Aust. N. Z. J. Stat. **49**(3), 2007, 241–250 doi: 10.1111/j.1467-842X.2007.00478.x

A GRAPHICAL DIAGNOSTIC FOR VARIANCE FUNCTIONS

IAIN PARDOE[1]* AND R. DENNIS COOK[2]

University of Oregon and University of Minnesota

Summary

This paper proposes diagnostic plots for regression variance functions. It shows how to extend graphical methodology that uses Bayesian sampling for checking the regression mean function to also check the variance function. Plots can be constructed quickly and easily for any model of interest. These plots help to identify model weaknesses and can suggest ways to make improvements. The proposed methodology is illustrated with two examples: a simple linear regression model to fix ideas, and a more complex study involving count data to demonstrate the potential for wide application.

Key words: Bayesian methodology; diagnostic plot; marginal model plot; model criticism; posterior predictive distribution.

1. Introduction

Pardoe (2001) and Pardoe & Cook (2002) describe 'Bayes marginal model plots' for checking regression model mean functions. Such plots enhance the 'marginal model plots' of Cook & Weisberg (1997) by displaying model uncertainty. This paper shows how to extend these ideas to situations where competing models differ mainly with respect to the regression variance function. While the proposed graphical diagnostic maintains the intuitive motivation of the papers just cited, the details of exactly how to extend the methodology are non-trivial; the purpose of this paper is to elucidate these details for general regression settings. The resulting diagnostic plots clearly display features of the fitted model that derive from the assumed variance function, and enable a visual assessment of whether the data support this variance function. The plots include a Bayesian band of uncertainty for the variance function derived from the fitted model, which seems to be a unique feature among graphical diagnostics for variance functions. Despite some work in this area (for example, Diblasi & Bowman, 1997), there do not appear to be any competing graphical methods available with all of the strengths of the proposed methodology. Improving specification of the variance function can be important for correctly assessing prediction uncertainty both for standard linear models and more complex regression models. The remainder of this section provides sufficient details on the use of Bayes marginal model plots for assessing mean functions to allow extensions to variance functions in Section 2. Section 3 contains an example based on a classic chemical data set which illustrates the use of the plots in a simple linear regression setting. A more complex application is presented in Section 4, which follows an analysis

*Author to whom correspondence should be addressed.

[1]Charles H. Lundquist College of Business, University of Oregon, Eugene, OR 97403, USA.
e-mail: ipardoe@lcbmail.uoregon.edu

[2]School of Statistics, University of Minnesota, Minneapolis, MN 55455, USA.

Acknowledgements. The authors thank Dr. Catherine A. Durham, the Associate Editor, and a reviewer for helpful comments and suggestions. R. D. Cook's research was supported in part by the U.S. National Science Foundation.

by Durham, Pardoe & Vega (2004) of restaurant wine demand, while Section 5 contains a discussion.

Suppose the exploratory stage of a regression analysis of a univariate response y on a vector of predictors $\boldsymbol{x}$ has been completed, and there is a reasonable first model to assess, $\mathrm{M}_{\boldsymbol{\theta}}(y \mid \boldsymbol{x})$, where $\boldsymbol{\theta}$ is a vector of unknown parameters. Assume that $\boldsymbol{\theta}$ can be consistently estimated under $\mathrm{M}_{\boldsymbol{\theta}}(y \mid \boldsymbol{x})$ with $\widehat{\boldsymbol{\theta}}$, or, for a Bayesian analysis, that inference will be based on a posterior distribution for the model, $\mathrm{M}_{\boldsymbol{\theta}}(y \mid \boldsymbol{x}, \boldsymbol{y}_d)$, where $\boldsymbol{y}_d$ is the n-vector of observed responses. The model should ideally provide a sufficiently accurate approximation to the unknown conditional cumulative distribution function $\mathrm{F}(y \mid \boldsymbol{x})$, relative to whatever practical issue is at hand. The nature of any model deficiency may indicate the need for model reformulation, or for separate consideration of poorly fitting or influential observations. Cook & Weisberg (1997) consider this problem from the perspective of the following:

$$\begin{aligned} &\mathrm{E}_{\mathrm{F}}(y \mid \boldsymbol{x}) = \mathrm{E}_{\widehat{\mathrm{M}}}(y \mid \boldsymbol{x}), \quad \forall\, \boldsymbol{x} \in \mathcal{X} \subset \mathbb{R}^p \\ \Longleftrightarrow\ &\mathrm{E}_{\mathrm{F}}(y \mid h) = \mathrm{E}_{\widehat{\mathrm{M}}}(y \mid h), \quad \forall\, h = h(\boldsymbol{x}) : \mathbb{R}^p \to \mathbb{R}^1, \end{aligned} \tag{1}$$

where E_{F} is model-free expectation under F, $\mathrm{E}_{\widehat{\mathrm{M}}}$ is model-based expectation under $\mathrm{M}_{\widehat{\boldsymbol{\theta}}}$, $\mathcal{X}$ is the sample space of $\boldsymbol{x}$, and h is any measurable function of $\boldsymbol{x}$.

