Plots for Generalized Additive Models

David J. Olive
Department of Mathematics
Southern Illinois University
Carbondale, Illinois 62901-4408
dolive@math.siu.edu

Keywords Diagnostics; Generalized linear models; Goodness of fit; Overdispersion; Response transformations.

Mathematics Subject Classification Primary 62G08; Secondary 62G35.

Abstract

Several useful plots for generalized linear models (GLMs) can be applied to generalized additive models (GAMs) with little modification. A plot for a GLM using the estimated sufficient predictor $ESP = \hat{\alpha} + \hat{\boldsymbol{\beta}}^T \boldsymbol{x}$ can be extended to a GAM by replacing the ESP by the estimated additive predictor $EAP = \hat{\alpha} + \sum_{j=1}^p \hat{S}_j(x_j)$. The residual plot, response plot and transformation plots are examples. Since a GLM is a special case of a GAM, a plot of EAP versus ESP is useful for checking goodness of fit of the GLM.

1. Introduction

Regression is the study of the conditional distribution $Y|\boldsymbol{x}$ of the scalar response Y given the predictors $\boldsymbol{x}$. In a *1D regression model*, Y is conditionally independent of $\boldsymbol{x}$ given a single linear combination of the predictors, called the linear predictor or *sufficient predictor* $SP = \alpha + \boldsymbol{\beta}^T \boldsymbol{x}$, written $Y \perp\!\!\!\perp \boldsymbol{x}|SP$. See Olive and Hawkins (2005).

In a *generalized additive model* (GAM), Y is conditionally independent of $\boldsymbol{x}$ given the *additive predictor* $AP = \alpha + \sum_{j=1}^p S_j(x_j)$, written $Y \perp\!\!\!\perp \boldsymbol{x}|AP$, for some functions S_j. See Hastie and Tibshirani (1990), Wood (2006) and Zuur, Ieno, Walker, Saveliev and Smith (2009). This definition of the GAM is an extension of the 1D regression model rather than the more restrictive extension of the generalized linear model (GLM). Notice that the 1D regression model is a GAM where $S_j(x_j) = \beta_j x_j$.

The following examples are important, and the GLM or 1D regression analog of the GAM can be obtained by replacing AP by SP. Often the notation "GAM" can be replaced by "regression model" to obtain the GLM analog of the GAM. Hence the binary logistic regression model is the GLM analog of the binary logistic GAM.

1) The *additive model*

$$Y|AP = AP + e \tag{1}$$

has conditional mean function $E(Y|AP) = AP$ and conditional variance function $V(Y|AP) = \sigma^2 = V(e)$. *Linear models*, including the *multiple linear regression model*, are the 1D regression analogs of the additive model.

2) The *response transformation model* is

$$Z = t^{-1}(AP + e) \ \text{ where } \ \mathrm{Y = t(Z) = AP + e.} \tag{2}$$

Here, as is often the case when the error is additive, the conditioning $Y|AP$ is suppressed.

3) The *binary logistic GAM* states that $Y_1, ..., Y_n$ are independent with

$$Y|AP \sim \mathrm{binomial(1, \rho(AP))} \ \text{ where } \ \mathrm{P(success|AP) = \rho(AP) = \frac{exp(AP)}{1 + exp(AP)}}. \tag{3}$$

This model has $E(Y|AP) = \rho(AP)$ and $V(Y|AP) = \rho(AP)(1 - \rho(AP))$.

4) The *binomial logistic GAM* states that $Y_1, ..., Y_n$ are independent with

$$Y_i|AP_i \sim \mathrm{binomial(m_i, \rho(AP_i))}. \tag{4}$$

This model has $E(Y_i|AP_i) = m_i\rho(AP_i)$ and $V(Y_i|AP_i) = m_i\rho(AP_i)(1 - \rho(AP_i))$. The binary model is a special case with $m_i \equiv 1$.

5) Some notation is needed for the beta-binomial GAM. Let $\delta = \rho/\theta$ and $\nu = (1 - \rho)/\theta$, so $\rho = \delta/(\delta + \nu)$ and $\theta = 1/(\delta + \nu)$. Let $B(\delta, \nu) = \dfrac{\Gamma(\delta)\Gamma(\nu)}{\Gamma(\delta + \nu)}$. If Y has a beta–binomial distribution, $Y \sim \mathrm{BB(m, \rho, \theta)}$, then the probability mass function of Y is

$$P(Y = y) = \binom{m}{y} \frac{B(\delta + y, \nu + m - y)}{B(\delta, \nu)}$$

for $y = 0, 1, 2, ..., m$ where $0 < \rho < 1$ and $\theta > 0$. Hence $\delta > 0$ and $\nu > 0$. Then $E(Y) = m\delta/(\delta + \nu) = m\rho$ and $\mathrm{V}(Y) = m\rho(1 - \rho)[1 + (m - 1)\theta/(1 + \theta)]$.

The *beta-binomial GAM* states that $Y_1, ..., Y_n$ are independent random variables with

$$Y_i|AP_i \sim \mathrm{BB}(\mathrm{m_i}, \rho(\mathrm{AP_i}), \theta). \tag{5}$$

This model has $E(Y_i|AP_i) = m_i\rho(AP_i)$ and

$$V(Y_i|AP_i) = m_i\rho(AP_i)(1 - \rho(AP_i))[1 + (m_i - 1)\theta/(1 + \theta)].$$

Following Agresti (2002, pp. 554-555), as $\theta \to 0$, it can be shown that the beta-binomial GAM converges to the binomial GAM.

6) The *Poisson GAM* states that $Y_1, ..., Y_n$ are independent random variables with

$$Y|AP \sim \mathrm{Poisson}(\exp(\mathrm{AP})). \tag{6}$$

This model has $E(Y|AP) = V(Y|AP) = \exp(AP)$.

7) Some notation is needed for the negative binomial GAM. If Y has a (generalized) negative binomial distribution, $Y \sim NB(\mu, \kappa)$, then the probability mass function of Y is

$$P(Y = y) = \frac{\Gamma(y + \kappa)}{\Gamma(\kappa)\Gamma(y + 1)} \left(\frac{\kappa}{\mu + \kappa}\right)^{\kappa} \left(1 - \frac{\kappa}{\mu + \kappa}\right)^{y}$$

for $y = 0, 1, 2, ...$ where $\mu > 0$ and $\kappa > 0$. Then $E(Y) = \mu$ and $\mathrm{V}(Y) = \mu + \mu^2/\kappa$.

The *negative binomial GAM* states that $Y_1, ..., Y_n$ are independent random variables with

$$Y|AP \sim \mathrm{NB}(\exp(\mathrm{AP}), \kappa). \tag{7}$$

This model has $E(Y|AP) = \exp(AP)$ and

$$V(Y|AP) = \exp(AP)\left(1 + \frac{\exp(AP)}{\kappa}\right) = \exp(AP) + \tau\exp(2\ AP).$$

Following Agresti (2002, p. 560), as $\tau \equiv 1/\kappa \to 0$, it can be shown that the negative binomial GAM converges to the Poisson GAM.

