 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 Rice (1981), Green, Jennison and Seheult (1983), Wahba (1984), Green (1985), Green and Yandell (1985) and Wahba (1986). Another direction has been taken by Hastie and Tibshirani (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 upon the generalized linear model (GLM). We begin with the log-likelihood, which can be written as

$$L(\theta(\beta,\gamma)) = \sum_{i=1}^{n} [z_i \theta_i - a(\theta_i) + b(z_i)]/\sigma^2 \ , \qquad (1)$$

with $\mathbf{z} = \{z_i\}_{i=1}^n$ the observed responses, σ^2 a measure of spread (possibly 1), and a and b known functions (Nelder and Wedderburn (1972); McCullagh and Nelder (1983)). The parameters $\theta = \{\theta_i\}_{i=1}^n$ are related to the expected responses $\mu = \{\mu_i\}_{i=1}^n$ 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

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

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

$$n^{-1}\sigma^2 L(\theta(\beta,\gamma)) - \tfrac{1}{2}\lambda J(\gamma) \qquad (3)$$

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. Wahba's generalized cross-validation (GCV) method (Wahba, 1977) uses an invariant modification of a predictive mean-squared error criterion to choose λ. Other approaches to the automatic choice of λ would be possible, for example the empirical Bayesian methods proposed by Leonard (1982).

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 and known properties. 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 detailed analysis of the data on phones.

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 we can write $\gamma(\mathbf{t}_i)=\{\mathbf{E}\,\xi\}_i$ for an $n\times q$ matrix $\mathbf{E}$. This may limit γ to some smooth subspace, *e.g.* one spanned by B-splines (see Green and Yandell (1985)), or may provide no restriction. We can write θ as

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

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

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

with partials evaluated at $\hat\theta$ being the score vector and matrix of pseudo-weights, respectively,

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

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, suppose there are m_i individuals, with z_i owning phones. Then $\{\mathbf{A}\}_{ii}=m_i\mu_i(1-\mu_i)$ and $\{\mathbf{u}\}_i=z_i-m_i\mu_i$, leading to pseudo-values $y_i=\theta_i+(z_i-m_i\mu_i)/[m_i\mu_i(1-\mu_i)]$.

The penalty can often be written in a quadratic form, for some $q\times q$ symmetric $\mathbf{K}$ satisfying certain conditions outlined below. Thus the problem of maximizing (3) can be approximated by a locally linearized problem which involves minimizing a quadratic form

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

leading to estimates for β and ξ. These estimates are used to update θ, $\mathbf{u}$ and $\mathbf{A}$, and hence the pseudo-values (4), with iteration until convergence. The MPLEs have the form

$$\hat\beta=\mathbf{M}_1^{-1}\mathbf{D}^{\mathrm{T}}\mathbf{A}(\mathbf{I}-\mathbf{S})\mathbf{y} \ ,$$
$$\mathbf{E}\hat\xi=\mathbf{S}\,(\mathbf{y}-\mathbf{D}\beta) \ , \qquad (5)$$

with $\mathbf{M}_k=\mathbf{D}^{\mathrm{T}}\mathbf{A}(\mathbf{I}-\mathbf{S})^k\mathbf{D}$, $k=1,2$. The matrix $\mathbf{S}=\mathbf{E}(\mathbf{E}^{\mathrm{T}}\mathbf{A}\mathbf{E}+\lambda\mathbf{K})^{-1}\mathbf{E}^{\mathrm{T}}\mathbf{A}$ is sometimes referred to as the "smoother". The MPLE of $\hat\theta=\theta(\hat\beta,\hat\xi)$ has the form

$$\begin{aligned}\hat\theta&=\mathbf{D}\hat\beta+\mathbf{E}\hat\xi\\&=[\mathbf{S}+(\mathbf{I}-\mathbf{S})\mathbf{D}\mathbf{M}_1^{-1}\mathbf{D}^{\mathrm{T}}\mathbf{A}(\mathbf{I}-\mathbf{S})]\mathbf{y} \qquad (6)\\&=\mathbf{A}_\lambda^{-1/2}\mathbf{H}_\lambda\mathbf{A}_\lambda^{1/2}\mathbf{y} \ .\end{aligned}$$

Note especially in the last expression that all terms may depend on β as well as on λ.

We need a few conditions to ensure that this problem is well-defined (Green, 1985). The matrices $\mathbf{D}$ and $\mathbf{E}$ must be of full rank p and q, respectively. Decompose $\mathbf{K}$ as $\mathbf{L}^{\mathrm{T}}\mathbf{L}$, where $\mathbf{L}$ is $r\times q$ of full rank r. If $q=r$, we need $[\mathbf{D}:\mathbf{E}\mathbf{L}^{-1}]$ of rank $p+q$. If $q<r$, as is often the case, then there is a $q\times(q-r)$ matrix $\mathbf{T}$ such that $\mathbf{L}\mathbf{T}=\mathbf{0}$ and $[\mathbf{L}^{\mathrm{T}}:\mathbf{T}]$ is non-singular. We then need $[\mathbf{D}:\mathbf{E}\mathbf{T}]$ of rank $p+q-r$.

At the MPLEs we have approximate, nominal covariances for the pseudo-values, $COV(\mathbf{y})=\mathbf{A}^{-1}$, and for the parameters, $COV(\hat\beta)=\mathbf{M}_1^{-1}\mathbf{M}_2\mathbf{M}_1^{-\mathrm{T}}$. Goodness-of-fit can be assessed globally, as in generalized linear models, by the deviance

$$\begin{aligned}D^2(\hat\theta)&=2\{sup_\theta L(\theta)-L(\hat\theta)\} \ ,\\&=2\{L(g(\mathbf{z}))-L(\hat\theta)\} \ ,\end{aligned}$$

or by the chi-square statistic

$$\chi^2(\hat\theta)=(\mathbf{y}-\hat\theta)^{\mathrm{T}}\mathbf{A}(\mathbf{y}-\hat\theta) \ .$$

For the binomial model, these are simply

$$\chi^2(\theta)=\sum_{i=1}^n(z_i-m_i\mu_i)^2/[m_i\mu_i(1-\mu_i)] \ ,$$

$$D^2(\theta)=2\sum_{i=1}^n z_i\log(z_i/(m_i\mu_i))+(m_i-z_i)\log((1-z_i/m_i)/(1\cdot$$

The degrees-of-freedom ν, generally not an integer, can be approximated by (Green, 1985)

$$\begin{aligned}\nu&=n-tr(\mathbf{S})-tr(\mathbf{M}_1^{-1}\mathbf{M}_2) \ , \qquad (7)\\&=n-p-tr(\mathbf{S})+tr((\mathbf{D}^{\mathrm{T}}\mathbf{A}(\mathbf{I}-\mathbf{S})\mathbf{D})^{-1}\mathbf{D}^{\mathrm{T}}\mathbf{A}(\mathbf{I}-\mathbf{S})\mathbf{S}\mathbf{D})\end{aligned}$$

which approximates the asymptotic expectation of $D^2(\hat\theta)$ and reduces to the usual $n-p$ when the non-parametric part of the model is omitted. This ν has been used informally for linear models (Eubank, 1985; Eubank, 1984) and generalized linear models (O'Sullivan, Yandell and Raynor, 1986; O'Sullivan, 1985; Green and Yandell, 1985; Yandell, 1985, 1986).

