

A GRAPHICAL METHOD FOR ASSESSING THE FIT OF A REGRESSION MODEL

Iain Pardoe, Department of Decision Sciences, Charles H. Lundquist College of Business,
1208 University of Oregon, Eugene, OR 97403–1208

Key Words: Bayesian methods, Model criticism, Predictive distribution, Graphics, Diagnostic plot, Marginal model plot

Abstract

Before a regression model is used to address questions about the relationship between a response variable and predictors, the fit of the model to the data should be assessed. For example, consider a logistic regression model for explaining the dependence of a binary outcome variable on a set of predictor variables or for predicting the outcome variable based on the predictors. Checking the fit of the model before it is used in a practical setting is of critical importance. If a model is found to be deficient, the nature of the deficiency may indicate a need for some aspect of the model to be reformulated or that poorly fitting observations need to be considered separately. I propose graphical methodology based on a Bayesian framework to help address issues such as this. Plots can be constructed quickly and easily for any model of interest, and goodness of fit assessed. These plots are more intuitive and easy-to-use than traditional graphical diagnostic methods for regression such as residual plots.

1 Introduction

Cook and Pardoe (2000) suggested a graphical technique for assessing the fit of a regression model which they called a “Gibbs marginal model plot.” This methodology was developed for linear and additive models in Pardoe (2001b), in which the plots were renamed “Bayes marginal model plots” (BMMPs). The plots provide a way for visualizing model uncertainty in the “marginal model plots” (MMPs) of Cook and Weisberg (1997).

This article describes the BMMP methodology in the context of a binary logistic regression analysis. Section 2 highlights the issues involved in assessing the fit of logistic regression models; it also introduces an example dataset on breast cancer diagnosis that will be used to illustrate the proposed methodology, and reviews how MMPs can help diagnose the fit of a model. Without guidance on the level of uncertainty in the model, MMPs can be difficult to interpret however—Section 3

provides details on how Bayesian model checking ideas can be used to address this problem, and illustrates how BMMPs can guide model improvement for the breast cancer data. Section 4 contains a discussion.

2 Background

How can we assess the fit of a regression model used to explain the dependence of y on \mathbf{x} or to predict y from \mathbf{x} ? Let the unknown conditional distribution of y given \mathbf{x} be represented by its cumulative distribution function, $F(y|\mathbf{x})$. Suppose we have derived a model for $F(y|\mathbf{x})$, and denote this by its cumulative distribution function, $M_{\theta}(y|\mathbf{x})$, where θ is a vector of unknown parameters.

For a frequentist analysis, assume that θ can be consistently estimated under $M_{\theta}(y|\mathbf{x})$ with $\hat{\theta}$. For a Bayesian analysis, assume that inference will be based on a posterior distribution for the model denoted by $M_{\theta}(y|\mathbf{x}, \mathbf{y}_d)$, where \mathbf{y}_d is the n -vector of data, that is the observed responses. Before using $M_{\hat{\theta}}(y|\mathbf{x})$ or $M_{\theta}(y|\mathbf{x}, \mathbf{y}_d)$ to address a practical issue, we need to be confident that the model provides a *sufficiently accurate* approximation to $F(y|\mathbf{x})$, where the accuracy is gauged relative to the practical issue. If the model is found to be deficient, the nature of the deficiency may indicate a need for some aspect of the model to be reformulated or that poorly fitting observations need to be considered separately. However, identifying the nature of a model deficiency for logistic regression is not an easy task (see Pregibon, 1981; Landwehr, Pregibon, and Shoemaker, 1984).

An example of a logistic model assessment problem is the “Wisconsin Breast Cancer Data” (Bennett and Mangasarian, 1992). This consists of 681 cases of potentially cancerous tumors in Wisconsin in the 1980s, 238 of which turned out to be malignant, and 443 of which were benign. Determining whether a tumor is malignant or benign is traditionally accomplished with an invasive surgical biopsy procedure. An alternative, less invasive technique, allowing examination of a small amount of tissue from the tumor, is “Fine Needle Aspiration” (FNA). FNA provides nine cell features for each case; the biopsy determines the tumor status as malignant or benign.

Features of the tissue cells can be used as predictors in a model with tumor status as the response. The hope is to

use the model to successfully predict tumor status based only on the FNA predictors. Of critical importance is whether the model can provide an accurate alternative to the biopsy procedure for future patients.