8) Suppose Y has a gamma $\text{G}(\nu, \lambda)$ distribution so that $E(Y) = \nu\lambda$ and $V(Y) = \nu\lambda^2$. The *gamma GAM* states that $Y_1, ..., Y_n$ are independent random variables with

$$Y|AP \sim G(\nu, \lambda = \mu(AP)/\nu). \tag{8}$$

Hence $E(Y|AP) = \mu(AP)$ and $V(Y|AP) = [\mu(AP)]^2/\nu$. The choices $\mu(AP) = AP$, $\mu(AP) = \exp(AP)$ and $\mu(AP) = 1/AP$ are common. Since $\mu(AP) > 0$, gamma GAMs that use the identity or reciprocal link run into problems if $\mu(EAP)$ is negative for some of the cases.

Table 1: Some Useful Plots for GAMs

Plot	Use	Extension of
Response Plot	Visualize $Y\|AP$	Cook and Weisberg (1997)
Residual Plot	Check for Lack of Fit	Plot of ESP vs. Residuals
Transformation Plot	Find Response Transformation	Olive (2004)
EE Plot	Compare EAPs of 2 GAMs	Olive and Hawkins (2005)
OD Plot	Check for Overdispersion	Winkelmann (2000, p. 110)

For a GLM, the estimated sufficient predictor $ESP = \hat{\alpha} + \hat{\boldsymbol{\beta}}^T \boldsymbol{x}$ while for a GAM, the estimated additive predictor $EAP = \hat{\alpha} + \sum_{j=1}^p \hat{S}_j(x_j)$. Table 1 lists five plots that can be extended from GLMs to GAMs with little modification, often by replacing the ESP by the EAP. A plot of w versus z will have w on the horizontal axis and z on the vertical axis. Section 2 considers the response plot, section 3 considers plots for response transformations, section 4 considers the OD plot for detecting overdispersion and section 5 considers the EE plot for comparing the $EAPs$ of two competing models.

2. Response Plots

For a GAM, a response plot, also called an estimated sufficient summary plot, is the plot of the EAP versus the response Y with the estimated conditional mean function and a scatterplot smoother often added to the plot as visual aids. The response plot is also a special case of model checking plots. See Brillinger (1983), Chambers, Cleveland, Kleiner and Tukey (1983, p. 280), Cook and Weisberg (1997, 1999: pp. 396-442) and Olive and Hawkins (2005).

The response plot is used to visualize $Y|AP$ in the background of the data for generalized additive models. Assume that the EAP takes on many values and $EAP \approx AP$. Then visualize $Y|AP = h$ by examining plotted points in a narrow vertical slice centered at $EAP = h$ in the response plot. For example, consider the single index GAM with additive error, $Y = m(AP) + e$, and suppose the zero mean constant variance errors $e_1, ..., e_n$ are iid from a unimodal distribution that is not highly skewed. Then the plotted points in the response plot should scatter about the curve formed by the estimated conditional mean function $\hat{m}(AP)$ in an evenly populated band. Note that if $EAP = AP$ and the errors $e \equiv 0$, then the plotted points would lie exactly on the curve $Y = m(AP)$. If the errors e_i are iid $N(0, \sigma^2)$, then $Y|AP \sim N(m(AP), \sigma^2)$, and plotted points in a narrow vertical slice centered at $EAP = h$ should look roughly like a sample from a $N(m(h), \sigma^2)$ distribution. For model (1) the estimated conditional mean function $\hat{m}(AP) = m(EAP) = EAP$ is the identity line with unit slope and zero intercept. If the sample size n is large, then the plotted points should scatter about the identity line and the residual $= 0$ line in an evenly populated band for the response and residual plots, with no other pattern. Note that the residual plot of EAP versus the residual is used to visualize $e|AP$.

In the GLM and 1D regression literature, AP is replaced by SP and EAP by ESP, but the 1D regression models are a special case of generalized additive models. To avoid overfitting, assume $n > 5d$ where d is the model degrees of freedom. Hence $d = p$ for multiple linear regression.

In the following examples, plots made with *Splus* used the gam function. See MathSoft (1999, ch. 10) and Chambers and Hastie (1993, ch. 7). Plots made in *R* used the gam function from the mgcv library. See Wood (2006) and R Development Core Team (2008).

Example 2.1. Chambers and Hastie (1993, pp. 251, 516) examine an environmental study that measured the four variables $Y =$ *ozone concentration, solar radiation, temperature*, and *wind speed* for 111 consecutive days. Figure 1 gives the response and residual plots for the additive model (1) that were made using *Splus.* These plots suggest that the additive model is reasonable since the plotted points follow the identity line and $r = 0$ line in roughly evenly populated bands.

If $Z_i = Y_i/m_i$, then the conditional distribution $Z_i|\boldsymbol{x}_i$ of the binomial GAM can be visualized with a response plot of the EAP versus Z_i with the estimated conditional mean function of the Z_i,

$$\hat{E}(Z|AP) = \rho(EAP) = \frac{\exp(EAP)}{1 + \exp(EAP)},$$

and a scatterplot smoother added to the plot as visual aids. Cook and Weisberg (1999, p. 515) add a lowess curve to the plot for the binomial GLM. Alternatively, divide the EAP into J slices with approximately the same number of cases in each slice. Then compute $\hat{\rho}_s = \sum_s Y_i / \sum_s m_i$ where the sum is over the cases in slice s. Then plot the resulting step function. For binary data the step function is simply the sample proportion in each slice. The response plot for the beta-binomial GAM is similar.

The lowess curve and step function are simple nonparametric estimators of the conditional mean function $\rho(AP)$. If the lowess curve or step function tracks the logistic curve (the estimated conditional mean function) closely, then the logistic conditional mean function is a reasonable approximation to the data. For the GLM, this plot is a graphical approximation of the logistic regression goodness of fit tests described in Hosmer and Lemeshow (2000, pp. 147-151).

Example 2.2. For binary data, Kay and Little (1987) suggest examining the two distributions $x|Y = 0$ and $x|Y = 1$. Use predictor x if the two distributions are roughly symmetric with similar spread. Use x and x^2 if the distributions are roughly symmetric with different spread. Use x and $\log(x)$ if one or both of the distributions are skewed. The log rule says add $\log(x)$ to the model if $\min(x) > 0$ and $\max(x)/\min(x) > 10$. The Gladstone (1905-6) data is useful for illustrating these suggestions. The response was *gender* with $Y = 1$ for male and $Y = 0$ for female. The predictors were *age, height* and the head measurements *circumference, length* and *size.* When the GAM was fit without $log(age)$ or $log(size)$, the $\hat{S}_j$ for *age*, *height* and $circumference$ were nonlinear. The log rule suggested adding $log(age)$, and $log(size)$ was added because $size$ is skewed. The GAM for this model had plots of $\hat{S}_j(x_j)$ that were fairly linear. Figure 2 shows the response plot, made in R, for this binary GAM. Note that the step function tracks the logistic curve closely. When $EAP = 0$, the estimated probability of $Y = 1$ (male) is 0.5. When $EAP > 5$ the estimated probability is near 1, but near 0 for $EAP < -5$. The response plot for the binomial GLM, not shown, is similar.