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 (Wahba, 1977). Choosing λ to minimize

$$V(\lambda) = n\, \nu_\lambda^{-2}\, \| (\mathbf{I}-\mathbf{H}_\lambda)\mathbf{A}_\lambda^{1/2}\mathbf{y} \|^2 = n\, \chi^2(\hat{\theta}_\lambda) / \nu_\lambda^2 \quad (8)$$

comes close to minimizing the predictive mean square error for linear models (Craven and Wahba, 1979; Speckman, 1985), and appears to serve the same purpose for generalized linear models (O'Sullivan, 1983). The argument has been made by several to use instead a variant based on the likelihood,

$$V_L(\lambda) = n\, D^2(\hat{\theta}_\lambda) / \nu_\lambda^2 \ . \quad (9)$$

3. Theoretical Questions

We can show that the penalized likelihood estimate of θ is its posterior mean, at least to the first two moments. Its distribution is approximately normal under certain situations similar to those for ordinary GLMs. In the process we show that the pseudo-values are approximately normal, which in part justifies the diagnostics proposed in the next section.

We first review results for the linear model, in which $\mathbf{z}=\mathbf{y}$, $\theta=\mu$ and $\mathbf{A}=\mathbf{I}$. This is commonly called the partial spline model (Rice (1981); Engle et al. (1986); Green, Jennison and Seheult (1983); Wahba (1984)).

Assume that there is a reproducing kernel Hilbert space H_Q of real valued functions of $\mathbf{x} \in I$ (*e.g.* $I = E^d$), and that H_Q has an M-dimensional subspace H_o spanned by $\{\phi_k\}_{k=1}^M$. One may have $\theta_i = L_i\theta$, with $\theta \in H_Q$ and $\{L_i\}_{i=1}^n$ bounded linear functionals on H_Q (Wahba, 1985), but we leave this out below. Suppose that γ follows a Bayesian model,

$$\gamma(\mathbf{x}) \sim \sum_{k=1}^{M} \alpha_k \phi_k(\mathbf{x}) + \sigma(n\lambda)^{-1/2} Z(\mathbf{x})$$

with $\{Z(\mathbf{x}), \mathbf{x} \in I\}$ a family of zero-mean Gaussian random variables with prior covariance

$$EZ(\mathbf{x})Z(\mathbf{x}') = Q_1(\mathbf{x}, \mathbf{x}') \ ,$$

where Q_1 is the reproducing kernel for H_{Q_1}, the orthocomplement of H_o .

In the case of no parametric piece ($\beta = \mathbf{0}$), we know that $\hat{\theta} = \hat{\gamma}$ is the best linear unbiased predictor for the model

$$z(\mathbf{x}) = \gamma(\mathbf{x}) + \sigma\varepsilon(\mathbf{x})$$

with $\varepsilon(\bullet)$ a zero-mean unit variance *iid* process independent of γ (Kimeldorf and Wahba, 1970). The residuals $\mathbf{z}-\hat{\theta}$ have covariance $\sigma^2(\mathbf{I}-\mathbf{H})$ (Wahba, 1978). Under a Bayesian prior for $\alpha = \{\alpha_j\}_{j=0}^{m-1}$, $\hat{\theta}$ is the posterior mean of θ, having covariance $\sigma^2\mathbf{H}$ (Wahba, 1983); see also Silverman (1985) and Wahba (1985). Cox (1983) developed strong approximations for the spline process by a Gaussian process under mild conditions, allowing for a non-normal but homoscedastic model. Speckman (1985) improved earlier results of Craven and Wahba (1979), showing that the automatic choice of λ via generalized cross-validation asymptotically minimizes the predictive mean square error. Recently, Cox and Koh (1986) developed a test for adequacy of the polynomial part of γ, that is, whether one could reduce the problem to a parametric problem.

For the semi-parametric linear, or partial spline, model, Heckman (1986b) considered the explanatory variables $\{\mathbf{x}_{in}\}_{i=1}^n$ in (2) as *iid* random variables from a zero-mean distribution with covariance Σ and finite fourth moments. Heckman showed that $\hat{\beta}$ is consistent provided that ε_i have uniformly bounded third moments and $n\lambda^{1/2m} \to \infty$ as $n \to \infty$. If in addition either $\lambda \to 0$ or $\gamma^{(m)}(\bullet) \equiv 0$, then

$$n^{1/2}(\hat{\beta}-\beta) \xrightarrow{\mathbf{D}} N(\mathbf{0}, \sigma^2\Sigma^{-1}) \ .$$

Heckman (1986b) also noted under a diffuse Bayesian prior as in Wahba (1978), $\hat{\beta}$ and $\hat{\theta}$ are the respective posterior means. Cox et al. (1986) extended the results of Cox and Koh (1986) to Heckman's model to test the adequacy of the parametric null model in the semi-parametric linear model.

Rice (1986) considered a variation of the model used by Heckman (1986b) in which the univariate explanatory x_i has the form $x_{in} = x_{in}^* + h(t_{in})$ with x_{in}^* as in Heckman (1986b), t_{in} "regular" and $h(\bullet)$ continuous. Rice showed that the variance is $O(n^{-1})$, as hoped, provided $\lambda n^{2m/(2m+1)-\delta} \to \infty$ and (roughly speaking) $\int h^2 < \infty$. However, the bias is $o(n^{-1/2})$ only if h is a polynomial of degree at most $m-1$. If h is of higher order, then the bias can be $O(\lambda)$, that is, decreasing at a non-parametric rate. Heckman (1986a) has found a way to circumvent this difficulty using a minimax procedure.

Wahba (1986) sets the partial spline model in a general context and develops the interaction spline model of Barry (1986). Chen (1986) provides theoretical justification for the interaction spline model, which allows one formally to test nested models with splines along the lines of additive models (Stone (1985); Hastie and Tibshirani (1986)).

In the parametric GLM, $\hat{\beta}$ is asymptotically normal, and the chi-square and deviance statistics are asymptotically χ^2 under appropriate conditions (McCullagh (1983); Jørgensen (1986b)). Nelder and Pregibon (1986) proposed an extended quasi-likelihood, which is asymptotically χ^2 when it corresponds to an exponential family distribution (Efron (1985); Jørgensen (1986ab, 1987)).

O'Sullivan (1983) showed for the non-parametric GLM that if (3) is convex and there is a unique MLE for (1), then there is a unique MPLE. O'Sullivan also generalized the result of Craven and Wahba (1979) on the automatic choice of λ.

For semi-parametric GLMs, we can show following McCullagh and Nelder (1983) that

$$E\mathbf{u} = \mathbf{0} \quad \text{and} \quad E\mathbf{u}\mathbf{u}^{\mathrm{T}} = \sigma^2 \mathbf{A} ,$$

that is, $\mathbf{u} \sim (\mathbf{0}, \sigma^2 \mathbf{A})$. In most cases, $\mathbf{u}$ is not normal. However, suppose n is fixed and that for each i, z_i tends to a normal distribution as $m_i \to \infty$. This occurs for the binomial model we mentioned earlier. Then if $min\{m_i\}_{i=1}^{n} \to \infty$, $\mathbf{u}$ is asymptotically normal. This follows from standard small dispersion arguments for GLMs (Jørgensen, 1986b; McCullagh, 1983).

