Graphical Aids for Regression

David Olive*

Southern Illinois University

January 13, 2004

Abstract

Regression is the study of the conditional distribution of the response Y given the predictors $\boldsymbol{x}$. In a 1D regression, Y is independent of $\boldsymbol{x}$ given a single linear combination $\alpha + \boldsymbol{\beta}^T\boldsymbol{x}$ of the predictors. Special cases of 1D regression include multiple linear regression, logistic regression, generalized linear models and single index models. Plots can be very useful for model description and as diagnostics. For example, a sufficient summary plot of $\alpha + \boldsymbol{\beta}^T\boldsymbol{x}_i$ versus Y_i can be used to explain the model to students and consulting clients. An estimated sufficient summary plot of $\hat{\alpha} + \hat{\boldsymbol{\beta}}^T\boldsymbol{x}_i$ versus Y_i can be used as a diagnostic for goodness of fit and as a diagnostic for the test $H_o : \boldsymbol{\beta} = \boldsymbol{0}$.

KEY WORDS: Diagnostics, Generalized Linear Models, Goodness of Fit, Logistic Regression, Loglinear Regression, Single Index Models.

*David J. Olive is Assistant Professor, Department of Mathematics, Southern Illinois University, Mailcode 4408, Carbondale, IL 62901-4408, USA. E-mail address: dolive@math.siu.edu. This research was supported by NSF grant DMS 0202922.

1 INTRODUCTION

Regression is the study of the conditional distribution $Y|\boldsymbol{x}$ of the response Y given the $k \times 1$ vector of nontrivial predictors $\boldsymbol{x}$. In a *1D regression model* (or regression with 1–dimensional structure), Y is conditionally independent of $\boldsymbol{x}$ given a single linear combination $\boldsymbol{\beta}^T\boldsymbol{x}$ of the predictors, written

$$Y \perp\!\!\!\perp \boldsymbol{x}|\boldsymbol{\beta}^T\boldsymbol{x}. \tag{1.1}$$

The class of 1D regression models is very rich. An important class of 1D regression models has the form

$$Y = g(\alpha + \boldsymbol{\beta}^T\boldsymbol{x}, e) \tag{1.2}$$

where g is a bivariate (inverse link) function and e is a zero mean error that is independent of $\boldsymbol{x}$. See Li and Duan (1989) and Cook and Weisberg (1999, p. 414). A *single index model* uses

$$Y = g(\alpha + \boldsymbol{\beta}^T\boldsymbol{x}, e) \equiv m(\alpha + \boldsymbol{\beta}^T\boldsymbol{x}) + e, \tag{1.3}$$

and the *multiple linear regression model* is an important special case where m is the identity function: $m(\alpha+\boldsymbol{\beta}^T\boldsymbol{x}) = \alpha+\boldsymbol{\beta}^T\boldsymbol{x}$. Another important special case of 1D regression is the *response transformation model* where

$$g(\alpha + \boldsymbol{\beta}^T\boldsymbol{x}, e) = t^{-1}(\alpha + \boldsymbol{\beta}^T\boldsymbol{x} + e) \tag{1.4}$$

and t^{-1} is a one to one (typically monotone) function so that $t(y) = \alpha + \boldsymbol{\beta}^T\boldsymbol{x} + e$.

Generalized linear models (GLM's) are also a special case of 1D regression, and two of the most important GLM's are the logistic and loglinear regression models. The *logistic* *regression model* states that $Y_1, ..., Y_n$ are independent random variables with

$$Y_i \sim \text{binomial}(\mathrm{m_i}, \rho(\boldsymbol{x}_\mathrm{i})) \quad \text{where}$$

$$P(\text{success}|\boldsymbol{x}_\mathrm{i}) = \rho(\boldsymbol{x}_\mathrm{i}) = \frac{\exp(\alpha + \boldsymbol{\beta}^\mathrm{T}\boldsymbol{x}_\mathrm{i})}{1 + \exp(\alpha + \boldsymbol{\beta}^\mathrm{T}\boldsymbol{x}_\mathrm{i})}. \tag{1.5}$$

The *loglinear regression model* states that $Y_1, ..., Y_n$ are independent random variables with

$$Y_i \sim \text{Poisson}(\mu(\boldsymbol{x}_\mathrm{i})) \quad \text{where} \quad \mu(\boldsymbol{x}_\mathrm{i}) = \exp(\alpha + \boldsymbol{\beta}^\mathrm{T}\boldsymbol{x}_\mathrm{i}). \tag{1.6}$$

Assume that the data has been collected and that a 1D regression model (1.1) has been fitted. Suppose that the *sufficient predictor*

$$SP = \alpha + \boldsymbol{\beta}^T\boldsymbol{x} = \alpha + \boldsymbol{\beta}_R^T\boldsymbol{x}_R + \boldsymbol{\beta}_O^T\boldsymbol{x}_O \tag{1.7}$$

where the $r \times 1$ vector $\boldsymbol{x}_R$ consists of the predictors in the *reduced model*. Then the investigator will often want to check whether the model is useful and to perform inference. Several things to consider are listed below.

i) Explain the 1D regression model to consulting clients, students or researchers.

ii) Goodness of fit: show that the model provides a simple, useful approximation for the relationship between the response variable Y and the predictors $\boldsymbol{x}$.

iii) Check for lack of fit of the model.

iv) Test $H_o : \boldsymbol{\beta} = \mathbf{0}$, that is, check whether the predictors $\boldsymbol{x}$ are needed in the model.

v) Test $H_o : \boldsymbol{\beta}_O = \mathbf{0}$, that is, check whether the reduced model can be used instead of the full model.

vi) Use variable selection to find a good submodel.

vii) Estimate the mean function $E(Y_i|\boldsymbol{x}_i) = \mu(\boldsymbol{x}_i) = d_i \tau(\boldsymbol{x}_i)$ or estimate $\tau(\boldsymbol{x}_i)$ where the d_i are known constants.