2.1. Plots for the Poisson and Negative Binomial GAMs

For Poisson regression, the response plot is a plot of EAP versus Y with $\hat{E}(Y|AP) = \exp(EAP)$ and lowess added as visual aids. If the conditional mean function is a reasonable approximation to the data, then the lowess curve should be close to the exponential curve, except possibly for the largest values of the EAP. The response plot for the negative binomial GAM is similar.

For the Poisson models, judging the conditional mean function (exponential curve) from the response plot may be rather difficult for large counts for two reasons. First, the exponential curve increases rapidly. Secondly, for real and simulated Poisson GLM and GAM data, it was observed that lowess often underestimates the exponential curve in the upper right corner of the response plot because lowess downweights the largest Y values too much.

Two new plots for the Poisson GAM transform the data towards a linear model, then make the response plot and residual plot for the transformed data. The transformation is motivated by the minimum chi–square estimator $(\hat{\alpha}_M, \hat{\boldsymbol{\beta}}_M)$ for Poisson regression which is found from the weighted least squares (WLS) regression of $\log(Z_i)$ on $\boldsymbol{x}_i$ with weights $w_i = Z_i$ where $Z_i = Y_i$ if $Y_i > 0$ and $Z_i = 0.5$ if $Y_i = 0$. Equivalently, use the least squares (OLS) regression (without intercept) of $\sqrt{Z_i}\log(Z_i)$ on $\sqrt{Z_i}(1, \boldsymbol{x}_i^T)^T$. Then the plot of the "fitted values" $\sqrt{Z_i}(\hat{\alpha}_M + \hat{\boldsymbol{\beta}}_M^T \boldsymbol{x}_i)$ versus the "response" $\sqrt{Z_i}\log(Z_i)$ should have points that scatter about the identity line. Agresti (2002, pp. 611-612) discusses when the minimum chi–square estimator and Poisson regression maximum likelihood estimator (MLE) are consistent. Since the two estimators are often close for many data sets, the plotted points in a plot of $\sqrt{Z_i}ESP$ versus $\sqrt{Z_i}\log(Z_i)$ should also scatter about the identity line.

The above reasoning motivates the following two new plots. The *weighted forward re-*

sponse plot is a plot of $\sqrt{Z_i}EAP$ versus $\sqrt{Z_i}\log(Z_i)$. The *weighted residual plot* is a plot of $\sqrt{Z_i}EAP$ versus the "WLS" residuals $r_{Wi} = \sqrt{Z_i}\log(Z_i) - \sqrt{Z_i}EAP$. These plots can also be used for the negative binomial GAM.

If the counts Y_i are large and $\hat{E}(Y|AP) = \exp(EAP)$ is a good approximation to the conditional mean function $E(Y|AP)$, then the plotted points should scatter about the identity line and $r = 0$ lines in roughly evenly populated bands. When the counts Y_i are small, the WLS residuals can not be expected to be approximately normal. Often the larger counts are fit better than the smaller counts and hence the residual plots have a "left opening megaphone" shape. This fact makes residual plots for the Poisson GAM rather hard to use, but cases with large WLS residuals may not be fit very well by the model. Both the weighted forward response and residual plots perform better for simulated Poisson regression data with many large counts than for data where all of the counts are less than 10.

Example 2.3. The species data is from Cook and Weisberg (1999, pp. 285-286) and Johnson and Raven (1973). The response variable is the total *number of species* recorded on each of 29 islands in the Galápagos Archipelago. Predictors include *area* of island, *areanear* = the area of the closest island, the *distance* to the closest island, the *elevation*, and *endem* = the number of endemic species (those that were not introduced from elsewhere). A scatterplot matrix of the predictors suggested that log transformations should be taken. Exploration suggested that $log(endem)$ and $log(areanear)$ were the important predictors, and the corresponding Poisson GAM was fit with *R*. Figure 3 shows four plots, and the response plot and weighted forward response plot in Figure 3a and 3c suggest that the Poisson conditional mean function is a good approximation to the data. In the response plot, lowess is shown as a jagged curve to distinguish lowess from the exponential curve. The weighted residual plot in Figure 3d has the common left opening megaphone shape, and suggests that there may be two clusters of data. This example and the explanation of the OD plot in Figure 3b will be continued in section 4.

3. Plots for Response Transformations

The applicability of regression models can be expanded by allowing a response transformation, and this section extends the Olive (2004) graphical method for linear model response transformations to response transformations for regression models with additive errors $Y = m(\boldsymbol{x}) + e$, including the single index GAM with additive error $Y = m(AP) + e$.

An important class of *response transformation models* adds an additional unknown transformation parameter λ_o, such that

$$Y_i = t_{\lambda_o}(Z_i) \equiv Z_i^{(\lambda_o)} = m(\boldsymbol{x}_i) + e_i \tag{9}$$

where $m(\boldsymbol{x}_i) = E(Y_i|\boldsymbol{x}_i)$. If λ_o was known, then $Y_i = t_{\lambda_o}(Z_i)$ would follow model (9) with p predictors. The function m depends on λ_o, and the p predictors x_j are assumed to be measured with negligible error. Assume that the zero mean constant variance iid errors e_i follow a unimodal distribution that is not highly skewed, and assume that the fitted values $\hat{Y} = \hat{m}(\boldsymbol{x})$ take on many values. The residuals are $r = Y - \hat{Y}$.

Next, two important response transformation models are given. Assume that *all* of the values of the "response" Z_i are *positive*. A *power transformation* has the form $Y = t_\lambda(Z) = Z^\lambda$ for $\lambda \neq 0$ and $Y = t_0(Z) = \log(Z)$ for $\lambda = 0$ where

$$\lambda \in \Lambda_L = \{-1, -1/2, -1/3, 0, 1/3, 1/2, 1\}.$$

The *modified power transformation family*

$$t_\lambda(Z_i) \equiv Z_i^{(\lambda)} = \frac{Z_i^\lambda - 1}{\lambda} \tag{10}$$

for $\lambda \neq 0$ and $Z_i^{(0)} = \log(Z_i)$. Often $Z_i^{(1)}$ is replaced by Z_i for $\lambda = 1$. Generally $\lambda \in \Lambda$ where Λ is some interval such as $[-1, 1]$ or a coarse subset such as Λ_L. This family is a special case of the response transformations considered by Tukey (1957).

A graphical method for response transformations computes the "fitted values" $\hat{W}_i$ using $W_i = t_\lambda(Z_i)$ as the "response." Then a *transformation plot* of $\hat{W}_i$ versus W_i is made for each of the seven values of $\lambda \in \Lambda_L$ with the identity line added as a visual aid. Vertical deviations from the identity line are the "residuals" $r_i = W_i - \hat{W}_i$. Then a candidate response transformation $Y = t_{\lambda^*}(Z)$ is reasonable if the plotted points follow the identity line in a roughly evenly populated band. Then take $\hat{\lambda}_o = \lambda^*$, that is, $Y = t_{\lambda^*}(Z)$ is the response transformation. Curvature from the identity line suggests that the candidate response transformation is inappropriate. Note that this procedure can be modified to create a graphical diagnostic for a numerical estimator $\hat{\lambda}$ of λ_o by adding $\hat{\lambda}$ to Λ_L. For linear models, the method proposed by Box and Cox (1964) is widely used.