The dataset consists of response $y = \text{Class1} = 0$ if malignant, 1 if benign, and predictor variables: $x_1 = \text{Adhes} = \text{marginal adhesion}$, $x_2 = \text{BNucl} = \text{bare nuclei}$, $x_3 = \text{Chrom} = \text{bland chromatin}$, $x_4 = \text{Epith} = \text{epithelial cell size}$, $x_5 = \text{Mitos} = \text{mitoses}$, $x_6 = \text{NNucl} = \text{normal nucleoli}$, $x_7 = \text{Thick} = \text{clump thickness}$, $x_8 = \text{UShap} = \text{cell shape uniformity}$, and $x_9 = \text{USize} = \text{cell size uniformity}$.

The predictors, $\mathbf{x} = (x_1, \dots, x_9)^T$, are all integers between one and ten (one represents a normal state, ten an abnormal one), and are determined by a doctor assessing the tissue cells through a microscope. Together, the predictors provide a wealth of information on tumor status. In fact, it appears that a subset of the predictors can provide nearly all the information available. Subset selection on the full set of nine predictors, removing the least significant predictor at each stage, leads to the following model worthy of consideration:

Model 1: Five predictors (Adhes, BNucl, Chrom, NNucl and Thick).

Some traditional numerical measures of fit include Wald p -values for predictors in the model each less than 0.0005, p -values for adding one more predictor each greater than 0.05, and residual deviance of 96.5 on 675 degrees of freedom. Based on these numbers, the model appears to fit well. However, perhaps plots of the data can give us further information on the fit of this model. Figure 1 shows two residual plots with non-parametric smooths superimposed, one versus *Mitos*, a predictor *not* in the model, and the other versus the linear fit from the model. Residual plots are the traditional graphical method for assessing lack of fit of a regression model. The idea is to look at the residuals plotted against functions of the predictors in a series of two-dimensional scatterplots, and look for patterns that suggest violation of assumptions in the model.

The top plot in Figure 1 is a residual plot with horizontal axis equal to *Mitos*. If there are any unexpected patterns in this plot, then perhaps *Mitos* could usefully be added to the model. One problem with using residual plots in generalized linear models is that it can be difficult figuring out what kinds of patterns are unexpected and which are entirely to be expected. In this particular context, the main unexpected pattern to look for is a non-constant mean function in the plot. Here there appears to be nothing unusual since the smooth of the residuals is flat, so there is nothing to suggest that *Mitos* could usefully be included.

The lower plot has horizontal axis equal to the linear fit from the model. This looks a little strange, and can be

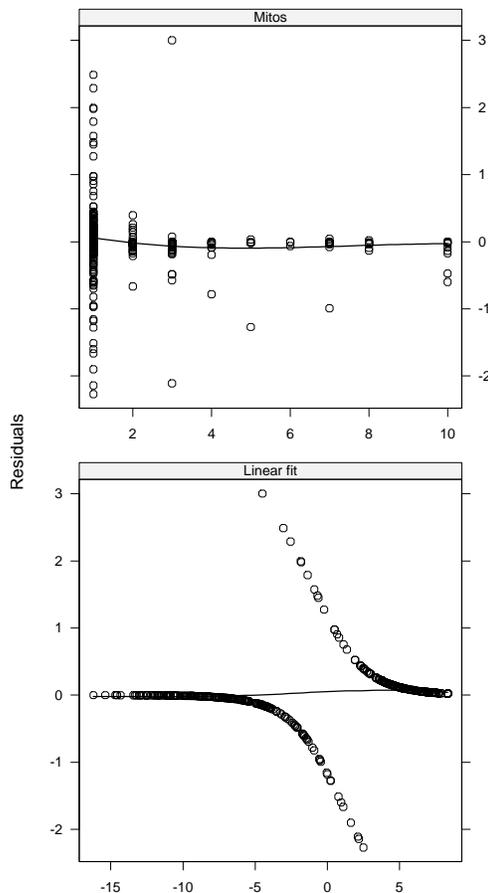


Figure 1: Residual plots for model 1.

hard to interpret. In particular, the way the residuals fall on two distinct curves is an *expected* pattern due entirely to the fact that the response values are either zero or one, and the fitted probabilities are a non-linear but monotone function of the linear fit. But again the smooth of the residuals is essentially a horizontal line, so there is no evidence of lack of fit from this plot either.

