Model Assessment Plots for Multilevel Logistic Regression

Iain Pardoe

Department of Decision Sciences, Charles H. Lundquist College of Business, University of Oregon, Eugene, OR 97403–1208, USA. Tel: (541) 346-3250. Fax: (541) 346-3341.

Abstract

This paper extends the Bayes marginal model plot (BMMP) model assessment technique from a traditional logistic regression setting to a multilevel application in the area of criminal justice. Convicted felons in the United States receive either a prison sentence or a less severe jail or non-custodial sentence. Researchers have identified many determinants of sentencing variation across the country, some individual such as type of crime and race, and some based on geographical units such as county crime rate. Multilevel rather than conventional regression should be used to quantify any interplay between such individual-level and county-level effects since the covariates have a hierarchical structure. Questions arise, however, as to whether a multilevel model provides an adequate fit to the data, and whether the computational burden of a multilevel model over a conventional model is justified. Residual plots, traditionally used to assess regression models, are difficult to interpret with a binary response variable and multilevel covariates, as in this case. BMMPs, an alternative graphical technique, can be used to visualize goodness of fit in such settings. The plots clearly demonstrate the need to use multilevel modeling when analyzing data such as these.

Key words: Bayesian methodology, Criminal justice, Diagnostic, Graphical method, Hierarchical model, Random effect

---

Email address: `ipardoe@lcbmail.uoregon.edu` (Iain Pardoe).
URL: `http://lcb1.uoregon.edu/ipardoe` (Iain Pardoe).

1 Introduction

In 2001, the United States imprisoned its citizens at a rate of 470 per 100,000, six to twelve times higher than in other western countries. Furthermore, there is large variation in imprisonment levels within the U.S. For example, in 2001, Louisiana's rate per 100,000 residents was 800, while Maine's was 127 [6, p. 4]. Studies of differences in prison use among the states have found various factors to play a key role, including: higher levels of crime [8], in particular violent crime [5]; percent of the population that is African American [8]; political conservatism [19,5]; and geographic region—Southern states appear to punish more severely [11]. There is also some empirical evidence of a relationship between state sentencing schemes and levels of incarceration, since such schemes often dictate which types of offense warrant prison time [22]. Other studies examining aggregate punishment variation using a county as the unit of analysis have found unemployment in urban counties and violent crime [7], and percent of the population that is African American and Southern region [20] to be significantly related to prison use.

By contrast, most sentencing studies focus on individuals, whereby effects of case characteristics, criminal history, and demographics are determined. However, effects of individual-level variables may vary according to the cultural, political, economic, and social contexts in which courts operate [3]. Studies of pooled statewide data have found several contextual variables to have an effect on sentencing, such as level of unemployment and crime rate [12] and racial composition [19]. However, these studies use conventional logistic regression which does not correctly account for individual-level effects that vary according to a jurisdiction's cultural context and organizational constraints [1,9]. To properly account for the hierarchical nature of individual-level covariates and county-level contextual covariates, multilevel modeling is more appropriate.

There has been some previous use of multilevel modeling in criminal justice research. For example, Ref. [17] uses a multilevel model for intra-city neighborhood differences in victimization risk, while Ref. [23] compares multilevel and conventional models for the impact of prison and inmate characteristics on misconduct. Ref. [1] investigates whether social context and racial disparities affected punishment decisions in Pennsylvania counties for 1991-1994. Controlling for urbanization, racial threat, economic threat, and crime control, punishment severity varies by race across jurisdictions, but measures of social context explains little of this variation.

Ref. [14] analyzes data from the Bureau of Justice Statistics' State Court Processing Statistics (SCPS) program, a biennial collection of data on felony defendants in state courts in 39 of the 75 most populous U.S. counties. That study uses the multilevel logistic regression model described in Section 3. Given the lack of consensus regarding determinants of variation in prison use, it is important to assess the fit of this model before it is used to inform policy. Furthermore, from a practical viewpoint, it is useful to gauge the relative worth of going beyond a conventional (non-multilevel) model with this more computationally intensive multilevel model. Theoretically, the "independent errors" assumption for a conventional model is violated with data having this hierarchical structure, but such a violation might be considered to be of little consequence if the model fits essentially as well as the multilevel model, and if results and conclusions are similar.