The 1D regression models offer a unifying framework for many of the most used regression models. By writing the model in terms of the sufficient predictor $SP = \alpha + \boldsymbol{\beta}^T \boldsymbol{x}$, many important topics valid for all 1D regression models can be explained compactly. For example, (1.7) can be used to motivate the test for whether the reduced model can be used instead of the full model. Similarly, the sufficient predictor can be used to explain models for variable selection and to unify the interpretation of coefficients, interactions and factors. For example, if $x_1, ..., x_{i-1}, x_{i+1}, ..., x_k$ can be held fixed, then a unit increase in x_i changes the sufficient predictor SP by β_i. If $SP = \alpha + \beta_1 x + \beta_2 x^2$, then

$$\frac{d}{dx} SP = \beta_1 + 2\beta_2 x.$$

Some notation from the regression graphics literature will be useful. *Dimension reduction* can greatly simplify our understanding of the conditional distribution $Y|\boldsymbol{x}$. If a 1D regression model is appropriate, then the k–dimensional vector $\boldsymbol{x}$ can be replaced by the 1–dimensional scalar $\boldsymbol{\beta}^T \boldsymbol{x}$ with "no loss of information." A *sufficient summary plot* (SSP) is a plot that contains all the sample regression information about the conditional distribution of $Y|\boldsymbol{x}$. For 1D regression, if $Y \perp\!\!\!\perp \boldsymbol{x}|\boldsymbol{\beta}^T \boldsymbol{x}$ then $Y \perp\!\!\!\perp \boldsymbol{x}|a\alpha + c\boldsymbol{\beta}^T \boldsymbol{x}$ for any constants a and $c \neq 0$. The quantity $a\alpha + c\boldsymbol{\beta}^T \boldsymbol{x}$ is called a *sufficient predictor* (SP), and a plot of a SP versus Y is a SSP. If a consistent estimator $\hat{\boldsymbol{b}}$ of $c\boldsymbol{\beta}$ can be found for some nonzero c, then an *estimated sufficient summary plot* (ESSP) is a plot of the *estimated sufficient predictor* (ESP) $a\hat{\alpha} + \hat{\boldsymbol{b}}^T \boldsymbol{x}_i$ versus Y_i. For parametric models such as GLM's, we will use $a = c = 1$, but for semiparametric models such as the single index model (1.3) where m is unknown, $\boldsymbol{\beta}$ is not identifiable although often many estimators of $c\boldsymbol{\beta}$ exist where c is unknown. For semiparametric 1D regression models, $a = 0$ is a common choice, and in the econometrics literature, often c is chosen so that $\hat{\boldsymbol{b}} = (1, \hat{b}_2, ..., \hat{b}_k)^T$. See Horowitz (1998, pp. 14-16).

Sections 2 and 3 will illustrate uses for graphs such as the SSP and ESSP. These plots are especially useful if the sufficient predictor $\alpha + \boldsymbol{\beta}^T \boldsymbol{x}_i$ takes on many values and if the mean function $E(Y_i|\boldsymbol{x}_i)$ is of interest.

2 A Graphical Aid for Model Description

To help explain the model, use the sufficient summary plot (SSP) of $SP = \alpha + \boldsymbol{\beta}^T \boldsymbol{x}_i$ versus Y_i with the mean function added as a visual aid. If $k = 1$, then $Y \perp\!\!\!\perp x|x$ and the plot of x_i versus Y_i is a SSP and has been widely used to explain the simple linear regression (SLR) model and the logistic regression model with one predictor. See Agresti (2002, cover illustration and p. 169) and Collett (1999, p. 74). Replacing x by SP has two major advantages. First, the plot can be made for $k \geq 1$ and secondly, the possible shapes that the plot can take is greatly reduced. For example, in a plot of x_i versus Y_i, the plotted points will fall about some line with slope β and intercept α if the SLR model holds, but in a plot of $SP = \alpha + \boldsymbol{\beta}^T \boldsymbol{x}_i$ versus Y_i, the plotted points will fall about the identity line with unit slope and zero intercept if the multiple linear regression model holds.

We used artificial data sets to illustrate the plots since if $k > 1$, $\boldsymbol{\beta}$ must be known to make a SSP. The multiple linear regression (MLR), logistic regression (LR) and the loglinear regression (LLR) models were used for illustration. The artificial MLR data used $\alpha = -1$, $\boldsymbol{\beta} = (1,2,3,0,0)^T$, $e_i \sim N(0,1)$, $\boldsymbol{x} \sim N_5(\mathbf{0}, \boldsymbol{I})$, and $n = 100$ cases. For the artificial LR data set, $P(Y = j) = 0.5$ and $\boldsymbol{x}|Y = j \sim N_k(\boldsymbol{\mu}_j, \boldsymbol{I})$ for $j = 0,1$. The data set used $\boldsymbol{\mu}_1 = (1,1,1,0,0)^T$, $\boldsymbol{\mu}_0 = \mathbf{0}$, and 200 cases, half of which had $Y_i = 1$. Then $\alpha = -0.5\boldsymbol{\mu}_1^T\boldsymbol{\mu}_1 = -1.5$ and $\boldsymbol{\beta} = (1,1,1,0,0)^T = \boldsymbol{\mu}_1$. See Hosmer and Lemeshow (2000, pp. 43-44). The artificial LLR data set used $n = 100$, $\boldsymbol{x} \sim N_5(\mathbf{1}, \boldsymbol{I}/4)$ and $Y_i \sim \text{Poisson}(\exp(\alpha + \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_{\mathrm{i}}))$ where $\alpha = -2.5$ and $\boldsymbol{\beta} = (1,1,1,0,0)^T$.