Alternatively, visualize *goodness* of fit in a marginal model plot (MMP), for example the MMP with horizontal axis $h = \text{Mitos}$ in Figure 2. Cook and Weisberg (1997) introduced these plots from a frequentist perspective, in which the solid line is a smooth of the data and the dashed line is a smooth of the fitted values from the model. The rationale for the MMP is this statement:

$$E_F(y|\mathbf{x}) = E_{\hat{M}}(y|\mathbf{x}), \quad \forall \mathbf{x} \in \mathcal{X} \quad (1)$$

$$\iff E_F(y|h) = E_{\hat{M}}(y|h), \quad \forall h = h(\mathbf{x}) \quad (2)$$

where E_F denotes expectation under F , $E_{\hat{M}}$ denotes expectation under $M_{\hat{\theta}}$, and \mathcal{X} is the sample space of \mathbf{x} : think of \mathbf{x} here in the same way that it is thought about in subset selection, i.e. predictors that are included in the model being considered, as well as potential predictors

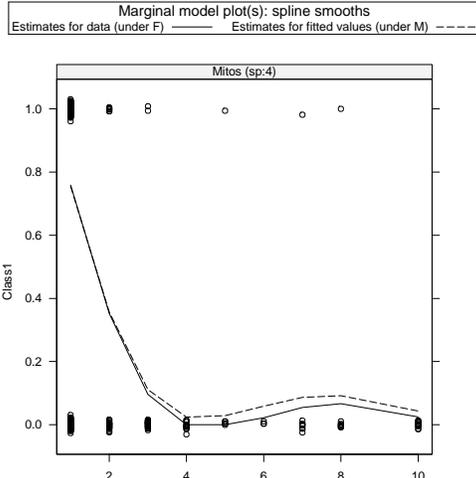


Figure 2: MMP for the mean, $h = \text{Mitos}$, model 1.

not in the current model. This result requires only that h be measurable with respect to \mathbf{x} .

Equality (1) is what we would like to check, but if the dimension of \mathbf{x} is greater than two, then $E(y|\mathbf{x})$ is difficult to visualize. However, if h is univariate, then $E(y|h)$ can be visualized in a 2-D scatterplot, and equality (2) can be checked. So, the idea in a MMP is to compare $E_F(y|h)$ and $E_{\hat{M}}(y|h)$ for various h to gain information about the relationship between $E_F(y|\mathbf{x})$ and $E_{\hat{M}}(y|\mathbf{x})$. The mean function based on F can be considered *model-free*, while the function using \hat{M} is *model-based*.

We can estimate the two mean functions with smooths. Obtain $\hat{E}_F(y|h)$ by smoothing y versus h using a non-parametric smooth such as a cubic smoothing spline. The corresponding model-based estimate of the mean function uses the relationship $E_{\hat{M}}(y|h) = E[E_{\hat{M}}(y|\mathbf{x})|h]$. So, obtain $\hat{E}_{\hat{M}}(y|h)$ by smoothing $E_{\hat{M}}(y|\mathbf{x})$ versus h ; note that $E_{\hat{M}}(y|\mathbf{x})$ is the (assumed) mean function from the fitted model, i.e. the *fitted values* from the model. Superimpose $\hat{E}_F(y|h)$ and $\hat{E}_{\hat{M}}(y|h)$ on a plot of y versus h to obtain a MMP for the mean in the (marginal) direction h . Using the same method and smoothing parameter for the mean function estimates under F and M allows pointwise comparison of the two estimates, since any estimation bias should approximately cancel. See Bowman and Young (1996) for elaboration of this point.

Ideas for selecting useful functions h to consider in practice are given in Cook and Weisberg (1997), and include fitted values, individual predictors in the model, potential predictors not in the model, linear combinations of the predictors, and random linear projections of the predictors. Other possibilities include functions h where lack of fit is most likely to be observed.

Now, if M is an accurate approximation to F , then for any quantity h the marginal mean function estimates

should agree, $\hat{E}_F(y|h) \approx \hat{E}_{\hat{M}}(y|h)$. Any indication that the estimated marginal mean functions do not agree for one particular h calls M into question; if they agree for a variety of plots, there is support for M .