If the dimension of $\boldsymbol{x}$ is greater than two, it is difficult to visualize $\mathrm{E}(y \mid \boldsymbol{x})$, and hence hard to check the first equality in (1). However, if h is univariate, then $\mathrm{E}(y \mid h)$ can be visualized in a 2-D scatterplot and the second equality in (1) can be checked using smoothed mean function estimates. In particular, smooth y versus h to obtain $\widehat{\mathrm{E}}_{\mathrm{F}}(y \mid h)$, and, since $\mathrm{E}_{\widehat{\mathrm{M}}}(y \mid h) = \mathrm{E}(\mathrm{E}_{\widehat{\mathrm{M}}}(y \mid \boldsymbol{x}) \mid h)$, smooth the model fitted values $\mathrm{E}_{\widehat{\mathrm{M}}}(y \mid \boldsymbol{x})$ versus h to obtain $\widehat{\mathrm{E}}_{\widehat{\mathrm{M}}}(y \mid h)$. A marginal model plot for the mean in the (marginal) direction h is a plot of y versus h with $\widehat{\mathrm{E}}_{\mathrm{F}}(y \mid h)$ and $\widehat{\mathrm{E}}_{\widehat{\mathrm{M}}}(y \mid h)$ superimposed. If $\mathrm{M}_{\widehat{\boldsymbol{\theta}}}$ is an accurate approximation to F, then $\widehat{\mathrm{E}}_{\mathrm{F}}(y \mid h) \approx \widehat{\mathrm{E}}_{\widehat{\mathrm{M}}}(y \mid h)$ and smooths that match closely for any function h provide support for $\mathrm{M}_{\widehat{\boldsymbol{\theta}}}$. Cook & Weisberg (1997) suggest possibilities for the checking functions h, including fitted values, individual predictors in the model, potential predictors not in the model, linear combinations of predictors, and random linear projections of $\boldsymbol{x}$.

Even if $\mathrm{M}_{\boldsymbol{\theta}}(y|\boldsymbol{x}) = \mathrm{F}(y|\boldsymbol{x})$ for some unknown value of $\boldsymbol{\theta}$, it should not be expected that $\mathrm{M}_{\widehat{\boldsymbol{\theta}}}(y \mid x) = \mathrm{F}(y \mid \boldsymbol{x})$ (because of sampling variability). From a Bayesian perspective, this variability can be represented using the posterior distribution for $\boldsymbol{\theta}$. By displaying this variability via the model smooth, it becomes easier to assess whether it seems reasonable for the data to have been generated by the model in question. In particular, Gelman, Meng & Stern (1996) suggest drawing a value of $\boldsymbol{\theta}$ from its posterior distribution, and then generating a $\boldsymbol{y}$-sample from the model M indexed by this $\boldsymbol{\theta}$. Then repeat this process a large number m of times and compare the data $\boldsymbol{y}_d$ to the m realizations from M. A graphical way to do this is to compare model-free predicted values with expected y-values based on sampled $\boldsymbol{\theta}$ values. A Bayes marginal model plot for the mean is then a scatterplot of y versus h with a mean function estimate under F superimposed, as well as mean function estimates for each model sample $\mathrm{M}_{\boldsymbol{\theta}_t}$, $t = 1, \ldots, m$. If enough samples are taken, say $m = 100$, the Bayes mean function estimates, $\widehat{\mathrm{E}}_{\mathrm{M}_{\boldsymbol{\theta}_t}}(y \mid h)$, $t = 1, \ldots, m$, form a mean function band under M. If, regardless of h, 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 $\mathrm{M}_{\widehat{\boldsymbol{\theta}}}$ provides an accurate description of the conditional distribution of $y \mid \boldsymbol{x}$ and is a useful model; otherwise, $\mathrm{M}_{\widehat{\boldsymbol{\theta}}}$ is called into question.