Ref. [13] describes a graphical technique for assessing the fit of a logistic regression model, called a "Bayes marginal model plot" (BMMP). The remaining sections of this paper describe an extension of the BMMP technique to assess the fit of the multilevel logistic regression model used in Ref. [14]. Section 2 describes the U.S. imprisonment data used, while the multilevel model employed is outlined in Section 3. Section 4 concerns assessment of the model using BMMPs, and compares the fit of a similar conventional model. The two models result in very different conclusions for this dataset, and the model assessment plots clearly show that the multilevel model fits the data well, whereas the conventional model fits poorly. Section 5 contains a discussion.

2 Data

Information collected in the SCPS program includes demographic characteristics, criminal history, and details of pretrial processing, disposition, and sentencing of felony defendants. Ref. [14] analyzes individual-level data for 8,446 felony convictions in 39 counties across 17 states during May 1998 linked to county-level variables using the Federal Information Processing Standards code. The number of individuals in each county ranged from 23 to 905, with median 186. Fourteen of the states represented contained just one or two sampled counties, while one contained four counties, one contained seven, and one contained eight.

In general, a jail sentence is less severe than a prison sentence, and so the factors influencing a decision to sentence to jail are likely to be markedly different from those affecting a decision to sentence to prison. Thus, in common with many studies that address cross-jurisdictional differences in punitiveness, the response variable for this study was "sentence severity", defined as $Y = 1$ if the offender received a prison sentence or 0 for a jail or non-custodial sentence, rather than "incarceration." Table 1 provides details of the twelve individual-level and six county-level covariates conjectured to affect sentencing severity.

[TABLE 1 ABOUT HERE]

3 Multilevel Logistic Regression

A multilevel logistic regression model, also referred to in the literature as a hierarchical model, can account for lack of independence across levels of nested data (i.e., individuals nested within counties). Conventional regression assumes that all experimental units (in this case, individuals) are independent in the sense that any variables affecting sentencing severity have the same effect in all counties. Multilevel modeling relaxes this assumption and allows these variables' effects to vary across counties. One way to do this uses a generalization of the model developed in Ref. [21]. First, each group of n_j individuals within $J = 39$ counties is assumed to follow a county-specific logistic regression model. For the ith individual in the jth county, observe a binary response,

$$\mathrm{Y}_{ij} = \begin{cases} 1 & \text{for a prison sentence} \\ 0 & \text{for a jail or non-custodial sentence} \end{cases}$$

$\mathrm{Y}_{ij}|p_{ij} \sim$ Bernouilli(p_{ij}), where $p_{ij} = \mathrm{Pr}(\mathrm{Y}_{ij} = 1)$, and

$$\mathrm{logit}(p_{ij}) = \log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = \boldsymbol{X}_i^T \boldsymbol{\beta}_j \quad (1)$$

where $\boldsymbol{X}_i$ represents measurements on K individual-level variables and $\boldsymbol{\beta}_j$ consists of K regression coefficients (specific to the jth county). Next, since each β-coefficient is likely to be related across counties, assume that each one can be explained by up to L county-level variables,

$$\boldsymbol{\beta}_j = \boldsymbol{G}_j \boldsymbol{\eta} + \boldsymbol{\alpha}_j \quad (2)$$

where $\boldsymbol{G}_j$ is a $K \times M$ block-diagonal matrix of measurements on L county-level variables, $\boldsymbol{\eta}$ consists of M regression coefficients, and $\boldsymbol{\alpha}_j$ is a $K \times 1$ vector of county-level errors. In particular, the kth row of $\boldsymbol{G}_j$ contains a non-zero block with a one for an intercept together with the county-level variables used to explain the kth β-coefficient. Thus, M is $K \times L$ if all county-level variables are used to explain each β-coefficient, or less than this otherwise. Combining (1) and (2) leads to

$$\mathrm{logit}(p_{ij}) = \boldsymbol{X}_i^T \boldsymbol{G}_j \boldsymbol{\eta} + \boldsymbol{X}_i^T \boldsymbol{\alpha}_j \quad (3)$$