Figure 1 corresponds to the SSP for the MLR data. Notice that the identity line with unit slope and zero intercept corresponds to the mean function since the identity line is the line $Y = SP = \alpha + \boldsymbol{\beta}^T\boldsymbol{x} = E(Y|\boldsymbol{x})$. The vertical deviation of Y_i from the line is equal to $e_i = Y_i - (\alpha + \boldsymbol{\beta}^T\boldsymbol{x}_i)$. For a given value of SP, $Y_i \sim N(SP, 1)$. Hence if $SP = 0$ then $Y_i \sim N(0,1)$, and if $SP = 5$ then $Y_i \sim N(5,1)$. Imagine superimposing the $N(SP, 1)$ curve at various values of SP. If all of the curves were shown, then the plot would resemble a road through a tunnel. For the artificial data, each Y_i is a sample of size 1 from the normal curve with mean $\alpha + \boldsymbol{\beta}^T\boldsymbol{x}_i$.

Figure 2 corresponds to the SSP for the LR data. Unlike the SSP for multiple linear regression where the mean function is always the identity line, the mean function

$$\rho(SP) = \frac{\exp(SP)}{1 + \exp(SP)}$$

can take a variety of shapes depending on the range of the SP. If Y is binary then $Y|SP = 0 \sim$ binomial(1,0.5), $Y|SP = -5 \sim$ binomial(1,$\rho \approx 0.007$), and $Y|SP = 5 \sim$ binomial(1,$\rho \approx 0.993$). Hence if the range of the SP is in the interval $(-\infty, -5)$, then the mean function is flat and $\rho(SP) \approx 0$. If the range of the SP is in the interval $(5, \infty)$, then the mean function is again flat but $\rho(SP) \approx 1$. If $-5 < SP < 0$ then the mean function increases slowly and then rapidly. If $-1 < SP < 1$ then the mean function looks roughly linear. If $0 < SP < 5$ then the mean function first increases rapidly and then slowly. Finally, if $-5 < SP < 5$ then the mean function has the characteristic "ESS" shape shown in Figure 2.

Figure 3 corresponds to the SSP for the LLR data. Notice that $Y|SP = 0 \sim$ Poisson(1), and in general, $Y|SP \sim$ Poisson(exp(SP)). The shape of the mean function $\mu(SP) = \exp(SP)$ for loglinear regression depends strongly on the range of the SP because the plotting software attempts to fill the vertical axis. Hence if $\max(Y_i)$ is less than 3 then the exponential function will be rather flat, but if there is a single large count, then the exponential curve will look flat in the left of the plot but will increase sharply in the right of the plot.

3 Graphical Aids for Goodness and Lack of Fit

The estimated sufficient summary plot (ESSP) is a plot of the estimated sufficient predictor $ESP = \hat{\alpha} + \hat{\boldsymbol{\beta}}^T \boldsymbol{x}_i$ versus Y_i. This plot can be used as a diagnostic for goodness of fit by adding the estimated parametric mean function and an estimated nonparametric mean function to the plot. The interpretation of the ESSP is almost the same as that of the SSP, but the SP is replaced by its estimator, the ESP. Adding the horizontal line $Y = \hat{\mu}$ (often using $\hat{\mu} = \overline{Y}$) to the plot is a diagnostic for the test of $H_o : \boldsymbol{\beta} = \mathbf{0}$ versus $H_A : \boldsymbol{\beta} \neq \mathbf{0}$. A plot of the residuals versus the ESP is often used as a diagnostic for lack of fit.

Figure 4 gives the ESSP, also called the *forward response plot*, for the MLR data. Ordinary least squares (OLS) is often used to estimate $(\alpha, \boldsymbol{\beta})$, and the estimated mean function is the identity line. Now the vertical deviation of Y_i from the line is equal to the residual $r_i = Y_i - (\hat{\alpha} + \hat{\boldsymbol{\beta}}^T \boldsymbol{x}_i)$. The residual plot is not shown but resembles Figure 10.

3.1 Logistic Regression

Figure 5 shows the ESSP for the LR data set with the estimated mean function

$$\hat{\rho}(ESP) = \frac{\exp(ESP)}{1 + \exp(ESP)}$$

added as a visual aid. This plot is very useful as a goodness of fit diagnostic. Divide the ESP into J "slices" each containing approximately n/J cases. Compute the sample mean = sample proportion of the $Y's$ in each slice and add the resulting step function to the ESS plot. This is done in Figure 5 with $J = 10$ slices. This step function is a simple nonparametric estimator of the mean function $\rho(SP)$. If the step function follows the estimated LR mean function (the logistic curve) closely, then the LR model fits the data well. The plot of these two curves is a graphical approximation of the goodness of fit tests described in Hosmer and Lemeshow (1980, 2000, pp. 147–156).

For binary data the Y_i only take two values, 0 and 1, and the residuals do not behave very well. Let V be a linear combination of the predictors that is (approximately) uncorrelated with the estimated sufficient predictor ESP. Then a *binary response plot* is a plot of the ESP versus V where different plotting symbols are used for $Y = 0$ and $Y = 1$. Instead of using residual plots, we suggest using the binary response plot, introduced by Cook (1996). Also see Cook (1998, Ch. 5) and Cook and Weisberg (1999, Section 22.2).