2. Bayes marginal model plots for the variance

The ideas discussed in Section 1 can be extended to compare variance functions. In particular, (1) can be recast in terms of variance functions rather than mean functions. To compare $\mathrm{Var}_{\mathrm{F}}(y|h)$ with $\mathrm{Var}_{\widehat{\mathrm{M}}}(y \mid h)$, first obtain $\widehat{\mathrm{Var}}_{\mathrm{F}}(y \mid h)$ as follows. First calculate $\widehat{\mathrm{E}}_{\mathrm{F}}(y \mid h)$ by smoothing y versus h. Then calculate squared 'model-free residuals', $\hat{e}_{\mathrm{F}}^2 = (y - \widehat{\mathrm{E}}_{\mathrm{F}}(y \mid h))^2$. Finally, smooth $\hat{e}_{\mathrm{F}}^2$ versus h to get the estimate $\widehat{\mathrm{Var}}_{\mathrm{F}}(y \mid h)$. The corresponding model-based estimate of the variance function $\mathrm{Var}_{\widehat{\mathrm{M}}}(y \mid h)$ uses the relationship

$$\mathrm{Var}_{\widehat{\mathrm{M}}}(y \mid h) = \mathrm{E}(\mathrm{Var}_{\widehat{\mathrm{M}}}(y \mid \boldsymbol{x}) \mid h) + \mathrm{Var}(\mathrm{E}_{\widehat{\mathrm{M}}}(y \mid \boldsymbol{x}) \mid h). \tag{2}$$

To estimate the first term on the right-hand side of (2), smooth the (assumed) variance function from the fitted model, $\mathrm{Var}_{\widehat{\mathrm{M}}}(y \mid \boldsymbol{x})$, versus h to obtain $\widehat{\mathrm{E}}(\mathrm{Var}_{\widehat{\mathrm{M}}}(y \mid \boldsymbol{x}) \mid h)$. To estimate the second term on the right-hand side of (2), first calculate $\widehat{\mathrm{E}}_{\widehat{\mathrm{M}}}(y \mid h)$ by smoothing $\mathrm{E}_{\widehat{\mathrm{M}}}(y \mid \boldsymbol{x})$ versus h. Then calculate squared 'model-based discrepancies', $\hat{e}_{\widehat{\mathrm{M}}}^2 = (\mathrm{E}_{\widehat{\mathrm{M}}}(y \mid \boldsymbol{x}) - \widehat{\mathrm{E}}_{\widehat{\mathrm{M}}}(y \mid h))^2$. Then smooth $\hat{e}_{\widehat{\mathrm{M}}}^2$ versus h to obtain $\widehat{\mathrm{Var}}(\mathrm{E}_{\widehat{\mathrm{M}}}(y \mid \boldsymbol{x}) \mid h)$. Finally, $\widehat{\mathrm{Var}}_{\widehat{\mathrm{M}}}(y \mid h) = \widehat{\mathrm{E}}(\mathrm{Var}_{\widehat{\mathrm{M}}}(y \mid \boldsymbol{x}) \mid h) + \widehat{\mathrm{Var}}(\mathrm{E}_{\widehat{\mathrm{M}}}(y \mid \boldsymbol{x}) \mid h)$. Now, if $\mathrm{M}_{\widehat{\boldsymbol{\theta}}}$ is an accurate approximation to F, then for any h the marginal variance function estimates should agree, $\widehat{\mathrm{Var}}_{\mathrm{F}}(y \mid h) \approx \widehat{\mathrm{Var}}_{\widehat{\mathrm{M}}}(y \mid h)$. Any indication that the function estimates do not agree for one particular h calls $\mathrm{M}_{\widehat{\boldsymbol{\theta}}}$ into question; if they agree for a variety of plots, there is support for $\mathrm{M}_{\widehat{\boldsymbol{\theta}}}$.

If the same method and smoothing parameter are used for the two smooths of model-free and model-based quantities in a scatterplot, any estimation bias should approximately cancel (see Bowman & Young, 1996), allowing their point-wise comparison. Thus each of the smooths used to obtain $\widehat{\mathrm{Var}}_{\mathrm{F}}(y \mid h)$ and $\widehat{\mathrm{Var}}_{\widehat{\mathrm{M}}}(y \mid h)$ should have the same smoothing parameter, γ. Therefore it is desirable to select γ so that the smooths are flexible enough to capture clear systematic trends in all of the corresponding scatterplots, while not over-fitting too much in any one scatterplot, over-reacting to individual points, or tracking spurious patterns. This is clearly impractical, but a viable alternative is to graphically select γ to capture the systematic trends in both a scatterplot of the data (y) versus h and a scatterplot of the fitted values from the model versus h. Further discussion of this issue is in Section 5 (see also Pardoe & Cook, 2002).

Although a variance function estimate should be non-negative, it is possible for the smooths involved to move into negative territory. One simple way to deal with this is to truncate any negative estimates for non-negative quantities at zero. More sophisticated methods might be preferable in certain cases, although the truncation method is fast and appears to work well in practice. In addition, changing the smoothing parameter to make the smooths more flexible can often get around this issue. This method for smoothing the variance function works well if the mean function is estimated well. However, if the mean function smooth is very biased, the resultant variance function smooth can be very poor. In such cases, Ruppert *et al.* (1997) suggest a bias adjustment—this, however, can require extensive further calculation.