Let $\mathbf{S}$ be an $n \times p$ matrix with $\mathbf{S}_{ij} = \psi_j(\xi : \mathbf{s}_i)$, and let $\mathbf{T}$ be $n \times M$ with $\mathbf{T}_{ik} = \phi_k(\xi)$. For convenience let $\mathbf{X} = [\mathbf{T} : \mathbf{S}]$ and $\delta^{\mathrm{T}} = (\alpha^{\mathrm{T}} : \beta^{\mathrm{T}})$. Let Σ be $n \times n$ with entries

$$E\gamma(\xi)\gamma(\mathbf{x}_j) = Q_1(\xi, \mathbf{x}_j) \ .$$

Thus if we assume δ is fixed,

$$\mathbf{A}^{1/2}\mathbf{y} \sim (\mathbf{A}^{1/2}\mathbf{X}\delta, b\,\mathbf{A}^{1/2}\Sigma\mathbf{A}^{1/2} + \sigma^2\mathbf{I}) \qquad (10)$$

or, if we adopt the improper prior $\delta \sim N(\mathbf{0}, \xi\mathbf{I})$ with $\xi \to \infty$,

$$\mathbf{A}^{1/2}\mathbf{y} \sim (\mathbf{0}, \xi\mathbf{A}^{1/2}\mathbf{X}\mathbf{X}^{\mathrm{T}}\mathbf{A}^{1/2} + b\,\mathbf{A}^{1/2}\Sigma\mathbf{A}^{1/2} + \sigma^2\mathbf{I}) \ . \quad (11)$$

For the normal model, $\mathbf{y}$ is asymptotically normal. Otherwise, we can use the arguments for $\mathbf{u}$ given earlier to yield asymptotic normality for the $\mathbf{y}$. Currently Dennis Cox and Yandell are extending the Cox et al. (1986) test of the parametic null hypothesis to semi-parametric GLMs.

4. Diagnostics

We present a variety of diagnostics in the spirit of Pregibon (1981,1982). While we have not formally justified these, they appear to be on the right track and to follow the type of generalizations to non-parametric problems developed by Eubank (1984,1985). It appears from work of Gay and Welsch (1986) that one could extend these tools to semi-parametric models arising from iteratively reweighted least squares or quasi-likelihood in an analogous manner, although we do not attempt this here. Throughout this section the dependence on λ is suppressed in the notation.

4.1. 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 to posit that

$$(\hat{\beta} - \beta_o)^{\mathrm{T}}\mathbf{M}_1^{\mathrm{T}}\mathbf{M}_2^{-1}\mathbf{M}_1(\hat{\beta} - \beta_o)$$

has approximately χ_p^2. This can be used to "test" for removal of parameters from a model, or for entering a new parameter in the usual fashion.

We now consider testing a full model $\theta_F = \mathbf{D}_1\beta_1 + \mathbf{D}_2\beta_2 + \mathbf{E}\xi$ against a reduced model $\theta_R = \mathbf{D}_1\beta_1 + \mathbf{E}\xi$, 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 deviance statistic

$$D^2(\hat{\theta}_R, \hat{\theta}_F) = D^2(\hat{\theta}_R) - D^2(\hat{\theta}_F) \ ,$$

or similarly use a chi-square statistic. If $q \ll n$, we are practically in a parametric model, with $\nu_R \approx n - p - q$ and $\nu_F \approx n - p - r - q$. 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. If the reduced model is correct, then even keeping the same tuning constant λ for the full and reduced model fits would not rectify this. A fixed λ locally fixes $\mathbf{A}$ (near $\hat{\theta}_R$), which in turn fixes the "smoother" $\mathbf{S}$. However, the degrees-of-freedom in (7) depends 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

$$\begin{aligned}\mathbf{u} &= \mathbf{A}(\mathbf{y} - \hat{\theta}_R) \\ &= [\mathbf{F}_0 - \mathbf{F}_0\mathbf{D}_1(\mathbf{D}_1^{\mathrm{T}}\mathbf{F}_0\mathbf{D}_1)^{-1}\mathbf{D}_1^{\mathrm{T}}\mathbf{F}_0]\mathbf{y} \\ &= \mathbf{F}_1\mathbf{y} \ .\end{aligned}$$

Following Pregibon (1982) the score statistic

$$S^2(\hat{\theta}_R, \hat{\theta}_F^1) = \mathbf{u}^{\mathrm{T}}\mathbf{D}(\mathbf{D}^{\mathrm{T}}\mathbf{F}_0\mathbf{D})^{-1}\mathbf{D}^{\mathrm{T}}\mathbf{u}$$

can be written as

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

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

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

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

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

with $\hat{\theta}_F^1 = (\mathbf{I} - \mathbf{A}^{-1}\mathbf{F})\mathbf{y}$ being the one-step estimate of θ_F. 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). [This can be seen by examination, noting that $(\mathbf{I}-\mathbf{S}) \neq (\mathbf{I}-\mathbf{S})^2$.] However, it has a very simple form and is easily computed.

4.2. Generalized Residuals

Following Pregibon (1981) and Eubank (1984, 1985), we propose examining diagnostics based on generalized chi-square and deviance residuals, *i.e.*,

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

in which $L(\theta_i)$ is the i th term of the log-likelihood (1). The matrix $\mathbf{H}$ of (6) 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 GLM. One can investigate the leverage values directly, or use them in conjunction with the ordinary residuals to develop cross-validated residuals. For instance, standardized chi-square residuals

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

can be plotted against i or one of the predictor variables. The dispersion σ^2 is commonly estimated by $D^2(\hat{\theta})/\nu$. With $\sigma^2 = 1$, τ_i^2 is approximately the decrease in $\chi^2(\hat{\theta})$ due to deleting the i th observation, which is Pregibon's "goodness-of-fit sensitivity". The deviance goodness-of-fit sensitivity is approximated by $d_i^2 + h_{ii}\tau_i^2$. Pregibon (1981) suggested a normal probability plot of the square root of this, or plotting d_i^2 against $h_{ii}\tau_i^2$.

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); Eubank (1985)). The cross-validated (CV) estimate of θ_i is to first order

$$\hat{\theta}_{(i)} = (\hat{\theta}_i + h_{ii}y_i) / (1 - h_{ii}) \ . \tag{12}$$

This can be used to construct CV residuals $r_{(i)} = r_i / (1 - h_{ii})$ which are equivalent to Pregibon's (1981) "coefficient sensitivity". For plotting, one may prefer the standardized CV residual $\tau_{(i)} = r_{(i)}\sigma_{(i)}^{-1}(1 - h_{ii})^{-1/2}$, with CV dispersion $\sigma_{(i)}^2$ given in (13) below. Eubank (1985) showed through Monte Carlo studies that, for the linear non-parametric model, τ_i and $\tau_{(i)}$ have the right level for the $N(0,1)$ approximation. Note that deviance analogues to τ_i and $\tau_{(i)}$ may be constructed since r_i and d_i allegedly have the same asymptotic distribution.

Influence measures can be developed to compare the fit with and without a particular point (Velleman and Welsch, 1981). In other words, we examine

$$DFIT_i(c_i) = (\hat{\theta}_i - \hat{\theta}_{(i)}) / c_i$$

for some c_i. Noting (12), we have (Eubank, 1985) among other possibilities,

$$DFITS_i = \tau_i h_{ii}^{1/2}(1 - h_{ii})^{-1/2} \text{ and } DFITS_{(i)} = \tau_{(i)} h_{ii}^{1/2}(1 - h_{ii}$$

The first is approximately the confidence interval displacement diagnostic of Pregibon (1981), which measures the effect of deleting the i th datum on the chi-square fit for all points.

Pregibon (1981) suggested examining the change in the j th contribution to deviance due to the deletion of the i th observation as a means to assess the effect of an influential point on fits at neighboring points. If we define

$$r_{(j \bullet i)} = r_j + h_{ij} r_{(i)} \ ,$$

then to first order the change in deviance from deleting the i th observation is $r_{(j \bullet i)}^2 - r_j^2$. This is exactly the change in the chi-square for the linear model.