So, how should Figure 2 be interpreted? In this plot, the smoothing splines have four effective degrees of freedom and the points have been jittered to aid visualization of data density. Most of the data is on the left where $\text{Mitos} = \text{one or two}$, and here the smooths match well. But, for Mitos three or higher, the model seems to predict higher probabilities of a tumor being benign than the data indicate. But, is the gap between the smooths so large that we should be concerned, or so small that we can just put it down to random variation?

The same issue of variability arises in residual plots also. The deviance, which is essentially a numerical summary of a residual plot for a logistic regression, provides one way to address this issue. However, if a model is identified as poorly fitting due to a high deviance in relation to the error degrees of freedom, there is no guidance available on how to improve the model. It would be helpful to *see* the nature of the lack-of-fit in a graphical display, and in this respect the MMP is to be preferred over the residual plot since it is easier to interpret.

Even if $M_\theta(y|\mathbf{x}) = F(y|\mathbf{x})$, the estimated marginal mean function estimates in an MMP would not match exactly. From a frequentist perspective, the data can be thought of as just one realization of many possible samples. So, a possible solution to the problem of comparing the estimates is to calculate a sampling-theory confidence band or perhaps generate replicate data by bootstrapping. Alternatively, from a Bayesian perspective, the data are fixed, but the variability in the model estimates is given explicitly by posterior distributions for the parameters. A possible solution to the assessment problem displays this variability in the model smooth, allowing the analyst to more easily judge whether it would be reasonable for the data to be generated by the particular model in question.

3 Bayes marginal model plots

To introduce ideas and keep notation concise, consider assessing how well a model $M = M(y|\theta)$ fits potential data $\mathbf{y} = (y_1, \dots, y_n)^T$, where θ is assumed to have a prior probability distribution. Box (1980) proposed a Bayesian diagnostic for checking M based on the marginal, or predictive, distribution of \mathbf{y} . He suggested assessing M by referring the value of the predictive density for the observed data, $f(\mathbf{y}_d|M)$, to the density function $f(\mathbf{y}|M)$, by calculating a tail area, say. A “small” tail area indicates that \mathbf{y}_d would be unlikely to have been generated by M , and thus calls M into question. More

generally, M can be assessed by referring the value of the predictive density of some relevant checking function, $g(\mathbf{y})$, at \mathbf{y}_d to its predictive density, for a variety of g . Examples of useful g in practice include residuals, order statistics, and moment estimators.

Rubin (1984) proposed an alternative approach that does not require proper priors, as Box's approach does, using the posterior predictive density

$$f(\mathbf{y}|\mathbf{y}_d, M) = \int f(\mathbf{y}|\boldsymbol{\theta}, M)\pi(\boldsymbol{\theta}|\mathbf{y}_d, M) d\boldsymbol{\theta}$$

where $f(\mathbf{y}|\boldsymbol{\theta}, M)$ is the likelihood for \mathbf{y} and $\pi(\boldsymbol{\theta}|\mathbf{y}_d, M)$ is the posterior density of $\boldsymbol{\theta}$. The posterior predictive distribution of \mathbf{y} can be thought of as a distribution for potential data that we might observe, if the model that we think produced \mathbf{y}_d , including the particular $\boldsymbol{\theta}$ value, was used to produce a new set of data. Since this particular $\boldsymbol{\theta}$ value is unknown, average over plausible values using its posterior distribution. Again, diagnostics similar to Box's tail area and checking functions g can be constructed. Use of the posterior predictive distribution in a goodness of fit test was first proposed by Guttman (1967). Rubin's approach has been extended by Gelman, Meng, and Stern (1996) to allow the checking function g to depend on $\boldsymbol{\theta}$ and nuisance parameters as well as on \mathbf{y} .

Another way to think about Rubin's approach is in terms of a sampling simulation. Gelman et al. (1996) provide references to many papers that discuss this interpretation. The idea is to draw a value of $\boldsymbol{\theta}$ from its posterior distribution, and then generate a sample of n realizations from the model M indexed by this $\boldsymbol{\theta}$. Repeat this process a large number m of times and then compare the data \mathbf{y}_d to the m realizations from M . Then, intuitively, if \mathbf{y}_d "looks like" a typical realization from M , there is no reason to doubt the fit of M . On the other hand, if \mathbf{y}_d appears to be very "unusual" with respect to the m realizations from M , then M is called into question. To do this in practice, methods for comparing \mathbf{y}_d to the m realizations from M and measures of "unusualness" need to be developed. But once done, the methodology can be applied in any situation where samples can be generated from the posterior distributions for $\boldsymbol{\theta}$.