When trying to assess variance functions, a graphical display based on the standard deviation scale is often more meaningful than the less-intuitive variance scale. Thus, one way to compare the variance function estimates visually is to superimpose lines representing $\widehat{\mathrm{E}}_{\mathrm{F}}(y \mid h) \pm \widehat{\mathrm{Var}}_{\mathrm{F}}(y \mid h)^{1/2}$ and $\widehat{\mathrm{E}}_{\widehat{\mathrm{M}}}(y \mid h) \pm \widehat{\mathrm{Var}}_{\widehat{\mathrm{M}}}(y \mid h)^{1/2}$ on a marginal model plot for the mean. The square roots of the variance function estimates can be thought of as 'standard deviation function' estimates. An alternative display for the variance function check is to plot the absolute values of the model-free residuals, $|\hat{e}_F| = |y - \widehat{E}_F(y \mid h)|$, versus h, and then superimpose the square roots of the variance function estimates, $\widehat{\mathrm{Var}}_F(y \mid h)^{1/2}$ and $\widehat{\mathrm{Var}}_{\widehat{M}}(y \mid h)^{1/2}$. The latter approach tends to provide a clearer, more easily interpreted plot. There is further discussion of this issue in Section 5.

To incorporate model variability in the plots, calculate model-based variance function estimates corresponding to individual samples from the posterior distribution of $\boldsymbol{\theta}$. Such estimates involve following the procedure based on (2), with the $\boldsymbol{\theta}$ samples used to obtain $\widehat{E}(E_{M_{\boldsymbol{\theta}_t}}(y \mid \boldsymbol{x}) \mid h)$ and $\widehat{E}(\mathrm{Var}_{M_{\boldsymbol{\theta}_t}}(y \mid \boldsymbol{x}) \mid h)$, $t = 1, \ldots, m$. A *Bayes marginal model plot for the variance* is then a plot of the absolute values of the model-free residuals, $|\hat{e}_F| = |y - \widehat{E}_F(y|h)|$, versus h, with a standard deviation function estimate under F superimposed. Then, superimpose a standard deviation function estimate for each model sample $M_{\boldsymbol{\theta}_t}$, $t = 1, \ldots, m$. In particular, $\widehat{\mathrm{Var}}_F(y \mid h)^{1/2}$ is compared with $\widehat{\mathrm{Var}}_{M_{\boldsymbol{\theta}_t}}(y \mid h)^{1/2}$, $t = 1, \ldots, m$.

3. Chemical experiment application

Franklin *et al.* (1956) described a chemical experiment to develop a catalyst in the vapour phase oxidation of naphthalene. The study authors considered how three variables, $AN = \log_{10}$(air to naphthalene ratio), $Ctime = \log_{10}$(contact time), and $Btemp = 0.01$(bed temperature-330), affect Yn = percentage mole conversion of naphthalene to naphthoquinone. The variable transformations result in approximate trivariate normality and appear to stabilize the yield surface. Cook (1998) suggested that the dependence of Yn on the predictors is through the single linear combination $x = 0.397\,AN + 0.445\,Ctime + 0.802\,Btemp$. Box-Cox response transformation methodology (Box & Cox, 1964) indicates that a simple linear regression model with transformed response $\log(Yn)$ and single predictor x provides a good fit to the data. It is straightforward to obtain posterior samples for the parameters of this model directly using standard non-informative priors (Lindley & Smith, 1972), and there is little evidence in the resulting Bayes marginal model plot for the mean with $h = x$ to call this model into question (plot not shown).

However, the constant variance function of this model appears, however, to be misspecified, as is clearly evident in the Bayes marginal model plot for the variance with $h = x$ in Figure 1. Since there is just one predictor in this model, there is an easy way to alter the model to attempt to clear up this problem. As the data indicate increasing variance with x, a weighted linear model using weights $\propto x^{-1}$ or x^{-2} may improve matters. Extrapolating the black data smooth in Figure 1 to the left, it would appear to cross the horizontal axis at about $x = 1$. So, setting weights equal to $(x - 1)^{-2}$ seems a sensible place to start.