After selecting the transformation, the usual checks on the model should be made. If more than one value of $\lambda \in \Lambda_L$ gives a linear plot, take the simplest or most reasonable transformation or the transformation that makes the most sense to subject matter experts. Also check that the corresponding "residual plots" of $\hat{W}_i$ versus $r_i = W_i - \hat{W}_i$ look reasonable.

Response transformations for the additive model $Y = AP + e$ are among the most difficult for regression models with additive errors since additive models are very flexible and tend to fit more than one candidate response transformation well. Rule out poor models with transformation and residual plots. For each remaining competing model, check the $\hat{S}_j$ and whether any of the predictors can be deleted.

Example 2.1 continued. Chambers and Hastie (1993, pp. 251, 516) examine the ozone data using additive models with $Z =$ *ozone concentration* or $Z^{1/3}$ as the response. Figure 4 shows four transformation plots made with *Splus*. The reciprocal transformation can be ruled out since the variability of the plotted points increases with EAP and one case is fit poorly. Similarly $\lambda = -1/2$ and $\lambda = -1/3$ can be ruled out. With the remaining transformations, the transformation and residual plots have plotted points that scatter about the identity line and the $r = 0$ line in roughly evenly populated bands except possibly for the case that appears in the lower left corner of the three remaining transformation plots in Figure 4. Figure 5 shows the residual plots of the four remaining models. No transformation $Y = Z$ may be best since the predictor *solar radiation* does not seem to be needed for this model, and the other transformations fit the case in the lower left corner poorly.

4. The OD Plot for Checking Overdispersion

Overdispersion occurs when the actual conditional variance function is larger than the model conditional variance function. Overdispersion can occur if the model is missing factors, if the response variables are correlated, if the population follows a mixture distribution, or if outliers are present.

A GAM has conditional mean and variance functions $E_M(Y|AP)$ and $V_M(Y|AP)$ where the subscript M indicates that the function depends on the model. Then overdispersion occurs if $V(Y|\boldsymbol{x}) > V_M(Y|AP)$. Let $E(Y|\boldsymbol{x})$ and $V(Y|\boldsymbol{x})$ denote the actual conditional mean and variance functions. Then the assumptions that $E(Y|\boldsymbol{x}) = E_M(Y|\boldsymbol{x}) \equiv m(AP)$ and $V(Y|\boldsymbol{x}) = V_M(Y|AP) \equiv v(AP)$ need to be checked.

First check that the assumption $E(Y|\boldsymbol{x}) = m(AP)$ is a reasonable approximation to the data using the response plot with lowess and the estimated conditional mean function $\hat{E}_M(Y|\boldsymbol{x}) = \hat{m}(AP)$ added as a visual aid.

If the conditional mean function is adequate, then we suggest checking for overdispersion using the *OD plot* of the estimated model variance $\hat{V}_M(Y|AP)$ versus the squared residuals $\hat{V} = [Y - \hat{E}_M(Y|AP)]^2$. The notation "OD" is used since the plot is a diagnostic for overdispersion, and this new plot is an extension of the plot that has been used by Winkelmann (2000, p. 110) for the Poisson regression model where $\hat{V}_M(Y|SP) = \hat{E}_M(Y|SP) = \exp(ESP)$. For binomial and Poisson regression, the OD plot can be used to complement tests and diagnostics for overdispersion such as those given in Cameron and Trivedi (1998), Collett (1999, ch. 6), and Winkelmann (2000).

For Poisson regression, Winkelmann (2000, p. 110) suggested that the plotted points in the OD plot should scatter about the identity line and that the OLS line should be approximately equal to the identity line if the Poisson regression model is appropriate. But in simulations, it was found that the following two observations make the OD plot much easier to use.

First, recall that a normal approximation is good for the Poisson distribution if the count Y is not too small. Notice that if $Y = E(Y|AP) + 2\sqrt{V(Y|AP)}$, then $[Y - E(Y|AP)]^2 = 4V(Y|AP)$. Hence if the estimated conditional mean and variance functions are both good approximations, the plotted points in the OD plot for a Poisson GAM will scatter about a wedge formed by the $\hat{V} = 0$ line and the line through the origin with slope 4: $\hat{V} = 4\hat{V}(Y|AP)$.

Only about 5% of the plotted points should be outside the wedge. Similar remarks apply to the negative binomial GAM, and to the binomial GAM if the counts are neither too big nor too small. OD plots can also be made for quasi-binomial and quasi-Poisson regression models.

Second, the evidence of overdispersion increases from slight to high as the scale of the vertical axis increases from 5 to 10 times that of the horizontal axis. (The scale of the vertical axis tends to depend on the few cases with the largest $\hat{V}(Y|AP)$, and $P[(Y - \hat{E}(Y|AP))^2 > 10\hat{V}(Y|AP)]$ can be approximated with a normal approximation or Chebyshev's inequality.) There is considerable evidence of overdispersion if the scale of the vertical axis is more than 10 times that of the horizontal, or if the percentage of points above the slope 4 line through the origin is much larger than 5%.

Hence the identity line and slope 4 line are added to the OD plot as visual aids, and one should check whether the scale of the vertical axis is more than 10 times that of the horizontal. It is easier to use the OD plot to check the variance function than the response plot since judging the variance function with the straight lines of the OD plot is simpler than judging two curves. Also outliers are often easier to spot with the OD plot.

Section 1 gives $E_M(Y|AP) = m(AP)$ and $V_M(Y|AP) = v(AP)$ for several models. Often $\hat{m}(AP) = m(EAP)$ and $\hat{v}(AP) = v(EAP)$, but additional parameters sometimes need to be estimated. Hence $\hat{v}(AP) = m_i \rho(EAP_i)(1 - \rho(EAP_i))[1 + (m_i - 1)\hat{\theta}/(1 + \hat{\theta})]$, $\hat{v}(AP) = \exp(EAP) + \hat{\tau}\exp(2\ EAP)$, and $\hat{v}(AP) = [m(EAP)]^2/\hat{\nu}$ for the beta-binomial, negative binomial and gamma GAMs, respectively. The beta-binomial regression model is often used if the binomial regression is inadequate because of overdispersion, and the negative binomial GAM is often used if the Poisson GAM is inadequate.

For GLMs, numerical summaries are also available. The deviance G^2 and Pearson goodness of fit statistic X^2 are used to assess the goodness of fit of the Poisson regression model much as R^2 is used for multiple linear regression. For Poisson regression (and binomial regression if the counts are neither too small nor too large), both G^2 and X^2 are approximately chi-square with $n - p - 1$ degrees of freedom. Since a χ^2_d random variable has mean

d and standard deviation $\sqrt{2d}$, the 98th percentile of the χ^2_d distribution is approximately $d+3\sqrt{d} \approx d+2.121\sqrt{2d}$. If G^2 or $X^2 > (n-p-1)+3\sqrt{n-p-1}$, then overdispersion may be present.

Example 2.3 continued. Figure 3b shows the OD plot for the Poisson GAM fails to indicate overdispersion. The Poisson GLM with $log(endem)$ and $log(areanear)$ was fit, but the deviance and Pearson X^2 statistics suggested overdispersion was present since both statistics were near 71.4 with 26 degrees of freedom. The residual plot (not shown) also suggested increasing variance with increasing fitted value. A negative binomial regression suggested that only $log(endem)$ was needed in the model, and had a deviance of 26.12 on 27 degrees of freedom. The residual plot for this model was roughly ellipsoidal. The negative binomial GAM with $log(endem)$ had an $\hat{S}$ that was linear.