A graphical way to do this is based on the MMPs introduced earlier. In regression, $\boldsymbol{\theta}$ provides "fitted values". So, instead of sampling y , compare model-free predicted values with expected y -values based on sampled $\boldsymbol{\theta}$ values. A Bayes marginal model plot (BMMP) is a scatterplot of y versus h with a mean function estimate under F superimposed. Then, superimpose a mean function estimate for each model sample $M_{\boldsymbol{\theta}_t}$, $t = 1, \dots, m$.

Recall that the smoothing parameters for the smooths in a particular MMP need to be equal to allow their pointwise comparison. Similarly, the smooths in a BMMP should all have the same smoothing parameter, γ . There-

fore it is desirable to select γ so that the smooths are flexible enough to capture systematic trends in all the corresponding scatterplots, while not over-fitting too much in any one scatterplot. This is clearly impractical, so a prudent compromise is to graphically select γ to capture the systematic trends only for the mean functions in the scatterplots for the data and for the model fitted values.

If enough samples are taken, say $m = 100$, the Bayes mean function estimates, $\hat{E}_{M_{\boldsymbol{\theta}_t}}(y|h)$, $t = 1, \dots, m$, will form a mean function *band* under M . The plot then provides a visual way of determining whether there is any evidence to contradict the possibility that $F(y|x) = M(y|x)$. If, for a particular h , the mean function estimate under F lies *substantially outside* the mean function band under M or it does not follow the general pattern shown by the model smooths, then M is called into question. If, no matter what the function h is, the mean function estimate under F lies *broadly inside* the mean function band under M and it follows the general pattern shown by the model smooths, then perhaps M provides an accurate description of the conditional distribution of $y|x$ and is a useful model.

The binary logistic regression model can be written

$$\begin{aligned} y_i | (x_i, p_i) &\sim \text{Bernoulli}(p_i) \\ p_i &= \Pr(y = 1 | x_i) = E(y | x_i) \\ \text{logit}(p_i) &= \log\left(\frac{p_i}{1 - p_i}\right) = \boldsymbol{\theta}^T \mathbf{x}_i \end{aligned}$$

One possible prior for this example is $\boldsymbol{\theta} \sim N(\mathbf{0}_p, k\mathbf{I}_p)$, where k can be set to reflect the degree of prior uncertainty for any particular dataset. It is not possible to sample directly from the posterior, so instead Markov chain simulation can be used to obtain the samples. In particular, posterior samples can be obtained by Gibbs sampling using "BUGS" software (Spiegelhalter, Thomas, Best, and Gilks, 2000). Checking convergence in Markov chain sampling is very important, and software is available from various sources to assist in this task, for example, some software that works well with BUGS output is "BOA" (Smith, 2000).

Constructing a BMMP for the mean in direction h requires model-free and model-based estimates of the mean function with respect to h . To obtain the model-free estimate $\hat{E}_F(y|h)$, smooth the data $\{y_i\}$ on $\{h_i\}$. To obtain the model-based estimates $\hat{E}_{M_{\boldsymbol{\theta}_t}}(y|h)$, smooth the fitted-values based on the posterior samples $\{E_{M_{\boldsymbol{\theta}_t}}(y|x_i)\}$ on $\{h_i\}$. The fitted values corresponding to posterior samples $\boldsymbol{\theta}_t$ are

$$\begin{aligned} E_{M_{\boldsymbol{\theta}_t}}(y|x_i) = p_{it} &= \frac{1}{1 + \exp(-\boldsymbol{\theta}_t^T \mathbf{x}_i)} \\ i &= 1, \dots, n; t = 1, \dots, m \end{aligned}$$

The BMMP equivalent to Figure 2 is shown in Figure 3. In this plot, the prior uncertainty parameter, k ,