Gunst and Eubank (1983) gave an expression for the CV dispersion $\sigma_{(i)}^2$, for the linear model case using the chi-square, which we write as

$$\sigma_{(i)}^2 = \nu_{(i)}^{-1} \sum_{j \neq i} r_{(j \bullet i)}^2 \tag{13}$$

in which the the CV degrees-of-freedom are

$$\nu_{(i)} = \nu - 1 + (1 - h_{ii})^{-1} \sum_{j=1}^{n} h_{ij}(\delta_{ij} - h_{ij}) \ ,$$

with $\delta_{ij} = 1$ if $i = j$, 0 if $i \neq j$. Thus to first order, a CV estimate based on the deviance rather than the chi-square would be

$$\sigma_{(i)}^2 = \nu_{(i)}^{-1} \sum_{j \neq i} (d_j^2 - r_j^2 + r_{(j \bullet i)}^2) \ .$$

Naturally, if the deviance and chi-square residuals are close, one can use either form in determining standardized CV residuals.

5. 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*''.

The analysis presented here is not intended to be complete, but to illustrate the use of the diagnostic methods developed in this paper. One might want to carry out a stepwise procedure to determine what factors should be modeled parametrically. However, two basic realities confronted us. First, most people have phones, and many subsets of the data had practically 100% phone ownership. Efforts to look at subsets or interactions frequently led to instability of the algorithm. Second, and somewhat related, such an effort on even a subset of the data requires days on a VAX. Yandell is currently exploring improvements of algorithms for multi-dimensional nonparametric piece and the use of supercomputers for this problem. We focus here on semi-parametric logistic regressions of phone use with *age* or *income4* being non-parametric and other socio-economic factors entering in a parametric fashion. Another factor of known importance which could be handled similarly is *educ*. This data has been analyzed independently by Ed Fowlkes (personal communication) and Trevor Hastie (1986).

We took a step-wise forward selection approach to adding variables one at a time after the non-parametric piece was in the model. We considered as criteria the largest reduction in GCV (9) and the largest score statistic for adding a single variable, which lead to essentially the same sequence. The semi-parametric ''age model'' entered 8 variables at a 90% confidence level, or only 7 at a 95% or 99% level, based on the score statistics and an (assumed) asymptotic χ_1^2 distribution (Table 1 has score statistics for factors as they were entered). The semi-parametric ''income4 model'' allowed 10 variables at a 90% level, 9 at a 95% level, and only 6 at a 99% level. It is reassuring that the same variables emerged in roughly the same signed order (see Table 2) for both models. *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 in part to having only 74 unique ages while there are 1557 unique income values.

Certain changes in parameter estimates when new factors were added (Table 2) and correlations among parameters (Table 3) were predictable from the choice of factors. For instance, the family size factors which were included in the ''income4 model'', *person*, *pch1* and *pch2*, were highly correlated, as were *nonf* and *nonff* in both models; *nonf* and the family size factors were also naturally correlated. We note in addition that *educ* and *olang* parameters were highly correlated in both models, leading to noticeable changes in the *educ* parameter values when *olang* was added. Similarly *educ* was highly correlated with *age* and slightly less so with *income4*. Noticeable but non-significant changes occurred in the slope of *income4* in the ''age model'' when *nonf* and *nonff* were added, reflecting a correlation seen in Table 3a. The *olang* coefficient was correlated with that for *black* in both models, probably reflecting the fact that 98% of the blacks surveyed spoke English, while only 81% of the non-blacks had English as their primary language. Other interesting relations can probably be found, but were not pursued in the interests of time.

The model fits were achieved to minimize the GCV (9). The GCV decreased with increasing number of model parameters, as expected (see Table 4). 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 (8), 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.

5.1. Diagnostics

In this section we graphically analyse the data using diagnostics. All plots and discussion in this section refer to the ''full'' semi-parametric *age* and *income4* models. While the diagnostics are justified by small dispersion results, requiring $min\{m_i\}_{i=1}^{n} \rightarrow \infty$, we choose here to examine individual cases. One may want to group cases by similar $\hat{\theta}_i$ to ensure expected counts of 1 or 5 as in the classical categorical data settings.

The leverage values h_{ii} were generally very small for the income4 model, with none above .08. However for the age model several values were over .3 (Figure 2). 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. The ranges were remarkable constant for the two models, reflecting the fact that points of relatively high leverage probably had small simple residuals (see Table 5). We present only the standardized CV deviance residuals $\tau_{(i)}$, which measure something like Pregibon's ''coefficient sensitivity'', in Figure 3. None of the cases seem to have a great influence on the fit for the 1-D models (see Table 5).

Goodness-of-fit sensitivity appears to be a useful tool here to ferret out points which have a large influence on the deviance (Figure 4). For the two models, deviance changes of 55 (5.4%) for the age model and 44 (3.5%) for the income4 model occurred when certain points were deleted.

We now highlight several of the points which stand out via the diagnostics for the age model (Table 6). Case 1106 had the highest leverage, while cases 1305 and 1679 had large negative deviances and high goodness-of-fit sensitivity. As one can see, 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 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.

6. 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 2-D and 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 from initial investigation of 2-D models 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 when the non-parametric piece is multi-dimensional. 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

Barry, D. (1986) Nonparametric Bayesian regression. *Ann. Statist.*, **14**, 934-953.

Chen, Z. (1986) A stepwise approach for the purely periodic interaction spline model. Technical Report#792, Dept. of Statistics, U. of Wisconsin.

Cox, D. D. (1983) Gaussian approximation of smoothing splines. Technical Report#743, Dept. of Statistics, U. of Wisconsin.

Cox, D. D. and Koh, E. (1986) A smoothing spline based test of model adequacy in polynomial regression. Technical Report#787, Dept. of Statistics, U. of Wisconsin.

Cox, D. D., Koh, E., Wahba, G. and Yandell, B. S. (1986) Testing the (parametric) null model hypothesis in (semi-parametric) partial and generalized spline models. Technical Report#790, Dept. of Statistics, U. of Wisconsin.

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.

Efron, B. (1985) Double exponential families and their use in generalized linear regression. Technical Report#239, Dept. of Statistics, Stanford U.

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, vol. 32, Springer-Verlag. (Technical Report#2847, Math. Res. Cen., U. of Wisconsin)

Gunst, R. F. and Eubank, R. L. (1983) Regression diagnostics and approximate inference procedures for penalized least squares estimators. Technical Report#181, Dept. of Statistics, So. Methodist U..

Hastie, T. (1986) Generalized additive models: a GAIM analyst's toolbox. Technical Report#28, AT&T Bell Laboratories, Murray Hill.

Hastie, T. J. and Tibshirani, R. J. (1986) Generalized additive models. *Statist. Science*, **1**, 297-318. (with discussion)

Heckman, N. (1986a) Minimax estimates in a semiparametric model. Technical Report#41, Dept. of Statistics, U. of British Columbia.

Heckman, N. E. (1986b) Spline smoothing in a partly linear model. *J. Roy. Statist. Soc. Ser. B*, **48**, 244-248.

Jørgensen, B. (1986a) Some properties of exponential dispersion models. *Scand. J. Statist.*, **13**, 187-198.

Jørgensen, B. (1986b) Small dispersion asymptotics. Technical Report, Odense University.

Jørgensen, B. (1987) Exponential dispersion models. *J. Roy. Statist. Soc. Ser. B*, **49**, 000-000.

Kimeldorf, G. S. and Wahba, G. (1970) A correspondence between Bayesian estimation of stochastic processes and smoothing by splines. *Ann. Math. Statist.*, **41**, 495-502.

Leonard, T. (1982) An empirical Bayesian approach to the smooth estimation of unknown functions. Technical Report#2339, Math. Research Center, U. of Wisconsin.

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

McCullagh, P. and Nelder, J. A. (1983) *Generalized Linear Models*. New York: Chapman and Hall.

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

Nelder, J. A. and Wedderburn, R. W. M. (1972) Generalised linear interactive models. *J. Roy. Statist. Soc. Ser. A*, **135**, 370-384.