The response plot with the exponential and lowess curves added as visual aids is shown in Figure 6. The interpretation is that $Y|\boldsymbol{x} \approx$ negative binomial with $E(Y|\boldsymbol{x}) \approx \exp(EAP)$. Hence if EAP $= 0$, $E(Y|\boldsymbol{x}) \approx 1$. The negative binomial and Poisson GAM have the same conditional mean function. If the plot was for a Poisson GAM, the interpretation would be that $Y|\boldsymbol{x} \approx$ Poisson($\exp(EAP)$). Hence if EAP $= 0$, $Y|\boldsymbol{x} \approx$ Poisson(1).

Figure 7 shows the OD plot for the negative binomial GAM with the identity line and slope 4 line through the origin added as visual aids. The plotted points fall within the "slope 4 wedge," suggesting that the negative binomial regression model has successfully dealt with overdispersion. Here $\hat{V}_M(Y|AP)$ used $\hat{\tau} = 1/37$.

5. EE Plots

An EE plot is a plot of EAP_1 versus EAP_2, and is useful for comparing two competing models. Olive and Hawkins (2005) used two EE plots for 1D regression variable selection. The EE plot of the submodel ESP versus the full model ESP was used to check whether the submodel could be used instead of the full model. If the EE plot of the OLS ESP versus the GLM ESP had plotted points that clustered tightly about some line, then some fast variable selection methods, originally meant for multiple linear regression, could be used to suggest interesting submodels for the GLM. Next, two applications of EE plots are described.

5.1. An EE Plot for Checking the GLM

One useful application of a GAM is for checking whether the corresponding GLM has the correct form of the predictors x_j in the model. Suppose a GLM and the corresponding GAM are both fit where at least one general $S_j(x_j)$ was used. Since the GLM is a special case of the GAM, the plotted points in the EE plot of EAP versus ESP should follow the identity line with very high correlation if the fitted GLM and GAM are roughly equivalent. If the correlation is not very high and the GAM has some nonlinear $\hat{S}_j(x_j)$, update the GLM, and remake the EE plot. For example, update the GLM by adding terms such as x_j^2 and possibly x_j^3, or add $\log(x_j)$ if x_j is highly skewed.

Example 5.1. Wood (2006, pp. 82-86) describes heart attack data where the response Y is the *number of heart attacks* for m_i patients suspected of suffering a heart attack. The enzyme *ck* (creatine kinase) was measured for the patients and it was determined whether the patient had a heart attack or not. A binomial GLM with predictors $x_1 = ck$, $x_2 = [ck]^2$ and $x_3 = [ck]^3$ was fit and had AIC = 33.66. The binomial GAM with predictor x_1 was fit in *R*, and Figure 8 shows that the EE plot for the GLM was not too good. The log rule suggests using *ck* and $\log(ck)$, but *ck* was not significant. Hence a GLM with the single predictor $\log(ck)$ was fit. Figure 9 shows the EE plot, and Figure 10 shows the response plot where the $Z_i = Y_i/m_i$ track the logistic curve closely. There was no evidence of overdispersion and the model had AIC = 33.45. The GAM using *log*(*ck*) had a linear $\hat{S}$, and the correlation of the plotted points in the EE plot, not shown, was one.

5.2. The EE Plot for Variable Selection

Variable selection is the search for a subset of variables that can be deleted without important loss of information. Olive and Hawkins (2005) make an EE plot of $ESP(I)$ versus ESP where $ESP(I)$ is for a submodel I and ESP is for the full model. This plot can also be used to complement the hypothesis test that the reduced model I (which is selected before gathering data) can be used instead of the full model. The obvious extension to GAMs is to make the EE plot of $EAP(I)$ versus EAP. If the fitted full model and submodel I are good, then the plotted points should follow the identity line with high correlation (≥ 0.95

as a benchmark).

To justify this claim, assume that there exists a subset S of predictor variables such that if $\boldsymbol{x}_S$ is in the model, then none of the other predictors is needed in the model. Write E for these ('extraneous') variables not in S, partitioning $\boldsymbol{x} = (\boldsymbol{x}_S^T, \boldsymbol{x}_E^T)^T$. Then

$$AP = \alpha + \sum_{j=1}^{p} S_j(x_j) = \alpha + \sum_{j \in S} S_j(x_j) + \sum_{k \in E} S_k(x_k) = \alpha + \sum_{j \in S} S_j(x_j). \tag{11}$$

The extraneous terms that can be eliminated given that the subset S is in the model have $S_k(x_k) = 0$ for $k \in E$.

Now suppose that I is a candidate subset of predictors and that $S \subseteq I$. Then

$$AP = \alpha + \sum_{j=1}^{p} S_j(x_j) = \alpha + \sum_{j \in S} S_j(x_j) = \alpha + \sum_{k \in I} S_k(x_k) = AP(I),$$

(if I includes predictors from E, these will have $S_k(x_k) = 0$). For any subset I that includes all relevant predictors, the correlation $\mathrm{corr(AP, AP(I))} = 1$. Hence if the full model and submodel are reasonable and if EAP and EAP(I) are good estimators of AP and AP(I), then the plotted points in the EE plot of EAP(I) versus EAP will follow the identity line with high correlation.

A referee pointed out that the S_j are estimated using backfitting with *Splus* and splines with *R*, and that the diagnostics are more likely to be useful when splines are used. For 1D regression, suppose $S \subseteq I$, the $\boldsymbol{x}_i$ are bounded in probability, and consistent estimators $\hat{\boldsymbol{\beta}} \xrightarrow{P} \boldsymbol{\beta}$ are used. Let $\boldsymbol{x}_i = (\boldsymbol{x}_{I,i}^T, \boldsymbol{x}_{O,i}^T)^T$ and $\boldsymbol{\beta} = (\boldsymbol{\beta}_I^T, \boldsymbol{0}^T)^T$. Then $\|\boldsymbol{x}_{I,i}^T \hat{\boldsymbol{\beta}}_I - \boldsymbol{x}_i^T \hat{\boldsymbol{\beta}}\| = \|\boldsymbol{x}_i^T [(\hat{\boldsymbol{\beta}}_I^T, \boldsymbol{0}^T)^T - \hat{\boldsymbol{\beta}}]\| \leq \|\boldsymbol{x}_i\| \ \ \|(\hat{\boldsymbol{\beta}}_I^T, \boldsymbol{0}^T)^T - \boldsymbol{\beta} + \boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\|$. Hence $\sup_{i=1,\ldots,n} \|\boldsymbol{x}_{I,i}^T \hat{\boldsymbol{\beta}}_I - \boldsymbol{x}_i^T \hat{\boldsymbol{\beta}}\| \xrightarrow{P} 0$ and $\mathrm{corr(ESP(I), ESP)} \xrightarrow{\mathrm{P}} 1$ as $n \to \infty$.