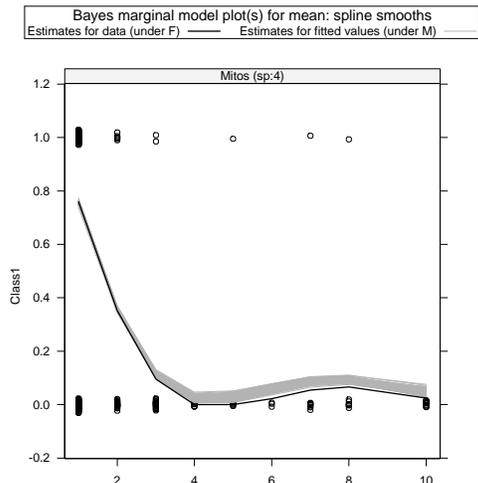


Figure 3: BMMP for the mean, $h = Mito$, model 1.

was set to be 10^6 , the smoothing splines have four effective degrees of freedom, and $m = 100$. Here, the black smooth of the data lies below the gray band of the fitted probability smooths for values of *Mito* three or higher. *Mito* clearly adds information on the probability of being benign not provided by the five predictors in the model. This plot, in contrast to a residual plot, can be interpreted straightforwardly, incorporates model uncertainty, and provides guidance on model improvement. Recall that the Wald p -values for adding one more predictor to the model were each greater than 0.05; the BMMP tells us that we should not be so hasty in not considering adding *Mito* because of this. So, let's add *Mito* to the model to see what happens.

Model 2: Six predictors (five from model 1 plus *Mito*).

For this model, the BMMP for the mean with $h = Mito$ is the upper plot of Figure 4. In this plot, the smoothing splines again have four effective degrees of freedom. This plot shows a big improvement over Figure 3, so it appears that adding *Mito* to the model is useful. However, recall that a *series* of BMMPs needs to be considered in order to be confident in the model. So, how about the BMMP for the mean with $h =$ the linear fit from the model? This is the lower plot of Figure 4. In this plot, the smoothing splines have twelve effective degrees of freedom—increased flexibility in the smooths is needed to fit the “logistic curve” shape of the fitted probabilities. The black smooth of the data lies mostly inside the gray band of the fitted probability smooths, but it gets very close to the edge of the band at one point. The model appears to fit most of the data very well, but has trouble with cases “in the middle” when the linear fit is close to

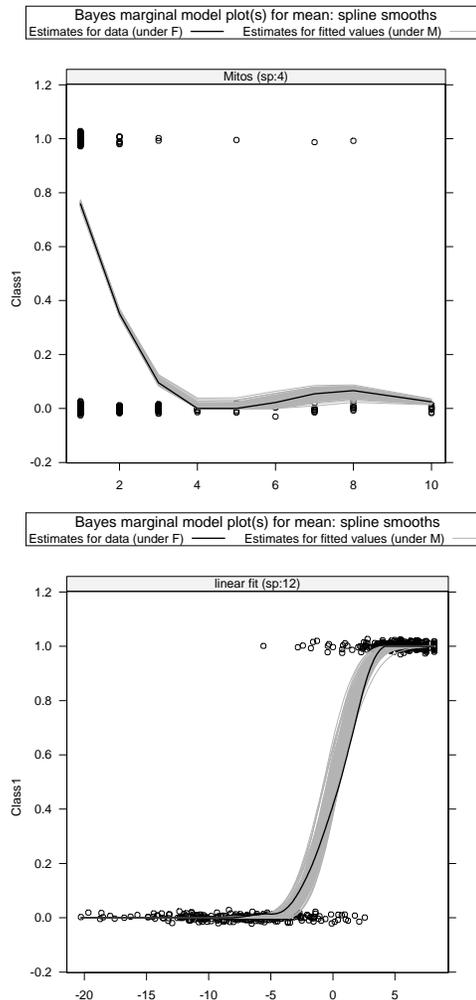


Figure 4: BMMP for the mean, model 2: $h = Mito$ (upper) and $h =$ the linear fit (lower).

zero. Should we worry when the data smooth is close to the edge of the “band” of model smooths in places as it is in this plot? The answer to this question is probably context dependent. Nonetheless, BMMPs have taken us much further in model assessment and understanding for this dataset than either residual plots or MMPs.

4 Discussion

BMMPs offer a quick and easy way to check models graphically. The sampling needs to be done only once for each model and cycling through BMMPs in a variety of directions h provides guidance on the fit of the model.

A discrepancy measure could perhaps provide a useful numerical complement to a BMMP to aid its interpretation. Discussion of a discrepancy measure based on the average squared distance between the model smooths and the data smooth is given in Pardoe (2001a).