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

O'Sullivan, F. (1985) Contribution to the discussion of the paper by Silverman. *J. Roy. Statist. Soc. Ser. B*, **47**, 39-40.

O'Sullivan, F., Yandell, B. S. and Raynor, Jr., W. J. (1986) Automatic smoothing of regression functions in generalized linear models. *J. Amer. Statist. Assoc.*, **81**, 96-103.

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.

Rice, J. R. (1981) An approach to peak area estimation. *J. Res. Nat. Bur. Stand.*, **87**, 53-65.

Rice, J. R. (1986) Convergence rates for partially splined models. *Statist. and Prob. Letters*, **4**, 000-000.

Silverman, B. W. (1985) Some aspects of the spline smoothing approach to non-parametric regression curve fitting (with discussion). *J. Roy. Statist. Soc. Ser. B*, **47**, 1-52.

Speckman, P. (1985) Spline smoothing and optimal rates of convergence in nonparametric regression models. *Ann. Statist.*, **13**, 970-983.

Stone, C. J. (1985) Additive regression and other nomparametric models. *Ann. Statist.*, **13**, 689-705.

Velleman, P. and Welsch, R. E. (1981) Efficient computing of regression diagnostics. *Amer. Statist.*, **35** , 234-242.

Wahba, G. (1977) A survey of some smoothing problems and the method of generalized cross-validation for solving them.. In *Applications of Statistics* (P. R. Krishnaiah, ed.), pp.507-523. Amsterdam: North Holland.

Wahba, G. (1978) Improper priors, spline smoothing and the problem of guarding against model errors in regression. *J. Roy. Statist. Soc. Ser. B*, **40**, 364-372.

Wahba, G. (1983) Bayesian "confidence intervals" for the cross-validated smoothing spline. *J. Roy. Statist. Soc. Ser. B*, **45**, 133-150.

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.

Wahba, G. (1985) A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline smoothing problem. *Ann. Statist.*, **13**, 1378-1402.

Wahba, G. (1986) Partial and interaction splines for the semiparametric estimation of functions of several variables. *Computer Science and Statistics: Proceedings of the 18th Symposium on the Interface*, Washington, D. C.pp.75-80.

Yandell, B. S. (1985) Graphical analysis of proportional Poisson rates. In *Proceedings of the 17th Symposium on the Interface, 17-19 March 1985, Lexington* (D. M. Allen, ed.), pp.283-287. New York: North Holland.