Figure 2 shows a Bayes marginal model plot for the variance with $h = x$ for the weighted model

$$y \mid x = E(y \mid x) + e/\sqrt{w},$$

where $E(y|x) = \theta_0 + \theta_1 x$, the errors e are normally distributed with mean 0 and variance σ^2, and the weights w are $(x - 1)^{-2}$. Now, the black data-based smooth follows the pattern of the gray model-based smooths much more closely, and this weighted model appears to improve on the unweighted model. Using appropriate weights is important here so that the level of uncertainty (whether quantified using frequentist or Bayesian methods) changes according to the variability in the data. The mean function estimate changes little whether weights are used

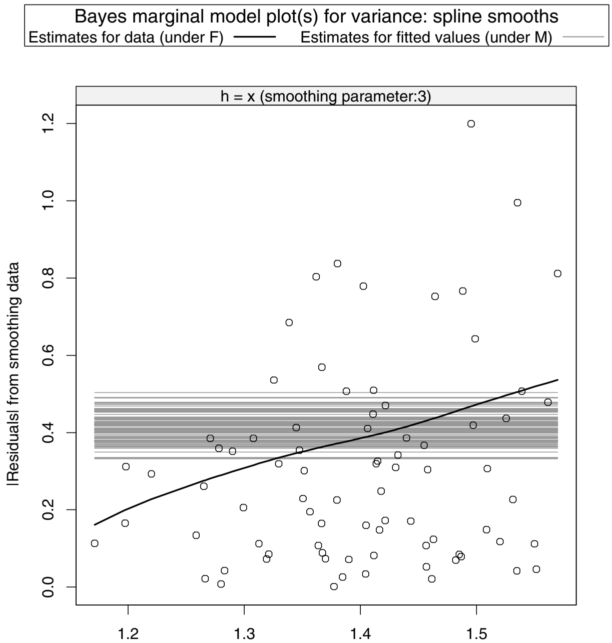

Figure 1. Bayes marginal model plot for the variance with $h = x$ for the simple linear regression model fitted to the naphthalene data.

or not (the Bayes marginal model plot for the mean for $h = x$ looks similar under weighted and unweighted models), but prediction intervals, for example, may vary considerably depending on whether or not weights are used.

Compare Figures 1 and 2 with the standard residual plots in Figure 3. The heterogeneity problem is less obvious in the residual plot for the unweighted model (on the left of the figure) than it was in Figure 1, and there is certainly no clear indication that using weights equal to $(x - 1)^{-2}$ might improve the model. In fact, using such weights results in a residual plot (on the right of the figure) that is virtually indistinguishable from the original plot. The Bayes marginal model plots on the other hand clearly demonstrate a heterogeneity problem with the unweighted model that can be corrected with a suitable choice of weights.

4. Wine demand application

Durham *et al.* (2004) presented the results of a Bayesian analysis of wine demand at a restaurant, using hedonic quantity models to evaluate the impact of objective factors (e.g., origin, varietal), sensory descriptors, and price, on the choice of 47 red and 29 white wines. The data were collected at a high-end restaurant over nineteen weeks in 1998. The restaurant offers a wide selection of wines detailed in an extensive menu that describes brand, vintage, origin, price, and sensory qualities. Sensory information includes aroma, flavours, and 'mouth feel' (e.g. dry, tannic, smooth, big, creamy, heavy), with typical descriptors for aroma and taste including fruits (berry, lemon), flowers (apple, rose), and other food associations (herbal, honey, chocolate). Wine prices are generally based on expert quality assessments,

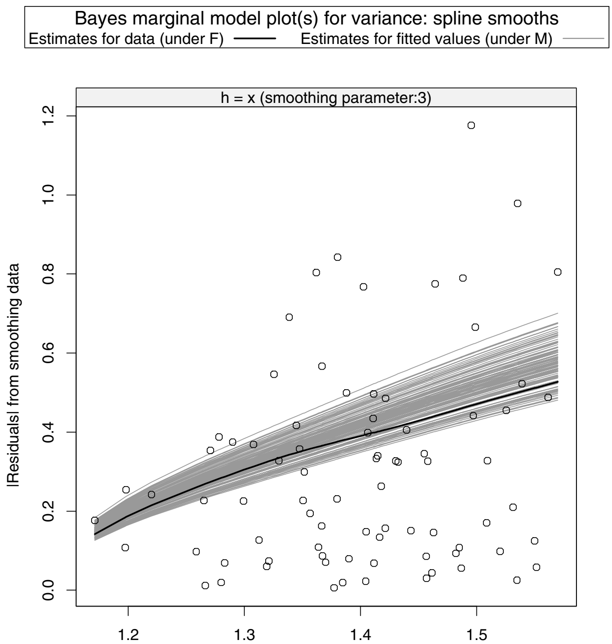

Figure 2. Bayes marginal model plot for the variance with $h = x$ for the weighted simple linear regression model fitted to the naphthalene data.