To make a binary response plot for logistic regression, sliced inverse regression (SIR) can be used to find V. SIR is a regression graphics method (see Li 1991 and Cook 2004) and the first SIR predictor $\hat{\boldsymbol{\beta}}_{SIR1}^T \boldsymbol{x}$ is used as the ESP while the second SIR predictor $\hat{\boldsymbol{\beta}}_{SIR2}^T \boldsymbol{x}$ is used as V. (Other regression graphics methods may provide a better plot, but the first SIR predictor is often highly correlated with the LR ESP $\hat{\alpha} + \hat{\boldsymbol{\beta}}^T \boldsymbol{x}$.) After fitting SIR and LR, check that

$$|\text{corr}(\text{SIRESP}, \text{LRESP})| > 0.95.$$

If the LR model holds, then Y is independent of $\boldsymbol{x}$ given the SP. If the absolute correlation is high, then this conditional independence is approximately true if the SP is replaced by either the SIR or LR ESP.

To check whether the LR model is good, consider the symbol density of +'s and 0's in a narrow vertical slice where 0 is used if $Y = 0$ and + is used if $Y = 1$. This symbol density should be approximately constant (up to binomial variation) from the bottom to the top of the slice. (Hence the +'s and 0's should be mixed in the slice.) The plot would be easier to interpret if the LR ESP was used on the horizontal axis instead of the SIR ESP since then the approximate probability of the symbol + could be computed. For example, if there are n_J points in narrow slice J and if the ESP ≈ -5, then the points in the slice resemble a sample of n_J cases from a binomial (1, 0.007) distribution and almost none of the points should be +'s. If the ESP ≈ 0, then the points resemble a sample of n_J cases from a binomial (1, 0.5) distribution and about half of the points should be

+'s. Moreover the proportion of +'s should be near 0.5 in the bottom, middle and top of the narrow slice. If the ESP ≈ 5, then the points resemble a sample of n_J cases from a binomial (1, 0.993) distribution and almost all of the points should be +'s.

The symbol density often changes greatly as the narrow slice is moved from the left to the right of the plot, e.g. from 0% to 100% if the correlation between the SIR and LR ESP is near 1 or from 100% to 0% if the correlation is near -1. If there are one or more wide slices where the symbol density is not constant from top to bottom, then the LR model may not be good (e.g. a more complicated model may be needed). If it is difficult to quickly find slices where the symbol density is not mixed, then the binary response plot should not be used as evidence that the model is bad. If only a few isolated points need to be changed to make a good plot, then the model is often good and the points correspond to an unlikely outcome; however, the isolated points could be "outliers" in that $\boldsymbol{x}$ is outlying or the value of Y was misclassified.

Figure 6 shows the binary response plot for the artificial data. The correlation between the SIR and LR ESP's was near -1. Hence the slice symbol density of +'s decreases from nearly 100% in the left of the plot to 0% in the right of the plot. The symbol density is mixed in most of the slices and the plot looks good. For contrast, Figure 7 shows the binary response plot when only X_2 and X_5 are in the model. Consider the slice where the ESP is between -2.4 and -1.7. At the bottom and top of the slice the proportion of +'s is near 1 but in the middle of the slice there are several 0's. In the slice where the ESP is between -1.7 and -0.8, the proportion of +'s increase as one moves from the bottom of the slice to the top of the slice. Hence there is a large slice from about -2.4

to -0.8 where the plot does not look good, suggesting that the logistic regression model may be poor.

3.2 Loglinear Regression

Figure 8 shows the ESSP for the LLR data with the estimated mean function

$$\hat{\mu}(ESP) = \exp(ESP)$$

added as a visual aid. This plot is very useful as a goodness of fit diagnostic. The lowess curve is a nonparametric estimator of the mean function called a "scatterplot smoother." The lowess curve is represented as a jagged curve to distinguish it from the estimated LLR mean function (the exponential curve) in Figure 8. If the lowess curve follows the exponential curve closely (except possibly for the largest values of the ESP), then the LLR model fits the data well.

Simple diagnostic plots for the loglinear regression model can also be made using weighted least squares (WLS). Let $Z_i = Y_i$ if $Y_i > 0$, and let $Z_i = 0.5$ if $Y_i = 0$. Then the minimum chi–square estimator of the parameters $(\alpha, \boldsymbol{\beta})$ in a loglinear regression model are $(\hat{\alpha}_M, \hat{\boldsymbol{\beta}}_M)$, and are found from the weighted least squares regression of $\log(Z_i)$ on $\boldsymbol{x}_i$ with weights $w_i = Z_i$. Equivalently, use the ordinary least squares (OLS) regression (without intercept) of $\sqrt{Z_i}\log(Z_i)$ on $\sqrt{Z_i}\boldsymbol{x}_i$. The minimum chi–square estimator tend to be consistent if n is fixed and all n counts Y_i increase to ∞ while the loglinear regression maximum likelihood estimator tends to be consistent if the sample size $n \to \infty$. See Agresti (2002, p. 611-612) and Powers and Xie (2000, p. 284). However, the two estimators are often close for many data sets. This result and the equivalence of the minimum chi–square estimator to an OLS estimator suggest the following diagnostic plots. Let $(\tilde{\alpha}, \tilde{\boldsymbol{\beta}})$ be an estimator of $(\alpha, \boldsymbol{\beta})$.

For a loglinear regression model, a *weighted forward response plot* is a plot of $\sqrt{Z_i} ESP = \sqrt{Z_i}(\tilde{\alpha} + \tilde{\boldsymbol{\beta}} \boldsymbol{x}_i)$ versus $\sqrt{Z_i} \log(Z_i)$. The *weighted residual plot* is a plot of $\sqrt{Z_i}(\tilde{\alpha} + \tilde{\boldsymbol{\beta}} \boldsymbol{x}_i)$ versus the WLS residuals $r_{Wi} = \sqrt{Z_i} \log(Z_i) - \sqrt{Z_i}(\tilde{\alpha} + \tilde{\boldsymbol{\beta}} \boldsymbol{x}_i)$.