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

Figure Captions

1. Non-parametric curve γ for full semi-parametric model: (a) age model, (b) income4 model.

2. Leverage values for age model.

3. Standardized CV deviance residuals $\tau_{(i)}$, with $\phi = 1$: (a) age model, (b) income4 model.

4. Goodness-of-fit sensitivity: (a) age model, (b) income4 model.

Figure 1a. nonparametric age

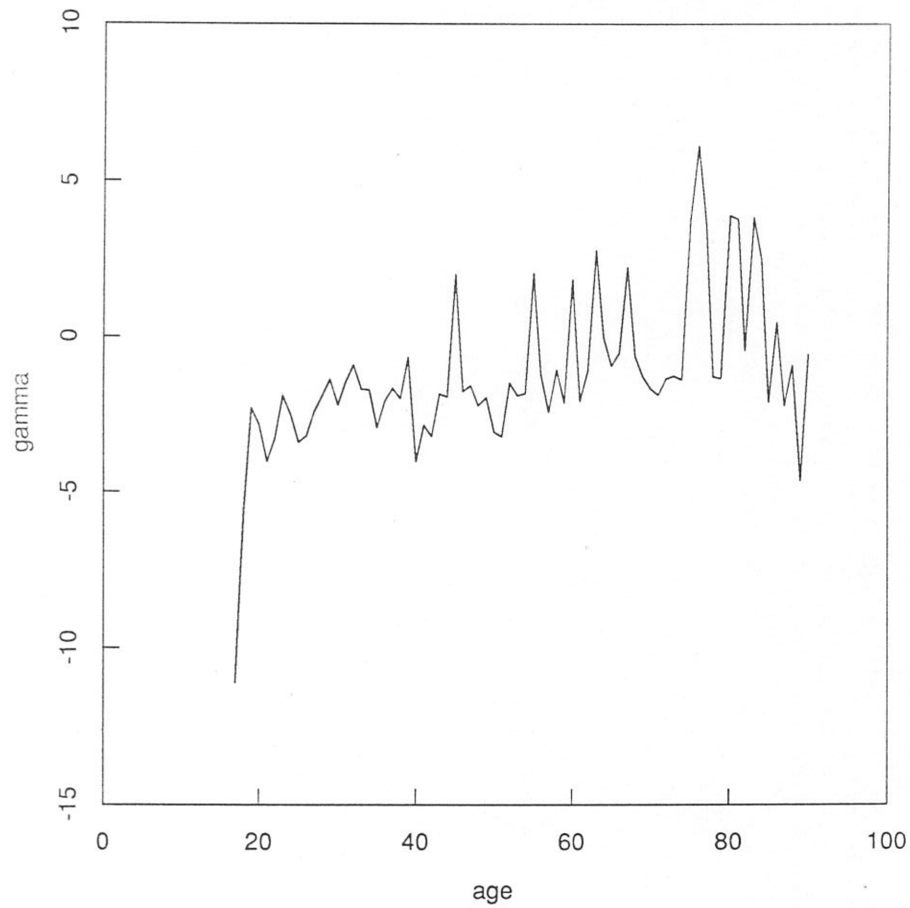

Figure 1b. nonparametric income4

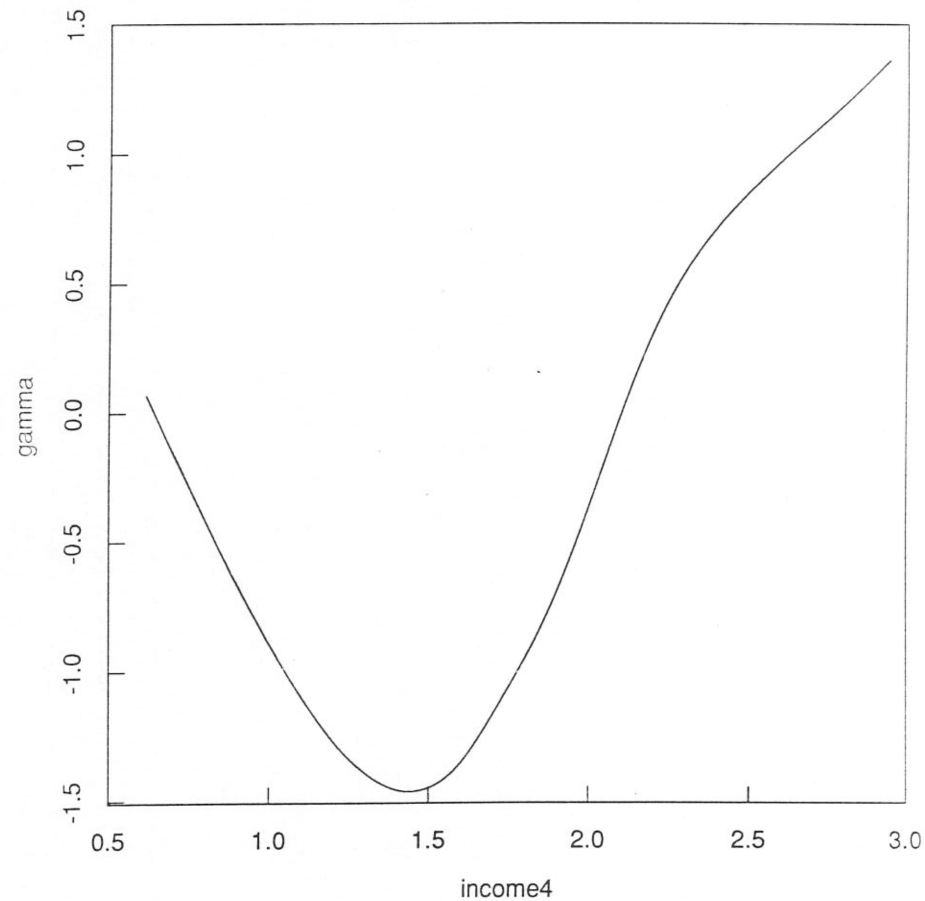

Figure 2. age leverage values

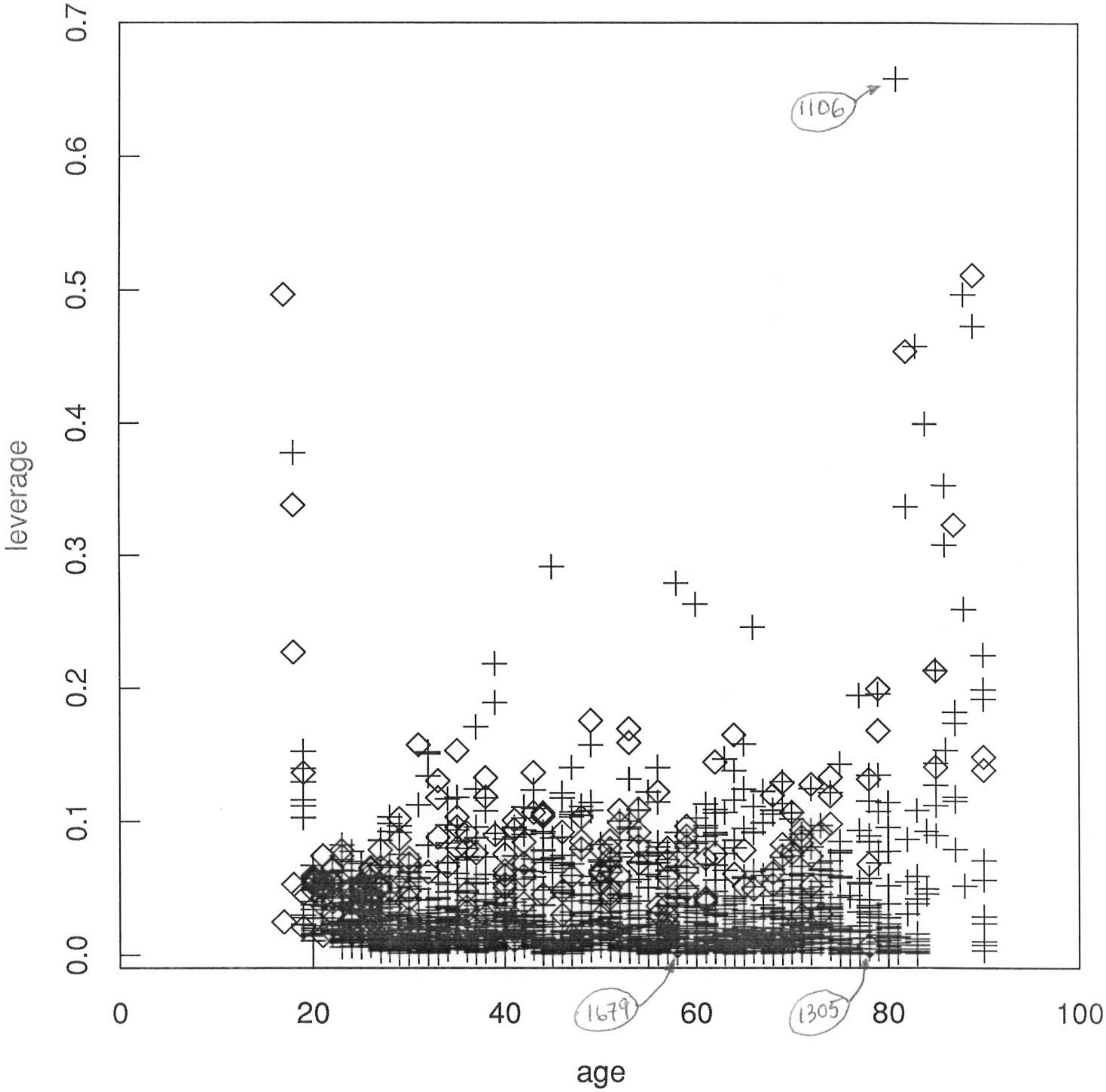

Figure 3a. age CV deviance residuals

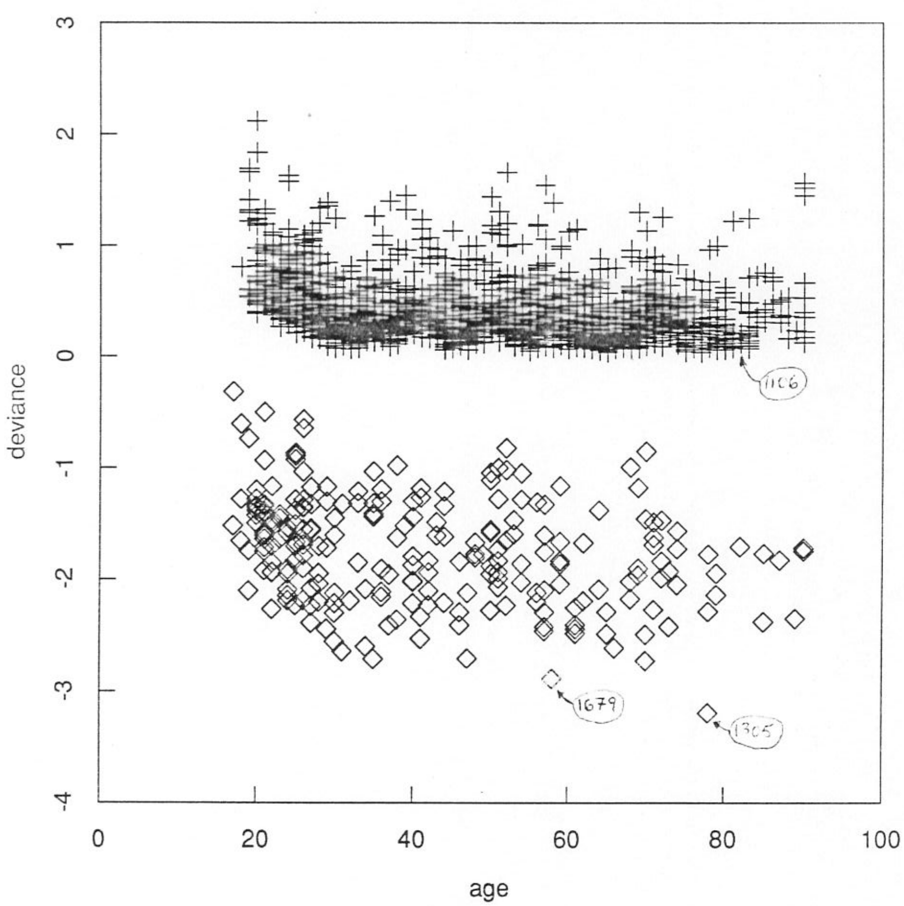

Figure 3b. income4 CV deviance residuals

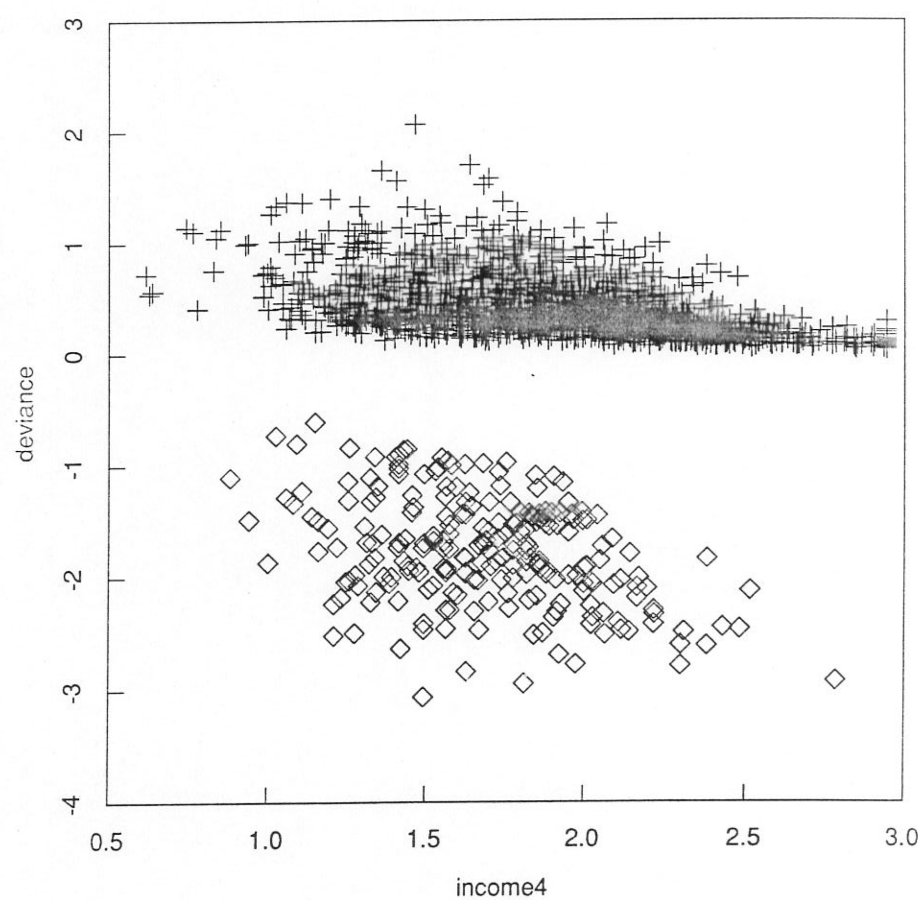

Figure 4a. age fit sensitivity