Conventionally, the η-parameters in (3) are fixed effects (they have no j-subscript and represent the same effect over all counties) while the α-parameters are random effects (they have a j-subscript and represent different effects across counties). The presence of both types of effects makes (3) a mixed model. Suppressing the county-level errors so that (3) becomes a fixed effects model and amenable to standard regression requires assuming that individual-level effects are the same across counties, an assumption unlikely to be satisfied in practice.

Mixed models can be fit using specialized software such as “MLwiN” [15] and “HLM” [16]. Alternatively, by putting the model into a Bayesian framework, the distinction between fixed and random effects disappears (since all effects are now considered random), and the hierarchical structure is explicitly accounted for in the analysis. Ref. [14] follows this Bayesian route, giving $\boldsymbol{\eta}$ independent, zero-mean, normal priors with variances of 10 for the interactions and 100 for the main effects. An exchangeable prior was used for the county-level errors, $\boldsymbol{\alpha}_j \sim N(\mathbf{0}, \boldsymbol{\Gamma}^{-1})$, where $\mathbf{0}$ is a K-vector of zeros and $\boldsymbol{\Gamma}^{-1}$ is a $K \times K$ covariance matrix. A hyper-prior distribution was specified for the inverse covariance matrix, $\boldsymbol{\Gamma} \sim \text{Wishart}(\boldsymbol{R}, K)$, where $\boldsymbol{R}$ can be considered a prior estimate of $\boldsymbol{\Gamma}^{-1}$ based on K observations, and, to represent vague prior knowledge, degrees of freedom for the Wishart distribution was set as small as possible to be K (the rank of $\boldsymbol{\Gamma}$). $\boldsymbol{R}$ was set to have values ten along the diagonal and zero elsewhere (sensitivity analysis, discussed in Ref. [14], confirmed that the choice of $\boldsymbol{R}$ has little effect on the results).

The software package WinBUGS [18] was used to generate posterior samples for $\boldsymbol{\eta}$ and $\boldsymbol{\alpha}_j$; this free software enables Bayesian analysis of complex statistical models using Gibbs sampling, a Markov chain Monte Carlo (MCMC) technique. Since there are 3,876 cases with some missing data (which, based on the patterns of missingness, it seems reasonable to assume is missing at random), additional Gibbs steps were used to impute missing values. Further details are provided in Ref. [14]. After running four chains for 20,000 iterations, discarding 10,000 burn-in samples, and thinning to retain every tenth sample to reduce autocorrelation (leaving a total of 4,000 posterior samples), trace plots showed a good degree of mixing and MCMC convergence diagnostics indicated convergence.

Before interpreting and using posterior samples from this model, model fit and underlying assumptions need to be assessed. Posterior samples of county-level errors, $\boldsymbol{\alpha}_j$, are a form of residual, and so conceivably could lend themselves to the usual kinds of model diagnostics. For this application, the fact that they averaged close to zero across counties is reassuring, but unsurprising. More open to doubt are the normality and exchangeability assumptions, which suggests that normal probability plots or plotting posterior means of the $\boldsymbol{\alpha}_j$ against county-level covariates could be useful. However, since multilevel modeling shrinks the $\boldsymbol{\alpha}_j$ estimates from individual (within-county) estimates to the population average, it is not clear that such plots can tell us anything useful about these assumptions. Traditional residual plots might also provide some insight into model fit, but it is unclear how to even define a residual plot here given the hierarchical nature of the model. Furthermore, Ref. [13] discusses how the use of residual plots can be problematic in logistic regression settings.