If the loglinear regression model is appropriate and if the minimum chi–square estimators are reasonable, then the plotted points in the weighted forward response plot should follow the identity line. Cases with large WLS residuals may not be fit very well by the model. When the counts Y_i are small, the WLS residuals can not be expected to be approximately normal.

Figure 9 shows the diagnostic plots for the artificial data using both the minimum chi–square estimator and the LLR maximum likelihood estimator. Even though the counts Y_i are small for this data set, the points in both weighted forward response plots follow the identity line, and neither residual plot has outlying residuals. Also notice that the larger counts are fit better than the smaller counts and hence the residual plots have a "left opening megaphone" shape. More research is needed to determine if these plots are useful for contingency tables.

3.3 A Diagnostic for Testing $H_o : \boldsymbol{\beta} = \boldsymbol{0}$

The ESSP is also a useful visual aid for describing the ANOVA F or deviance test for $H_o : \boldsymbol{\beta} = \boldsymbol{0}$ versus $H_A : \boldsymbol{\beta} \neq \boldsymbol{0}$, that is, the test for whether the predictors $\boldsymbol{x}$ are needed in the given model. For MLR, LLR and the binary LR models, if the predictors are not needed in the model, then $E(Y_i|\boldsymbol{x}_i)$ should be estimated by the sample mean $\overline{Y}$. If the predictors are needed, then $E(Y_i|\boldsymbol{x}_i)$ should be estimated by the appropriate function of the $ESP = \hat{\alpha} + \hat{\boldsymbol{\beta}}^T \boldsymbol{x}_i$. If it is clear that no horizontal line fits either the data or the estimated nonparametric mean function as well as the estimated mean function (as in Figures 4, 5 and 8), then the predictors are needed.

Figures 10, 11 and 12 show the ESSP when only X_4 and X_5 were used as predictors for the MLR, LR and LLR data sets respectively. Since Y is independent of these two predictors by construction, the horizontal line should fit either the data or the nonparametric estimated mean function about as well as the estimated mean function. In Figure 10, the horizontal line $Y = \overline{Y}$ fits the data about as well as the identity line. In Figure 11, the horizontal line fits the step function about as well as the logistic curve, suggesting that $\hat{\rho}(\boldsymbol{x}_i) \equiv \hat{\rho} = \overline{Y}$ should be used instead of the LR estimator

$$\hat{\rho}(\boldsymbol{x}_i) = \frac{\exp(\hat{\alpha} + \hat{\boldsymbol{\beta}}^T \boldsymbol{x}_i)}{1 + \exp(\hat{\alpha} + \hat{\boldsymbol{\beta}}^T \boldsymbol{x}_i)}.$$

In Figure 12, the horizontal line fits the lowess curve about as well as the exponential curve, suggesting that $\hat{\mu}(\boldsymbol{x}_i) \equiv \hat{\mu} = \overline{Y}$ should be used instead of the LLR estimator

$$\hat{\mu}(\boldsymbol{x}_i) = \exp(\hat{\alpha} + \hat{\boldsymbol{\beta}}^T \boldsymbol{x}_i).$$

It is easy to find data sets where the ESSP looks like Figure 10, 11 or 12 but the p–value for the ANOVA F or deviance test is very small. In this case, the model is statistically significant, but the investigator needs to decide whether the model is practically significant.

4 Conclusions

Plots for goodness of fit and lack of fit should be made immediately after fitting a 1D regression model and before performing inference. The plots in Section 3 are illustrative, and work for other models as well. For example, if a single index model holds where m is unknown, the OLS estimator $\hat{\boldsymbol{b}} \approx c\boldsymbol{\beta}$ for many data sets (see Li and Duan, 1989). The ESSP is a plot of $\hat{\boldsymbol{b}}^T \boldsymbol{x}_i$ versus Y_i and can be used to visualize m. Add a lowess smooth to the plot. If no horizontal line fits the data as well as the lowess smooth, then $\boldsymbol{x}$ is needed in the single index model. See Simonoff and Tsai (2002) for formal tests of hypotheses.

Similar plots may be useful in other settings. For example, if the reduced model is good, then the *EE plot* of $ESP(R) = \hat{\alpha}_R + \hat{\boldsymbol{\beta}}_R^T \boldsymbol{x}_{Ri}$ versus $ESP = \hat{\alpha} + \hat{\boldsymbol{\beta}}^T \boldsymbol{x}_i$ should follow the identity line closely. As a rule of thumb, models with $\text{corr}(ESP, ESP(R)) > 0.95$ are interesting. After performing variable selection, often there are several competing submodels that all look good. Making a scatterplot matrix of the different ESP's along with the response Y may be useful.

Of course the plots need to be used with caution. If the number of parameters k is too large, then the model may "overfit" the data. If all of the predictors are factors, as in experimental design and contingency tables, the ESP may not take on many values and the interpretation of the plots changes. If the model or fitting method (e.g. OLS or SIR or maximum likelihood) is inappropriate, then the ESSP may be greatly inferior to an ESSP made with appropriate methods.

Using the ESSP as a goodness of fit diagnostic is not new. Several authors suggested using the forward response plot to visualize the coefficient of determination R^2 in MLR.

For example, see Chambers, Cleveland, Kleiner and Tukey (1983, p. 280). Brillinger (1983) suggested that the forward response plot of the OLS fitted values $\hat{\alpha} + \hat{\boldsymbol{\beta}}^T \boldsymbol{x}_i$ versus Y_i can be used to look for the functional form m for single index models while Cook and Weisberg (1997, 1999, p. 397, 514, and 541) call the ESSP a model checking plot. Cook and Weisberg (1999, p. 121) adds a horizontal line to a plot of x_i versus Y_i as a diagnostic for $\beta = 0$ in the simple linear regression model.