For the binary logistic GAM, the EAP will not be a consistent estimator of the AP if the estimated probability $\hat{\rho}(AP) = \rho(EAP)$ is exactly zero or one. The following example will show that GAM output and plots can still be used for exploratory data analysis. The example also illustrates that EE plots are useful for detecting cases with high leverage and clusters of cases. Numerical diagnostics, such as analogs of Cook's distances (Cook 1977), tend to fail if there is a cluster of two or more influential cases.

Example 5.2. The ICU data, available from the STATLIB URL (http://lib.stat.cmu.edu/DASL/Datafiles/ICU.html), is used to study the survival of 200 patients following admission to an intensive care unit with logistic regression. Also see Hosmer and Lemeshow (2000, pp. 23-25). The response variable was STA (0 = Lived, 1 = Died). The 19 predictors were primarily indicator variables describing the health of the patient at time of admission, including CAN = is cancer part of the present problem? (0 = No, 1 = Yes), and TYP = type of admission (0 = Elective, 1 = Emergency). Two factors had 3 levels and were fit with two indicator variables: RACE (1 = White, 2 = Black, 3 = Other) and LOC = level of consciousness at admission (0 = no coma or stupor, 1 = deep stupor, 2 = coma). The three continuous predictors were AGE, SYS = systolic blood pressure at ICU admission, and HRA = heart rate at ICU admission.

A binary logistic GAM was fit in *R* with unspecified functions for AGE, SYS and HRA and linear functions for the remaining 16 variables. Output suggested that functions for SYS and HRA are linear but the function for AGE may be slightly curved. Several cases had $\hat{\rho}(AP)$ equal to zero or one, but the response plot in Figure 11 suggests that the full model is useful for predicting survival. Note that the ten slice step function closely tracks the logistic curve.

A binary logistic regression was also fit. The response plot, not shown, was similar to Figure 11, and Figure 12 shows the EE plot of EAP versus ESP. The plot shows that the near zero and near one probabilities are handled differently by the GAM and GLM, but the estimated success probabilities $\hat{P}(Y = 1|\boldsymbol{x})$ for the two models are similar. Note that four clusters of data are present in the EE plot.

Hence we used the GLM, and variable selection suggested the submodel using AGE, CAN, SYS, TYP and LOC. Several estimated success probabilities were zero or one. Hence the full model and submodel maximum likelihood estimators did not exist. The EE plot of ESP(sub) versus ESP(full) in Figure 13 shows 4 clusters of data and that the plotted points did not cluster tightly about the identity line. The lowest cluster of points and the case on the right nearest to the identity line correspond to black patients. The main cluster and upper right cluster correspond to patients who are not black. Figure 14 shows the EE plot when RACE is added to the submodel. Then all of the points cluster about the identity line. Although variable selection did not suggest that RACE is important, possibly since the GLM MLEs corresponding to the full model and submodels do not exist, the two EE plots suggest that RACE is important. Also the RACE variable could be replaced by an indicator for black. This example shows the plots can be used to quickly improve and check the models obtained from variable selection.

6. Conclusions

Following Cook and Weisberg (1999, p. 396), a residual plot is a plot of a function of the predictors versus the residuals, while a model checking plot is a plot of a function of the predictors versus the response. Hence any residual or response plot for GAMs can be regarded as a special case of known plots. Residual plots are widely used, but model checking plots are rarely used unless there is only one predictor.

Several useful plots for 1D regression can be extended to generalized additive models by replacing the ESP by the EAP. Although residual plots are important, the response plot is more important since regression is the study of $Y|AP$. Hence response plots are also called estimated sufficient summary plots. See Cook (1998, p. 10).

The graphical response transformation method in section 3 is similar to the Cook and Olive (2001) method for linear models where the "transformation plot" of $\hat{Z}_i$ versus W_i is made for each of the seven values of $\lambda \in \Lambda_L$. Cook and Weisberg (2004) give a graphical method for multiple linear regression, noting that an *inverse response plot* of Z versus $\hat{Z}$ can often be used to visualize t_{λ_0}. Then the transformation plot of $\hat{Z}$ versus Z can be used to visualize $t_{\lambda_0}^{-1}$. An advantage of this procedure is that the family of transformations need not be picked in advance, but the predictors need to be well behaved, and it may be difficult to generalize this method to experimental designs and additive models.

1D regression analogs of the plots in this paper are discussed in Olive (2010), as are plots for experimental design models, generalized least squares models and survival regression models. Some data sets can be found at (www.math.siu.edu/olive/regbk.htm) and (www.math.siu.edu/olive/regdata.txt). The *regpack R/Splus* functions found at (www.math.siu.edu/olive/regpack.txt) include `lrplot` which makes response and OD plots for binomial regression; *lrplot2* which makes the response plot for binary regression; *prplot* which makes the response, weighted forward response, weighted residual and OD plots for Poisson regression; and *prsim* which makes the last 4 plots for simulated Poisson or negative binomial regression models.

R/Splus code to reproduce the figures of this paper can be found at (www.math.siu.edu/olive/ppgamcode.txt). The Venables and Ripley (2010) library `MASS` was used for the negative binomial family. The Lesnoff and Lancelot (2010) *R* package `aod` is useful for fitting negative binomial regression and has function `betabin` for beta-binomial regression.

Acknowledgments

The author thanks the two referees for their comments that improved this article. This research was partially supported by the National Science Foundation grant DMS 0600933.

References

Agresti, A. (2002). *Categorical Data Analysis.* second ed., Hoboken, NJ: Wiley.

Box, G. E. P., Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, B* 26:211-246.

Brillinger, D. R. (1983). A generalized linear model with "Gaussian" regressor variables. In: Bickel, P.J., Doksum, K.A., Hodges, J.L., eds., *A Festschrift for Erich L. Lehmann.* Pacific Grove, CA: Wadsworth, pp. 97-114.

Cameron, A. C., Trivedi, P. K. (1998). *Regression Analysis of Count Data.* Cambridge: Cambridge University Press.

Chambers, J. M., Cleveland, W. S., Kleiner, B., Tukey, P. (1983). *Graphical Methods for Data Analysis.* Boston: Duxbury Press.

Chambers, J. M., Hastie, T. J. (eds.) (1993). *Statistical Models in S.* New York: Chapman

& Hall.

Collett, D. (1999). *Modelling Binary Data.* Boca Raton, FL: Chapman & Hall/CRC.

Cook, R. D. 1977. Deletion of influential observations in linear regression. *Technometrics,* 19:15-18.

Cook, R. D. (1998). *Regression Graphics: Ideas for Studying Regression Through Graphics.* New York: Wiley.

Cook, R. D., Olive, D. J. (2001). A note on visualizing response transformations in regression. *Technometrics* 43:443-449.

Cook, R. D., Weisberg, S. (1994). Transforming a response variable for linearity. *Biometrika* 81:731-737.

Cook, R. D., Weisberg, S. (1997). Graphics for assessing the adequacy of regression models. *J. Amer. Statist. Assoc.* 92:490-499.

Cook, R. D., Weisberg, S. (1999). *Applied Regression Including Computing and Graphics.* New York: Wiley.

Gladstone, R. J. (1905-6). A study of the relations of the brain to the size of the head. *Biometrika* 4:105-123.

Hastie, T. J., Tibshirani, R. J. (1990). *Generalized Additive Models.* London: Chapman & Hall.