BMMPs utilize nonparametric scatterplot smoothers, and cubic smoothing splines and loess smoothers each perform well. However, care must be taken to select the smoothing parameter so that a truthful representation of the patterns in the BMMP is obtained. Further discussion of this issue, including simulation work, is given in Pardoe (2001a). Other smoother methods, including kernel smooths and Friedman's "super smoother," perform less well. The actual number of samples used to create the BMMP does not appear to have a large impact on this methodology, so there would appear to be no need to use any more than 100 samples for each plot; nevertheless, using any less than 100 would likely lead to poor resolution in the plots and make interpretation difficult.

The example considered here adopted Rubin's approach using posterior sampling. BMMPs based on Box's approach using prior sampling can be constructed similarly, although their interpretation is a little different. An intermediate approach using cross-validation/jackknifing ideas might also be useful, although implementation becomes trickier computationally.

Details for other types of regression model, such as linear and additive models, follow from the discussion for the binary logistic model. Other models, e.g. survival models, time series models, and random effects models, could no doubt benefit from the application of the ideas in this paper. One strength of the BMMP methodology is that it would appear to be broadly applicable to *any* regression situation, with just the details of obtaining samples and constructing the actual plots to worry about.

In addition, there are other plots used in the area of regression diagnostics that can be difficult to assess relative to the variation in the data. Examples include residual plots; CERES plots, which are a generalization of partial residual plots; and net-effect plots, which aid in assessing the contribution of a selected predictor to a regression. The ideas discussed above would appear to have a rôle to play in the analysis of such plots.

S-PLUS and R functions have been developed that can be used in conjunction with BUGS and BOA to construct BMMPs for the mean in any user specified direction h . The software is available at:

<http://lcb1.uoregon.edu/ipardoe/research/bmmpsoft.htm>

and further details are provided in Pardoe (2001c).

References

Bennett, K. P. and O. L. Mangasarian (1992). Robust linear programming discrimination of two linearly inseparable sets. In *Optimization Methods and Software*, Volume 1, pp. 23–34. Gordon and Breach Science Publishers.

- Bowman, A. W. and S. Young (1996). Graphical comparison of nonparametric curves. *Applied Statistics* 45, 83–98.
- Box, G. E. P. (1980). Sampling and Bayes' inference in scientific modelling and robustness (with discussion). *Journal of the Royal Statistical Society, Series A (General)* 143, 383–430.
- Cook, R. D. and I. Pardoe (2000). Comment on "Bayesian backfitting" by T. J. Hastie and R. J. Tibshirani. *Statistical Science* 15, 213–216.
- Cook, R. D. and S. Weisberg (1997). Graphics for assessing the adequacy of regression models. *Journal of the American Statistical Association* 92, 490–499.
- Gelman, A., X.-L. Meng, and H. Stern (1996). Posterior predictive assessment of model fitness via realized discrepancies (with discussion). *Statistica Sinica* 6, 733–807.
- Guttman, I. (1967). The use of the concept of a future observation in goodness-of-fit problems. *Journal of the Royal Statistical Society, Series B (Methodological)* 29, 83–100.
- Landwehr, J. M., D. Pregibon, and A. C. Shoemaker (1984). Graphical methods for assessing logistic regression models (with discussion). *Journal of the American Statistical Association* 79, 61–83.
- Pardoe, I. (2001a). *A Bayesian Approach to Regression Diagnostics*. Ph. D. thesis, School of Statistics, University of Minnesota.
- Pardoe, I. (2001b). A Bayesian sampling approach to regression model checking. *Journal of Computational and Graphical Statistics*. In press.
- Pardoe, I. (2001c). User's manual for bmmp S-PLUS and R software. Technical Report 639, School of Statistics, University of Minnesota.
- Pregibon, D. (1981). Logistic regression diagnostics. *The Annals of Statistics* 9, 705–724.
- Rubin, D. B. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *The Annals of Statistics* 12, 1151–1172.
- Smith, B. J. (2000). Bayesian Output Analysis Program (BOA). Version 0.5.0 for S-PLUS and R, available at <http://www.public-health.uiowa.edu/BOA>.
- Spiegelhalter, D. J., A. Thomas, N. G. Best, and W. Gilks (2000). Bayesian inference Using Gibbs Sampling (BUGS). Version 13 for Windows, available at <http://www.mrc-bsu.cam.ac.uk/bugs>.