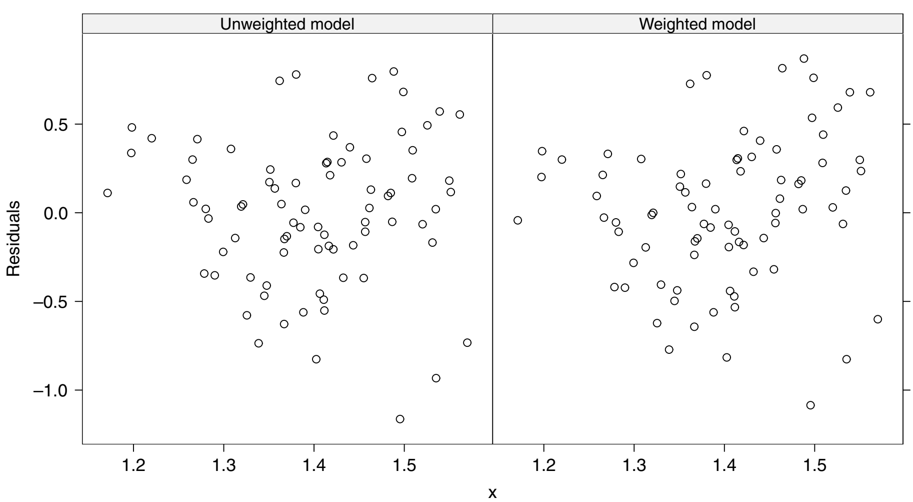

Figure 3. Residual plots for the unweighted model (left) and weighted model (right) fitted to the naphthalene data.

with adjustment for varietal, origin, and market factors, and only rarely can the price of a wine be said to reflect consumer valuation of its quality. Many wines can appear to be greatly over—or under-priced due to the great variety of wines available, supply variation, and a lack of good information on quality. This study was conducted to explore whether wine demand,

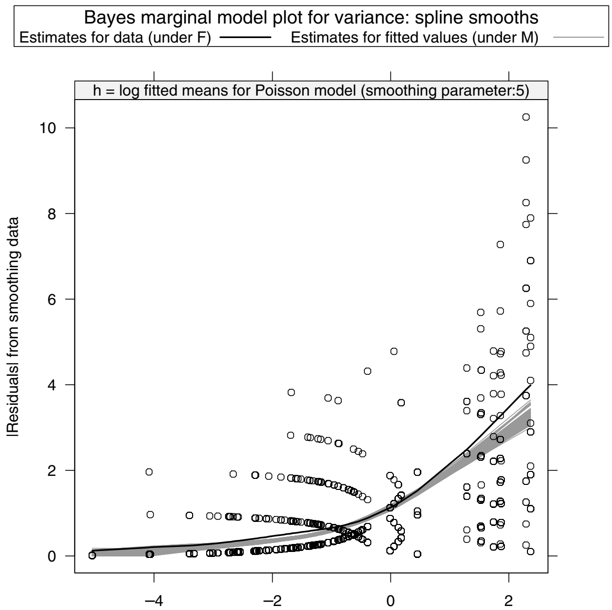

Figure 4. Bayes marginal model plot for the variance with h = log(fitted means) for the Poisson model fitted to all the wine data.

measured by quantity sold in each of the 19 weeks, was possibly driven by the objective and sensory descriptors on wine menus rather than simply price.

Count data such as these can be modelled using Poisson regression, with the (log) Poisson means dependent on the wine characteristics. However, these data exhibit over-dispersion, with more zero-counts than expected by a Poisson model. To address this, the study authors tried a number of alternative models; this paper presents Bayes marginal model plots for only the Poisson and zero-inflated Poisson (Lambert, 1992) models. Zero-inflated Poisson regression allows for over-dispersion by modelling the counts using a mixture distribution: Poisson with probability p_i or identically zero (i.e. a structural zero) with probability $1 - p_i$. The Poisson means and structural zero probabilities are modelled as functions of the wine characteristics. The study authors modified this set-up to account for wines available by the glass that almost always have positive sales. These wines were modelled as Poisson(μ_i) and other (non-glass) wines followed the zero-inflated model.

WinBUGS software (Spiegelhalter *et al.*, 2003) was used to generate posterior samples for the model parameters (using non-informative, zero-mean, Normal priors with standard deviations of ten, based on an assumption that it was implausible that model parameters would be more than about plus/minus 20). The residual deviance for the Poisson model was 1369 on 1381 degrees of freedom, ordinarily suggestive of a reasonable fit. Furthermore, Bayes marginal model plots for the mean show little difference in fit between the models considered. However, Bayes marginal model plots for the variance indicate a problem with the Poisson model. In particular, as the Poisson model restricts the variance to be the same as the mean, it fails to track the higher count variation at larger values of h = log(fitted means), as can be seen in Figure 4.