5 References

Agresti, A. (2002), *Categorical Data Analysis*, 2nd ed., John Wiley and Sons, Inc., Hoboken, NJ.

Brillinger, D.R. (1983), "A Generalized Linear Model with "Gaussian" Regressor Variables," in *A Festschrift for Erich L. Lehmann,* eds. Bickel, P.J., Doksum, K.A., and Hodges, J.L., Wadsworth, Pacific Grove, CA, 97-114.

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

Collett, D. (1991), *Modelling Binary Data,* Chapman & Hall/CRC, Boca Raton, Florida.

Cook, R.D. (1996), "Graphics for Regressions with Binary Response," *Journal of the American Statistical Association,* 91, 983-992.

Cook, R.D. (1998), *Regression Graphics: Ideas for Studying Regression Through Graphics,* John Wiley and Sons, Inc., NY.

Cook, R.D. (2004), "Testing Predictor Contributions in Sufficient Dimension Reduction," *The Annals of Statistics,* To Appear.

Cook, R.D., and Weisberg, S. (1997), "Graphs for Assessing the Adequacy of Regression Models," *Journal of the American Statistical Association,* 92, 490-499.

Cook, R.D., and Weisberg, S. (1999), *Applied Regression Including Computing and Graphics,* John Wiley and Sons, Inc., NY.

Horowitz, J.L. (1998), *Semiparametric Methods in Econometrics,* Springer-Verlag, NY.

Hosmer, D.W., and Lemeshow, S. (1980), "A Goodness of Fit Test for the Multiple Logistic Regression Model," *Communications in Statistics,* A10, 1043-1069.

Hosmer, D.W., and Lemeshow, S. (2000), *Applied Logistic Regression*, 2nd ed., John Wiley and Sons, Inc., NY.

Li, K.C. (1991), "Sliced Inverse Regression for Dimension Reduction," *Journal of the American Statistical Association,* 86, 316-342.

Li, K.C., and Duan, N. (1989), "Regression Analysis Under Link Violation," *The Annals of Statistics,* 17, 1009-1052.

Powers, D.A., and Xie, Y. (2000), *Statistical Methods for Categorical Data Analysis,* Academic Press, San Diego.

Simonoff, J.S. and Tsai, C.L. (2002), "Score Tests for the Single Index Model," *Technometrics,* 44, 142-151.