Hosmer, D. W., Lemeshow, S. (2000). *Applied Logistic Regression.* second ed., New York: Wiley.

Johnson, M. P., Raven, P. H. (1973). Species number and endemism, the Galápagos archipelago revisited. *Science* 179:893-895.

Kay, R., Little, S. (1987). Transformations of the explanatory variables in the logistic regression model for binary data. *Biometrika* 74:495-501.

Lesnoff, M., Lancelot, R. (2010). Aod: analysis of overdispersed data. R package version 1.2, (http://cran.r-project.org/package=aod).

MathSoft (1999). *S-Plus 2000 Guide to Statistics,* Vol. I, Data Analysis Products Division, MathSoft, Seattle, WA.

Olive, D. J. (2004). Visualizing 1D regression. In: Hubert, M., Pison, G., Struyf, A., Van Aelst S., eds. *Theory and Applications of Recent Robust Methods*, Series: Statistics for Industry and Technology. Basel: Birkhäuser, pp. 221-233.

Olive, D. J. (2010). *Multiple Linear and 1D Regression Models*, Unpublished Online Text available from (www.math.siu.edu/olive/regbk.htm).

Olive, D. J., Hawkins, D. M. (2005). Variable selection for 1D regression models. *Technometrics* 47:43-50.

R Development Core Team (2008). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, (www.R-project.org).

Tukey, J. W. (1957). Comparative anatomy of transformations. *Ann. Math. Statist.* 28:602-632.

Venables, W. N., Ripley, B. D. (2010). *Modern Applied Statistics with S.* fourth ed., New York: Springer.

Winkelmann, R. (2000). *Econometric Analysis of Count Data.* third ed., New York: Springer-Verlag.

Wood, S. N. (2006). *Generalized Additive Models: an Introduction with R.* Boca Rotan, FL: Chapman & Hall/CRC.

Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., Smith, G. M. (2009). *Mixed Effects Models and Extensions in Ecology with R.* New York: Springer-Science.