Thus, each of these diagnostic methods seems insufficient to assess the fit of a multilevel model of such complexity. Section 4 describes use of an alternative graphical diagnostic procedure that avoids these difficulties.

4 Bayes Marginal Model Plots

Ref. [2] proposes the use of "marginal model plots" (MMPs) to assess the goodness of fit of a regression model. Extending their rationale to multilevel regression with covariates $\boldsymbol{X}$ measured on units nested in clusters with covariates $\boldsymbol{G}$ leads to:

$$\mathrm{E}_{\mathrm{F}}(Y|\boldsymbol{X},\boldsymbol{G}) = \mathrm{E}_{\widehat{\mathrm{M}}}(Y|\boldsymbol{X},\boldsymbol{G}), \qquad \forall\, \boldsymbol{X} \in \mathcal{X} \subset \mathrm{I\!R}^K, \forall\, \boldsymbol{G} \in \mathcal{G} \subset \mathrm{I\!R}^L \tag{4}$$

$$\iff \mathrm{E}_{\mathrm{F}}(Y|h) = \mathrm{E}_{\widehat{\mathrm{M}}}(Y|h), \qquad \forall\, h = h(\boldsymbol{X},\boldsymbol{G}) :\ \mathrm{I\!R}^{K+L} \rightarrow \mathrm{I\!R}^1 \tag{5}$$

where E_{F} denotes *model-free* expectation, $\mathrm{E}_{\widehat{\mathrm{M}}}$ denotes *model-based* expectation, $\mathcal{X}$ and $\mathcal{G}$ are the sample spaces of $\boldsymbol{X}$ and $\boldsymbol{G}$ respectively, and h is any measurable function of $\boldsymbol{X}$ and $\boldsymbol{G}$. Ideas for selecting useful h-functions to consider in practice are given in Ref. [2], and include fitted values, individual covariates, and linear combinations of the covariates (for example, dimension reduction techniques can suggest linear combinations more likely to reveal lack-of-fit). Conditional expectations of Y in the logistic regression context correspond to the probabilities p_{ij} in (1).

Ideally, model assessment requires equality (4) to be checked, but when $K + L > 2$ then $\mathrm{E}(Y|\boldsymbol{X}, \boldsymbol{G})$ is difficult to visualize. However, if h is univariate, $\mathrm{E}(Y|h)$ can be visualized in a two-dimensional scatterplot, and equality (5) can be checked. So, to assess the relationship between $\mathrm{E}_{\mathrm{F}}(Y|\boldsymbol{X}, \boldsymbol{G})$ and $\mathrm{E}_{\widehat{\mathrm{M}}}(Y|\boldsymbol{X}, \boldsymbol{G})$, instead compare $\mathrm{E}_{\mathrm{F}}(Y|h)$ and $\mathrm{E}_{\widehat{\mathrm{M}}}(Y|h)$ for various h. $\mathrm{E}_{\mathrm{F}}(Y|h)$ and $\mathrm{E}_{\widehat{\mathrm{M}}}(Y|h)$ can be estimated with non-parametric smooths such as cubic smoothing splines, the former by smoothing Y versus h, the latter by smoothing fitted values (probabilities), $\mathrm{E}_{\widehat{\mathrm{M}}}(Y|\boldsymbol{X}, \boldsymbol{G})$, versus h. Superimpose $\widehat{\mathrm{E}}_{\mathrm{F}}(Y|h)$ and $\widehat{\mathrm{E}}_{\widehat{\mathrm{M}}}(Y|h)$ 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 both smooths allows their point-wise comparison, since any estimation bias approximately cancels. Smooths that match closely for any function h provide support for the model; otherwise model inadequacy is indicated.