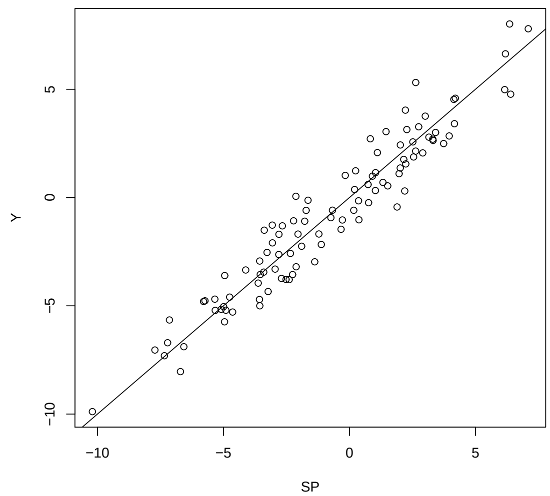

Figure 1: SSP for MLR Data

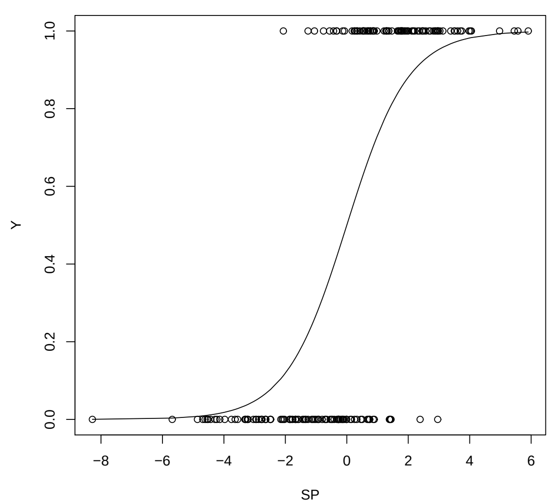

Figure 2: SSP for LR Data

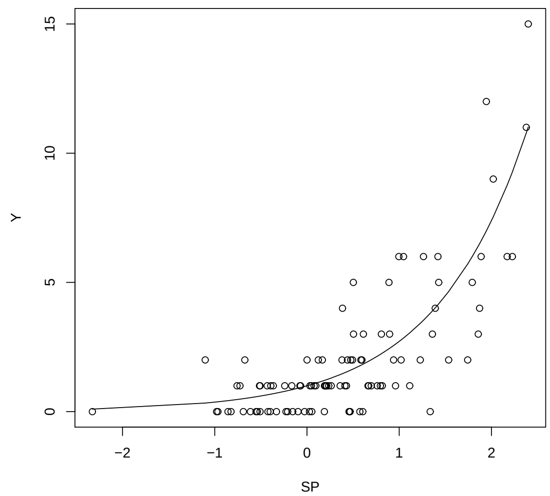

Figure 3: SSP for LLR Data

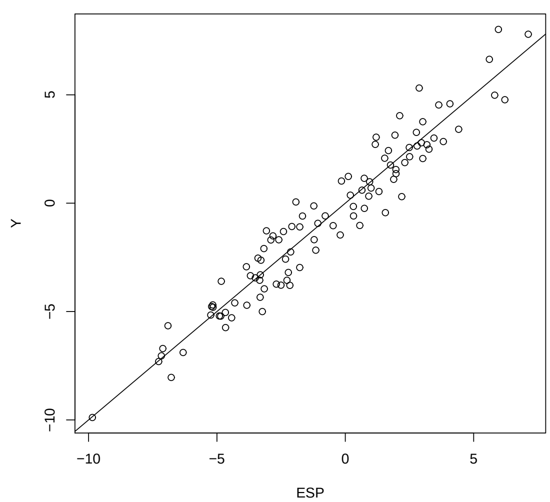

Figure 4: Forward Response Plot for MLR Data

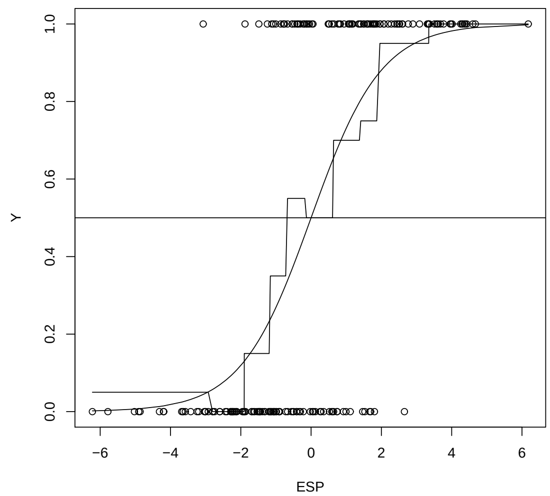

Figure 5: ESS Plot for LR Data

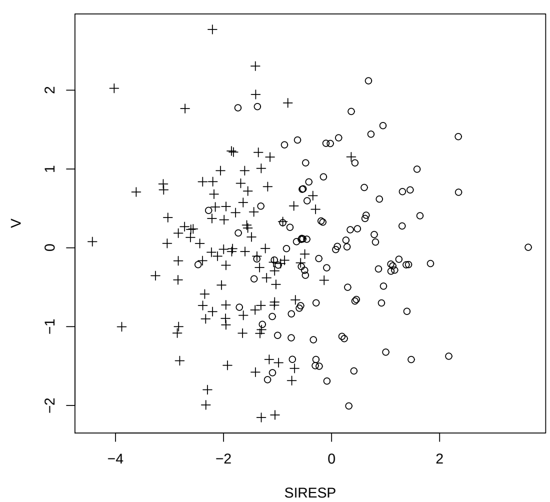

Figure 6: The Binary Response Plot for a Good LR Model

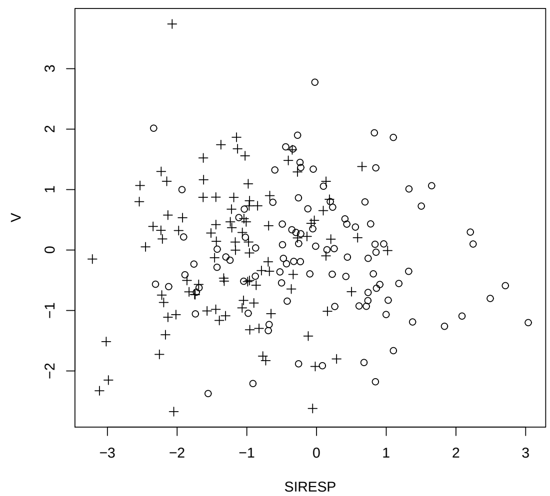