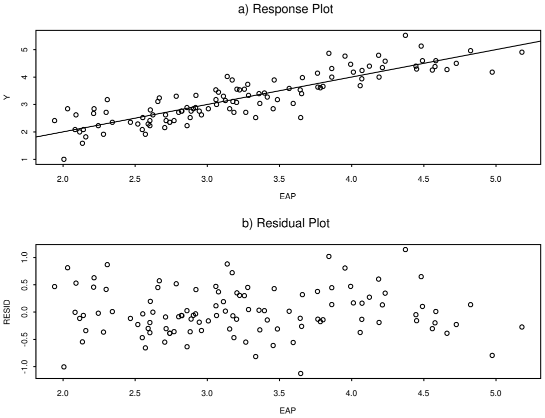

Figure 1: Visualizing the Additive Model for the Ozone Data

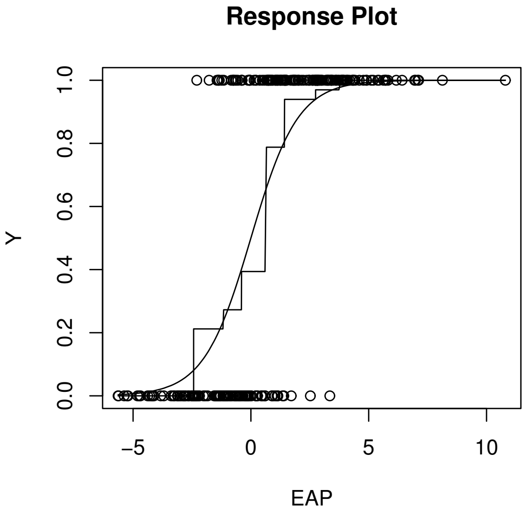

Figure 2: Visualizing the Binomial GAM for the Gladstone Data

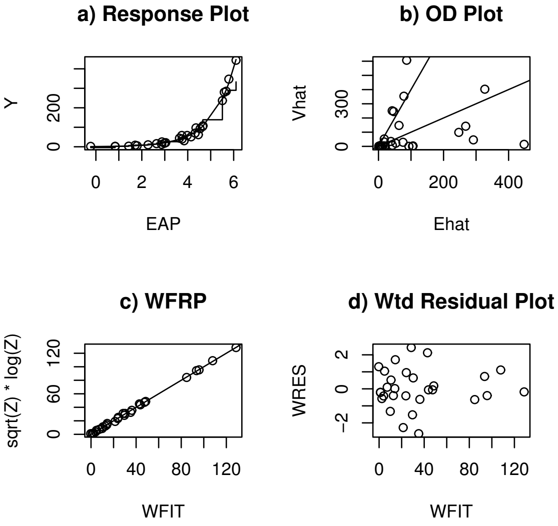

Figure 3: Plots for the Poisson GAM for the Species Data

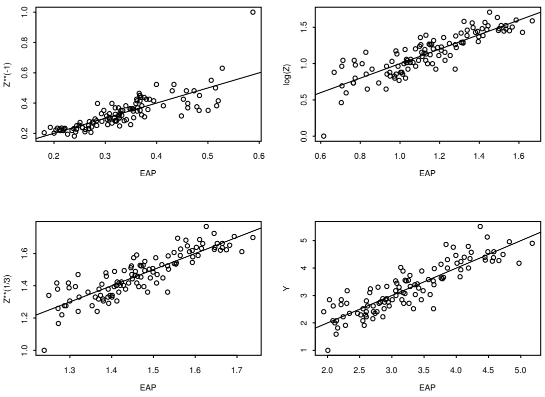

Figure 4: Transformation Plots for the Ozone Data

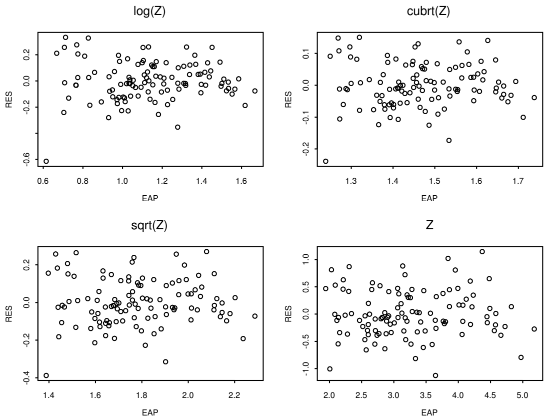

Figure 5: Residual Plots for the Ozone Data

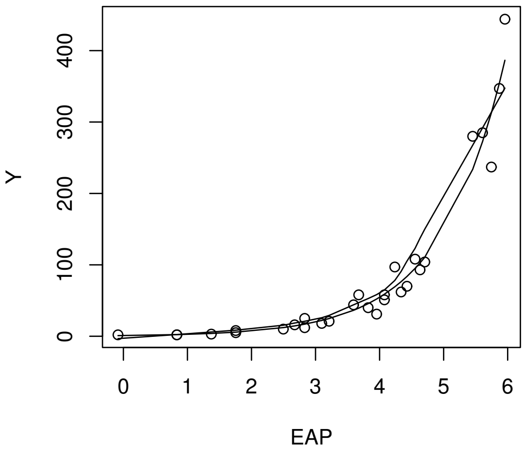

Figure 6: Response Plot for the Negative Binomial GAM for the Species Data

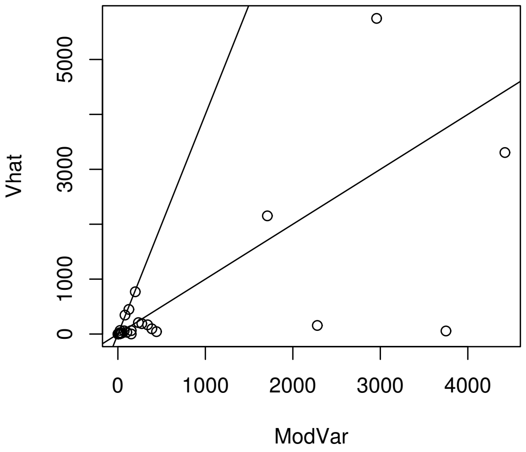

Figure 7: OD Plot for the Species Data

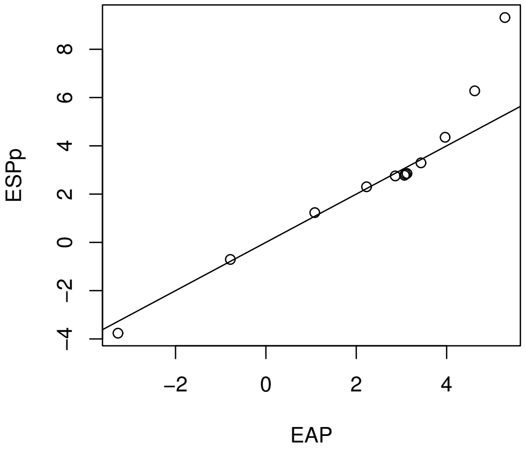

Figure 8: EE plot for Wood (2006) GLM for the Heart Attack Data

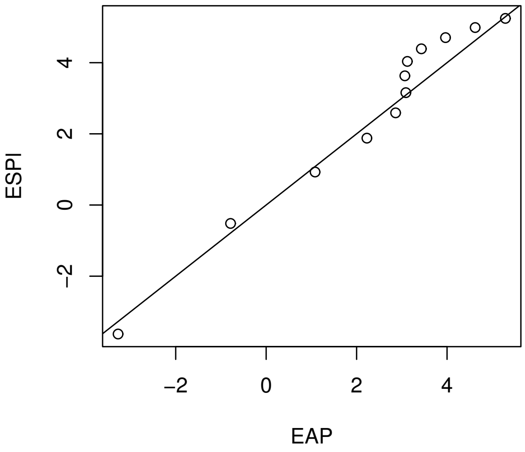

Figure 9: EE plot with log(ck) in the GLM

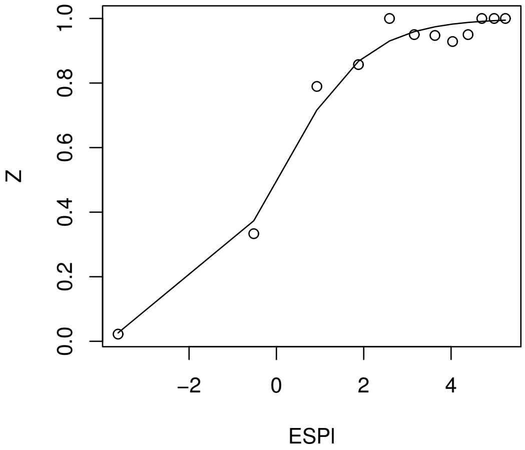

Figure 10: Response Plot for the Heart Attack Data

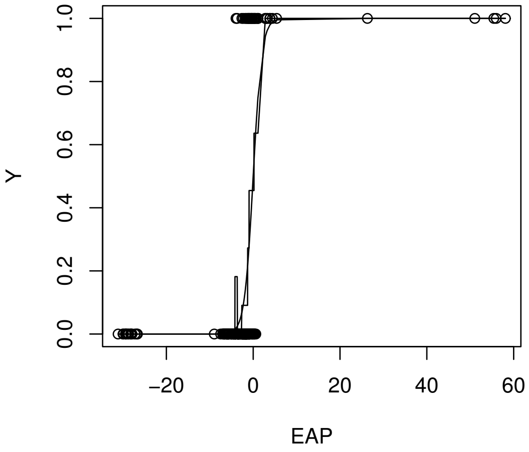

Figure 11: Visualizing the ICU GAM

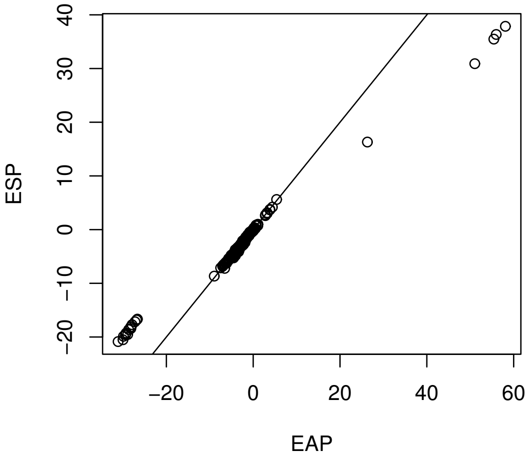

Figure 12: GAM and GLM give Similar Success Probabilities

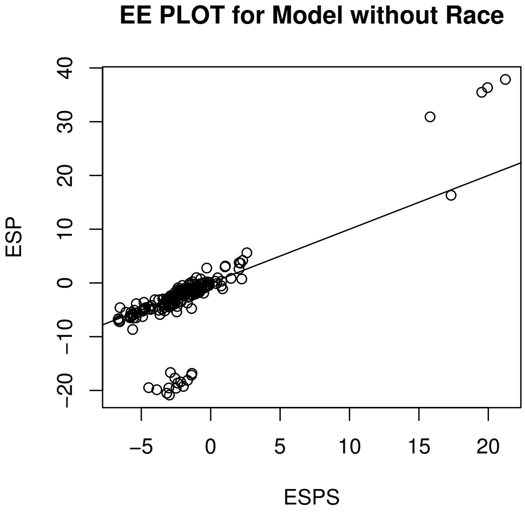

Figure 13: EE Plot Suggests Race is an Important Predictor

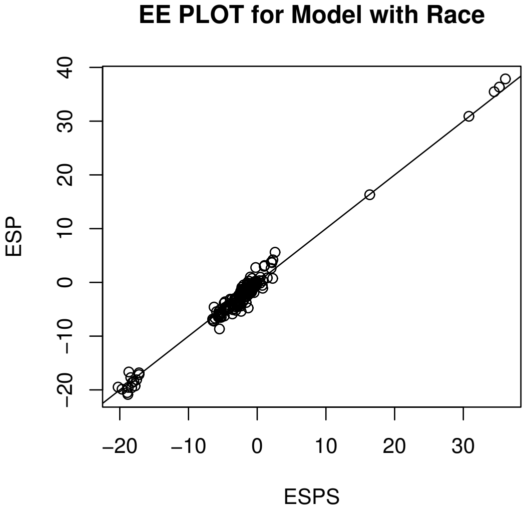

Figure 14: EE Plot Suggests Race is an Important Predictor