However, it can be difficult to judge whether smooths match closely without guidance on model uncertainty. Bayesian model assessment ideas in Ref. [4] provide one way to visualize this uncertainty. Consider drawing values of $\boldsymbol{\beta}_j$ in (1) from their posterior distributions, and generating a sample of realizations of Y from the model indexed by these $\boldsymbol{\beta}_j$. Repeat this process a large number m of times and compare the data Y-values to the m (posterior predictive) realizations from the model. Then, if the data "look like" a typical realization from the model there is no reason to doubt its fit. On the other hand, if the data appear to be very "unusual" with respect to the m model realizations, then the model is called into question. A graphical way to do this is to compare model-free smooths of data Y-values with model-based smooths of predicted probabilities (calculated using sampled $\boldsymbol{\beta}_j$ values). So, in a Bayes marginal model plot (BMMP), instead of superimposing just one model-based smooth, smooths for m model samples are superimposed; $m = 100$ provides good resolution in the plot without excessive computing overhead.

Figure 1 is a BMMP for the multilevel model for the imprisonment data with $h = \boldsymbol{X}_i^T \boldsymbol{G}_j \hat{\boldsymbol{\eta}} + \boldsymbol{X}_i^T \hat{\boldsymbol{\alpha}}_j$, where $\hat{\boldsymbol{\eta}}$ and $\hat{\boldsymbol{\alpha}}_j$ are posterior means.

[FIGURE 1 ABOUT HERE]

If, for a particular h, the black model-free smooth lies *substantially outside* the band of gray model-based smooths *or* it does not follow the general pattern of the gray model-based smooths, then the model is called into question. If, no matter what the function h is, the black model-free smooth lies *broadly inside* the gray model-based band *and* it follows the general pattern of the gray model-based smooths, then perhaps the model is a useful one. In Figure 1, the black smooth of the data passes close to the center of the gray band of model-based smooths of $1/(1 + \exp(-\boldsymbol{X}_i^T \boldsymbol{G}_j \boldsymbol{\eta}^* - \boldsymbol{X}_i^T \boldsymbol{\alpha}_j^*))$, where $\boldsymbol{\eta}^*$ and $\boldsymbol{\alpha}_j^*$ are 100 posterior samples. So, there is little indication of lack-of-fit from this plot.

Since fitting a multilevel logistic model requires substantially more computing time than a conventional (non-multilevel) model, it is instructive to compare BMMPs for a conventional model containing the same terms (main effects and interactions) as the multilevel model. If the conventional model fits essentially as well as the multilevel model, there would be little need to go to the trouble of fitting the latter. Furthermore, the different results produced by the multilevel and conventional models could impact the substantive conclusions. Nevertheless, a BMMP for the conventional model with $h = \boldsymbol{X}_i^T \boldsymbol{G}_j \hat{\boldsymbol{\eta}}$ (plot not shown), is qualitatively very similar to Figure 1. This is actually unsurprising since both models give similar predictions when averaging across counties, despite having very different posterior means for $\boldsymbol{\eta}$.

However, equality (5) should also match for subsets of the data, in particular within counties. Figure 2 contains BMMPs for one of the counties (number 13); the upper plot with $h = \boldsymbol{X}_i^T \boldsymbol{G}_j \hat{\boldsymbol{\eta}} + \boldsymbol{X}_i^T \hat{\boldsymbol{\alpha}}_j$ is for the multilevel model, the lower plot with $h = \boldsymbol{X}_i^T \boldsymbol{G}_j \hat{\boldsymbol{\eta}}$ is for the conventional model.

[FIGURE 2 ABOUT HERE]

Now the conventional model clearly appears to be inadequate, while the multilevel model continues to display no lack-of-fit. A similar assessment can be made for comparable BMMPs for the other counties (plots not shown).

Nevertheless, a series of BMMPs for various h-functions should be constructed to gain confidence in any particular model. In this application, since the models differ greatly on how county-level covariates are treated, consider the BMMPs with $h = \text{CCONS}$ in Figure 3.

[FIGURE 3 ABOUT HERE]

Again the conventional model appears inadequate, while the multilevel model shows promise. Differences are also apparent for the other county-level covariates (plots not shown).