Figure 7: The Binary Response Plot for a Poor LR Model

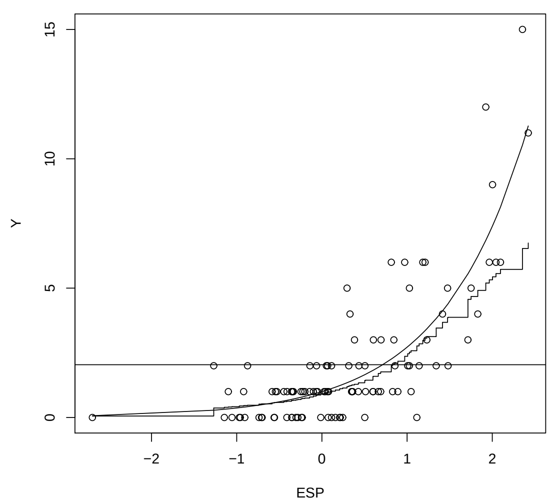

Figure 8: ESSP for Loglinear Regression

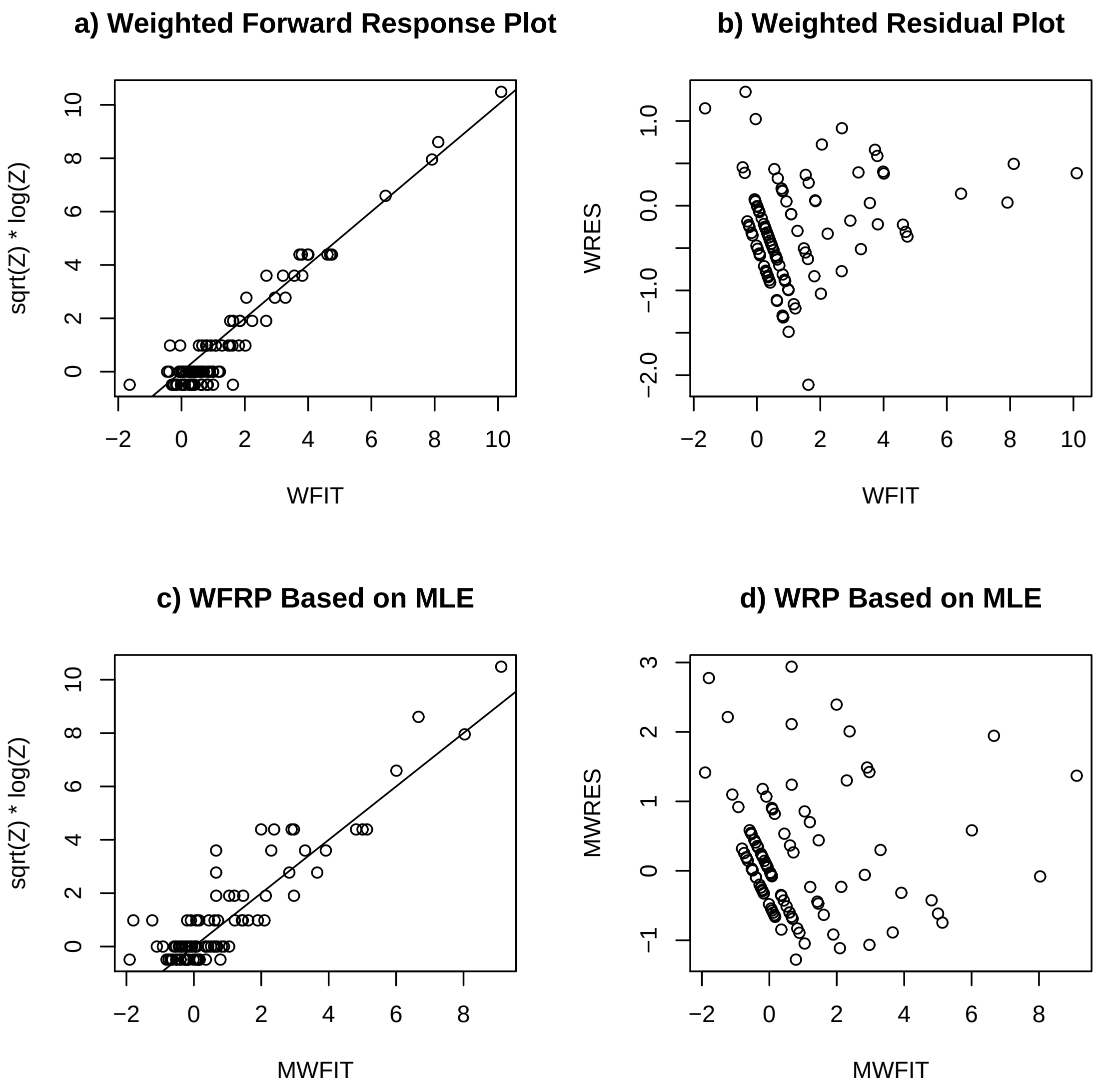

Figure 9: Diagnostic Plots for Loglinear Regression

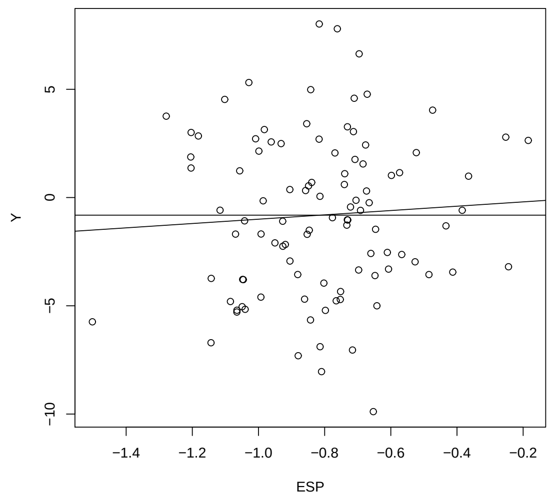

Figure 10: Forward Response Plot when Y is Independent of the Predictors

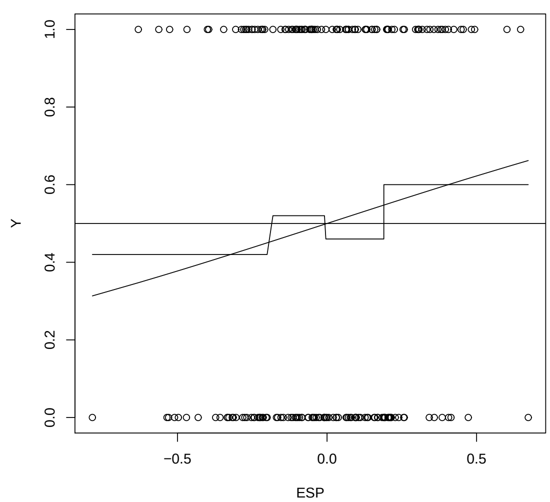

Figure 11: LR ESS Plot When Y Is Independent Of The Predictors

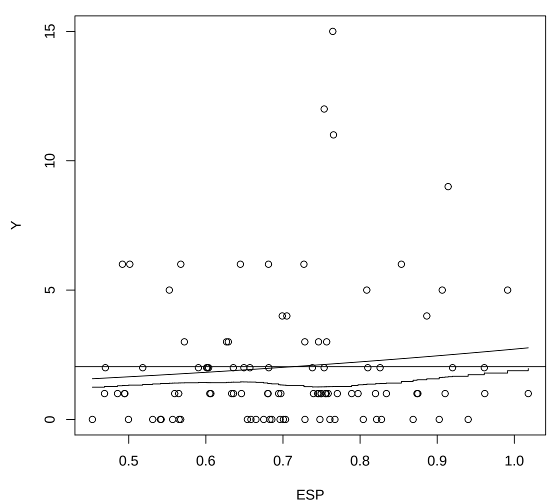

Figure 12: LLR ESSP when Y is Independent of the Predictors