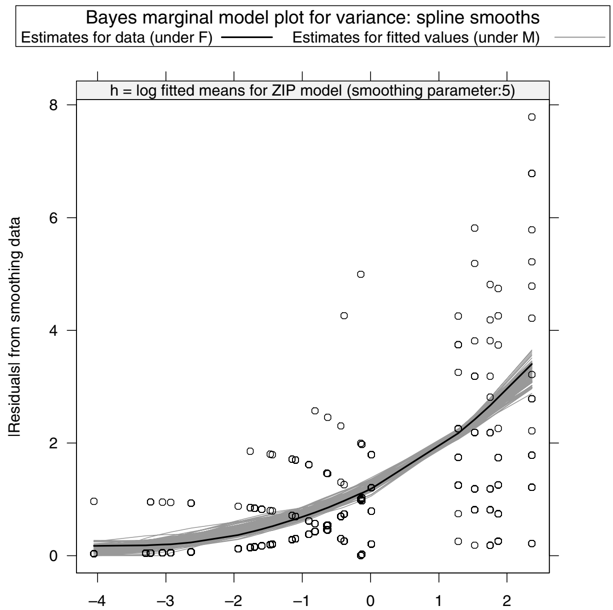

Figure 5. Bayes marginal model plot for the variance with h = log(fitted means) for the zero-inflated Poisson model fitted to white wines only.

The zero-inflated model fitted to all the data does track the variance more closely at larger values of h = log(fitted means), but the corresponding plot continues to indicate model discrepancies (plot not shown). However, the zero-inflated model does appear to fit the data satisfactorily for just the white wines. The Bayes marginal model plot methodology actually prompted this discovery as, to provide confidence in a particular model, there should be broadly matching smooths not only for a variety of h-functions but also for any subset of the data. Figure 5 shows a plot for the zero-inflated model fitted to the white wines only.

This plot shows a clear improvement on Figure 4, and provides support for this model (for white wines). A similar plot for red wines (not shown) is less convincing, although it too shows an improvement over the Poisson model. The study authors found that another model they considered (zero-inflated negative binomial) provided a more satisfactory fit for the red wines. Nevertheless, the improvement over the zero-inflated Poisson model was marginal and substantive results were essentially unchanged, so that the final results reported in Durham *et al.* (2004) use the zero-inflated Poisson model for all wines, red and white.

5. Discussion

Bayes marginal model plots for the variance offer an intuitive way to check regression variance functions graphically. After obtaining posterior samples for model parameters, a series of smoothing operations provides model-free and model-based variance estimates that can be compared in simple two-dimensional scatterplots. Cycling through plots in various directions h provides guidance on the fit of the model variance function. In an area where there appear to be no other competing graphical methods, this methodology is proposed as a viable aid to model assessment (as opposed to model selection). Publicly available software at `http://lcb1.uoregon.edu/ipardoe/research/bmmpsoft/bmmpsoft.htm` is available to construct these plots for standard models such as normal linear regression models. The software code can be adapted relatively easily for non-standard models such as the ones used for the wine example. The methods make use of common non-parametric methods for continuous data, such as cubic smoothing splines and loess smoothers. Experience suggests that use of more sophisticated smoothers usually changes the visual impression of a plot very little. Strictly speaking, local linear smoothers might be preferred, since it is not clear that other methods have the bias cancellation property mentioned in Section 2.

Whichever smoothing method is used, some care must be exercised in the selection of the smoothing parameter. Over-smoothing can result in smooths that miss clear patterns, while under-smoothing can produce highly variable smooths that track spurious patterns. Broadly speaking, experience using this methodology and previous simulation studies suggest that when the smoothing parameter provides a reasonable compromise between over- and under-smoothing, poorly-fitting and well-fitting models can be correctly identified most of the time. Although standard methods for smoothing parameter selection (for example, Hurvich, Simonoff & Tsai, 1998) sometimes work well, in general there appear to be no reliable short-cut methods for pre-selecting the smoothing parameter ahead of time, or recommended values that will work in most situations. Rather, the analyst should look at scatterplots of the data versus h and of the model fitted values versus h, and choose a smoothing parameter that produces reasonable smooths in both plots simultaneously. If in doubt, err on the side of over- rather than under-smoothing, since under-smoothing sometimes produces plots indicating model inadequacy in cases when a model is good, but over-smoothing rarely leads to plots indicating model adequacy for poor models. Furthermore, the amount of over- or under-smoothing has to be quite extreme to produce poor results. Although the suggested method for choosing the smoothing parameter relies on human perception of patterns in a scatterplot, experience indicates that, in the case of smoothing splines, variation of two or three effective degrees of freedom between different analysts is not unreasonable, but makes little (qualitative) difference to the plots. Varying the smoothing parameter any more than this can have adverse effects, but is unlikely to be a problem in practice since it would probably correspond to a very poor representation of the patterns in the scatterplot.