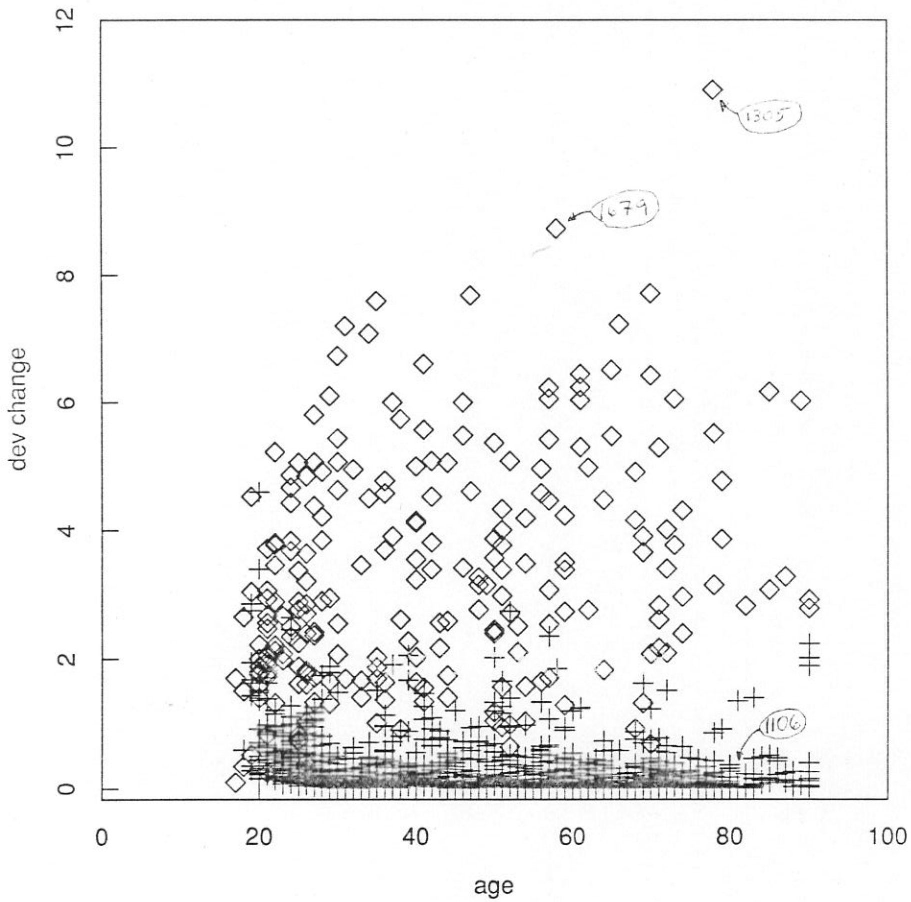

Figure 4b. income4 fit sensitivity

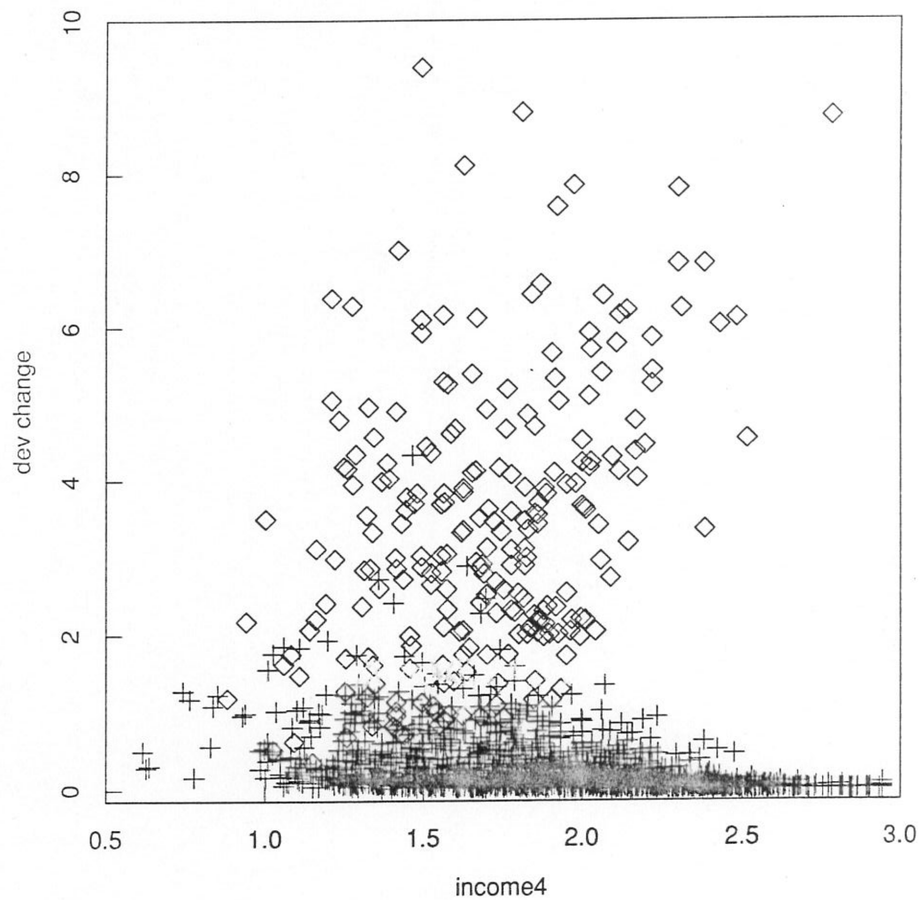

Table 1. Data Description and Score Statistics

			score for model	
name	range	description	age	income4
phone	0,1	1=household has phone		
income	.145-70	in thousands of dollars	119.96	-
age	17-90	in years (household head)	-	67.33
educ	0-20	education in years	57.36	70.85
black	0,1	1=black, 0=non-black	21.21	16.58
nonf	0,1	1=non-family household	14.11	16.65
nonff	0,1	1=nonf with female head	25.74	21.23
olang	0,1	1=English is not primary	11.19	8.52
person	1-13	number in household	8.48	4.46
pch1	0-6	number of young children	(2.18)	4.49
pch2	0-5	number of mid-age children	(0.44)	3.95
spff	0,1	single person female family	2.89	2.91
...		(10 variables not selected)	(2.09)	(2.31)

Table 2a. AGE model Parameters and SEs

educ	income4	black	nonf	nonff	olang	pch1	spff
0.236							
0.023							
0.199	1.71						
0.024	0.23						
0.208	1.60	-1.05					
0.024	0.24	0.23					
0.230	1.33	-1.09	-0.77				
0.025	0.25	0.24	0.21				
0.226	1.53	-1.05	-1.43	1.78			
0.026	0.25	0.24	0.24	0.36			
0.188	1.48	-1.24	-1.50	1.79	-0.72		
0.027	0.25	0.25	0.24	0.36	0.22		
0.183	1.46	-1.17	-1.71	1.90	-0.67	-0.31	
0.028	0.25	0.25	0.25	0.36	0.22	0.11	
0.182	1.56	-1.24	-1.64	1.93	-0.68	-0.30	0.58
0.028	0.26	0.25	0.25	0.37	0.22	0.11	0.35

Table 2b. INCOME4 model Parameters and SEs

age	educ	black	nonf	nonff	olang	pch1	pch2	person	spff
0.0330									
0.0041									
0.0478	0.177								
0.0048	0.022								
0.0485	0.183	-0.86							
0.0049	0.022	0.21							
0.0534	0.208	-0.93	-0.75						
0.0051	0.023	0.22	0.19						
0.0472	0.203	-0.88	-1.31	1.39					
0.0052	0.023	0.22	0.21	0.31					
0.0436	0.171	-1.03	-1.37	1.41	-0.57				
0.0054	0.025	0.23	0.21	0.31	0.20				
0.0402	0.171	-0.99	-1.51	1.48	-0.53	-0.205			
0.0056	0.025	0.23	0.22	0.31	0.20	0.098			
0.0413	0.173	-1.00	-1.42	1.47	-0.57	-0.213	0.27		
0.0056	0.025	0.23	0.23	0.31	0.20	0.098	0.13		
0.0416	0.166	-0.96	-1.61	1.46	-0.50	-0.061	0.47	-0.167	
0.0056	0.025	0.23	0.25	0.31	0.20	0.124	0.16	0.084	
0.0414	0.165	-1.03	-1.52	1.49	-0.51	-0.062	0.45	-0.157	0.54
0.0056	0.026	0.23	0.25	0.31	0.20	0.125	0.17	0.085	0.32

Table 3a. Parameter Correlations for full AGE model

	educ	income4	black	nonf	nonff	olang	pch1
income4	-0.165						
black	-0.009	0.065					
nonf	-0.153	0.161	0.031				
nonff	0.011	0.169	-0.009	-0.469			
olang	0.390	0.027	0.265	0.088	-0.023		
pch1	0.052	0.026	-0.104	0.303	-0.111	-0.069	
spff	-0.009	0.229	-0.150	0.133	0.054	-0.022	0.037

Table 3b. Parameter Correlations for full INCOME4 model

	age	educ	black	nonf	nonff	olang	pch1	pch2	person
educ	0.514								
black	-.055	-.0001							
nonf	0.025	-0.092	-0.027						
nonff	-0.243	-0.024	0.011	-0.430					
olang	0.172	0.379	0.284	-0.005	-0.023				
pch1	0.245	-0.080	-0.009	-0.028	-0.095	0.024			
pch2	0.096	-0.037	0.059	-0.133	-0.019	0.040	0.345		
person	-0.033	0.125	-0.116	0.394	0.009	-0.174	-0.613	-0.610	
spff	-0.016	-0.017	-0.181	0.177	0.045	-0.038	-0.012	-0.064	0.066

Table 4. Model Fit Statistics

model	dev	chi	df	λ	λJ	GCV
null	1454.8	2134.0	2133	-	-	0.682
age						
0	1272.0	1858.5	2065.8	0.00238	3.89	0.636
1 +educ	1155.4	1909.6	2065.1	0.00251	3.83	0.578
2 +income4	1098.0	1697.2	2063.7	0.00177	3.25	0.550
3 +black	1079.1	1762.2	2062.6	0.00162	3.13	0.541
4 +nonf	1065.6	1658.3	2061.9	0.00181	3.39	0.535
5 +nonff	1038.9	1722.2	2060.5	0.00146	3.00	0.522
6 +olang	1028.1	1651.0	2059.5	0.00137	2.88	0.517
7 +pch1	1020.0	1632.0	2058.5	0.00142	2.93	0.514
8 +spff	1017.0	1655.3	2057.5	0.00139	2.84	0.513
income4						
0	1338.4	2111.1	2126.2	0.00868	2.16	0.632
1 +age	1272.4	2118.7	2126.9	0.0249	2.37	0.600
2 +educ	1203.7	2006.3	2126.2	0.0292	2.13	0.568
3 +black	1189.5	2048.0	2125.4	0.0346	2.08	0.562
4 +nonf	1174.1	1889.5	2124.9	0.0510	2.29	0.555
5 +nonff	1152.9	1905.1	2123.9	0.0492	2.24	0.545
6 +olang	1144.6	1865.3	2122.8	0.0461	2.21	0.542
7 +pch1	1140.4	1848.2	2121.8	0.0463	2.10	0.541
8 +pch2	1135.7	1910.2	2120.9	0.0483	2.07	0.539
9 +person	1131.6	1884.8	2119.7	0.0414	2.11	0.537
10 +spff	1128.3	1890.0	2118.6	0.0370	2.07	0.536

Table 5. Ranges of Residuals

	age		income4	
simple				
dev	-3.18	2.08	-3.06	2.03
chi	-12.58	2.77	-10.32	2.63
stan.				
dev	-4.43	2.91	-4.17	2.79
chi	-17.49	3.88	-14.07	3.61
CV				
dev	-3.19	2.12	-3.06	2.07
chi	-12.61	2.82	-10.32	2.67
CV stan.				
dev	-4.45	2.92	-4.18	2.80
chi	-17.57	3.89	-14.12	3.62
DFITS				
dev	-1.19	0.92	-0.57	0.44
chi	-1.00	0.80	-0.73	0.45
CV DFITS				
dev	-1.19	0.92	-0.57	0.44
chi	-1.00	0.80	-0.74	0.45

Table 6. Points Influencing AGE and 2-D models

	age model			center
case	1106	1305	1679	
phone	1	0	0	.89
income	2.05	4.99	10.8	14
age	81	78	58	48
educ	0	17	17	12
black	0	0	0	.076
nonf	1	1	1	.23
nonff	0	1	1	.14
olang	1	0	0	.18
person	1	1	1	2.8
pch1	0	0	0	.35
pch2	0	0	0	.27