This paper suggests that the variance function check should be performed graphically on the standard deviation scale. However, smoothing squared residuals and then taking the square root is not quite the same as smoothing the corresponding absolute residuals, as by Jensen's inequality, $\mathrm{E}(|\text{residual}|) = \mathrm{E}\ ((\text{residual}^2)^{1/2}) \leq\ (\mathrm{E}\ (\text{residual}^2))^{1/2}$. Ideally, as it is standard deviation function estimates rather than variance function estimates that are being displayed in the suggested plots, all smoothing should be done on the standard deviation scale. For the model-free estimate, this would involve smoothing the absolute values of the model-free residuals. As noted above, this is not the same as smoothing the squared model-free residuals and then taking the square root, although they will probably be close. It is not clear, however, what the corresponding model-based estimate should be on the standard deviation scale—there is no relationship equivalent to (2) on this scale. Thus, the methodology described above of smoothing variances and taking square roots is the only possibility if the variance function check is to be done on the standard deviation scale. It is possible to instead use the variance scale, thus avoiding the square root difficulties, but this advantage has to be weighed against the disadvantage of working on an unintuitive scale.

Bayes marginal model plots can be summarized numerically using a discrepancy measure. First calculate the average squared distance from the model-free smooth to the model-based smooths (a). Then calculate the average squared distances from each model-based smooth to all the other model-based smooths ($b_m, m = 1, \ldots, 100$). The discrepancy measure is the proportion of b_ms larger than a; if near one then the model-free smooth passes close to the center of the model-based smooths, while if near zero then the model-free smooth differs greatly from the model-based smooths (e.g., the discrepancy measure for Figure 4 is 0.0, while that for Figure 5 is 0.51). See Hart (1997) for similar ideas in this vein.

As suggested by a referee, it would be interesting to see whether or not these graphical methods work as well with prediction sum of squares (PRESS) residuals, which are sometimes used in model comparison. See, for example, Tarpey (2000) for a recent paper on these in a constrained setting.

References

BOWMAN, A. W. & YOUNG, S. (1996). Graphical comparison of nonparametric curves. *Appl. Statist.* **45**, 83–98.

BOX, G. E. P. & COX, D. R. (1964). An analysis of transformations (with discussion). *J. Roy. Statist. Soc. Ser. B* **26**, 211–246.

COOK, R. D. (1998). *Regression Graphics: Ideas for Studying Regressions through Graphics*. New York: Wiley.

COOK, R. D. & WEISBERG, S. (1997). Graphics for assessing the adequacy of regression models. *J. Amer. Statist. Assoc.* **92**, 490–499.

DIBLASI, A. & BOWMAN, A. E. (1997). Testing for constant variance in a linear model. *Statist. Probab. Lett.* **33**, 95–103.

DURHAM, C. A., PARDOE, I. & VEGA, E. (2004). A methodology for evaluating how product characteristics impact choice in retail settings with many zero observations: An application to restaurant wine purchase. *J. Agric. Resource Econ.* **29**, 112–131.

FRANKLIN, N. L., PINCHBECK, P. & POPPER, F. (1956). A statistical approach to catalyst development, part 1: The effect of process variables in the vapor phase oxidation of naphthalene. *Trans. Instit. Chem. Eng.* **34**, 280–293.

GELMAN, A., MENG, X.-L. & STERN, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies (with discussion). *Statist. Sinica* **6**, 733–807.

HART, J. D. (1997). *Nonparametric Smoothing and Lack-of-fit Tests*. New York: Springer-Verlag.

HURVICH, C. M., SIMONOFF, J. S. & TSAI, C.-L. (1998). Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *J. Roy. Statist. Soc. Ser. B* **60**, 271–293.

LAMBERT, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* **34**, 1–14.

LINDLEY, D. V. & SMITH, A. F. M. (1972). Bayes estimates for the linear model (with discussion). *J. Roy. Statist. Soc. Ser. B* **34**, 1–41.

PARDOE, I. (2001). A Bayesian sampling approach to regression model checking. *J. Comput. Graph. Statist.* **10**, 617–627.

PARDOE, I. & COOK, R. D. (2002). A graphical method for assessing the fit of a logistic regression model. *Amer. Statist.* **56**, 263–272.

RUPPERT, D., WAND, M. P., HOLST, U. & HÖSSJER, O. (1997). Local polynomial variance-function estimation. *Technometrics* **39,** 262–273.

SPIEGELHALTER, D. J., THOMAS, A., BEST, N. G. & LUNN, D. (2003). *WinBUGS Version 1.4 User Manual*. Cambridge, UK: MRC Biostatistics Unit.

TARPEY, T. (2000). A note on the prediction sum of squares statistic for restricted least squares. *Amer. Statist.* **54**, 116–118.