Finally, consider a BMMP from the perspective of a county as the unit of analysis. The data Y-values now become proportions of individuals in the counties sentenced to prison. The model-based probabilities of receiving a prison sentence in each county can be obtained by averaging individual probabilities. County-level BMMPs can then be constructed with h now a function of county-level covariates, $\boldsymbol{G}$, only. BMMPs based on this premise with $h = \text{CBLPCT}$ are shown in Figure 4.

[FIGURE 4 ABOUT HERE]

Again the multilevel model seems better than the conventional one. Differences are also apparent for the other county-level covariates (plots not shown).

Using missing data imputation in fitting the model presents no additional difficulties in using BMMPs. The technique utilizes fitted values from the model, which were calculated for missing data cases in this application using posterior means of the regression coefficients used for imputation. Since equality (5) should hold for subsets of the data, BMMPs can also be constructed for just the non-missing data cases to provide a further check on model adequacy. Such plots for this analysis are qualitatively very similar to those above.

5 Discussion

In conclusion, the multilevel model appears to fit the U.S. imprisonment data well, and certainly improves on the conventional model which fits poorly. Ref. [14] provides a detailed discussion of the results of the multilevel analysis of this dataset.

The multilevel and conventional models produce some conflicting conclusions for this application, so using one model over the other has important policy implications. For example, the sum of the posterior means for the η-coefficients for CUNEMP and the CUNEMP by ICDRUG interaction is -0.31 for the multilevel model and 0.19 for the conventional model. The conventional model would appear to support the notion that punishment (for drug possession offenses) will be more severe in jurisdictions with greater proportions of individuals perceived as posing a threat because of their economic circumstances [10], whereas the multilevel model contradicts this. As noted by a referee however, unemployment may have more of an effect on the incarceration decision (jail or prison) so that an analysis with an indicator of “jail/prison” versus “non-custodial” as the response variable may produce different results.

As a further example of conflicting results from the multilevel and conventional analyses, the conventional model results suggest a positive CBLPCT by IBLACK interaction of 0.45 (indicating that African Americans may be punished more severely in counties with higher proportions of African Americans). However, the CBLPCT by IBLACK interaction is negligible (0.02) under the multilevel analysis, lending support to the notion that higher percentages of African Americans in these jurisdictions might function to increase African Americans’ political power, making racial bias less likely [5].

This paper has demonstrated how Bayes marginal model plots (BMMPs) can be extended to assessment of multilevel models containing random effects. Plots can be constructed at different levels of the hierarchy, for example at the individual level and the cluster level. The use of these plots for the imprisonment data illustrates the need to use multilevel modeling when covariates are measured at different levels in a hierarchical structure.

The BMMP methodology is not limited to logistic regression, and is generally applicable to any regression model. References to normal linear and additive model applications can be found in Ref. [13], which also contains further discussion of technical aspects of BMMPs such as calibration and smoothing. S-PLUS and R functions that can be used in conjunction with WinBUGS and BOA to construct BMMPs are available from the author's web-site.

Table 1

Individual-level covariates (May 1998) and county-level covariates (1998 unless specified)

IMALE	1: men, 0: women
IBLACK	1: African American, 0: otherwise
IACTCJS	1: active criminal justice status at time of offense
IPPRIS	1: prior stay(s) in state prison
IDETAIN	1: detained after being charged
IREVOKE	1: pretrial release was revoked
ITRIAL	1: convicted by trial, 0: convicted by plea
Most serious conviction charge (reference category includes weapons, driving-related, and other public order offenses):	
ICVIOL1	murder, rape or robbery (“more severe” violent)
ICVIOL2	assault, other violent crime (“less severe” violent)
ICTRAF	drug trafficking offense
ICDRUG	drug possession offense
ICPROP	burglary or theft (property offense)
CCRIME	index (known to police) crime rate per 10,000 residents
CUNEMP	unemployment rate (%)
CBLPCT	census estimate of African American population (%)
CCONS	share of vote for Bush in 2000 (%)
CSOUTH	1: located in a Southern state, 0: otherwise
CGUIDE	1: voluntary or mandatory state sentencing guidelines, 0: otherwise

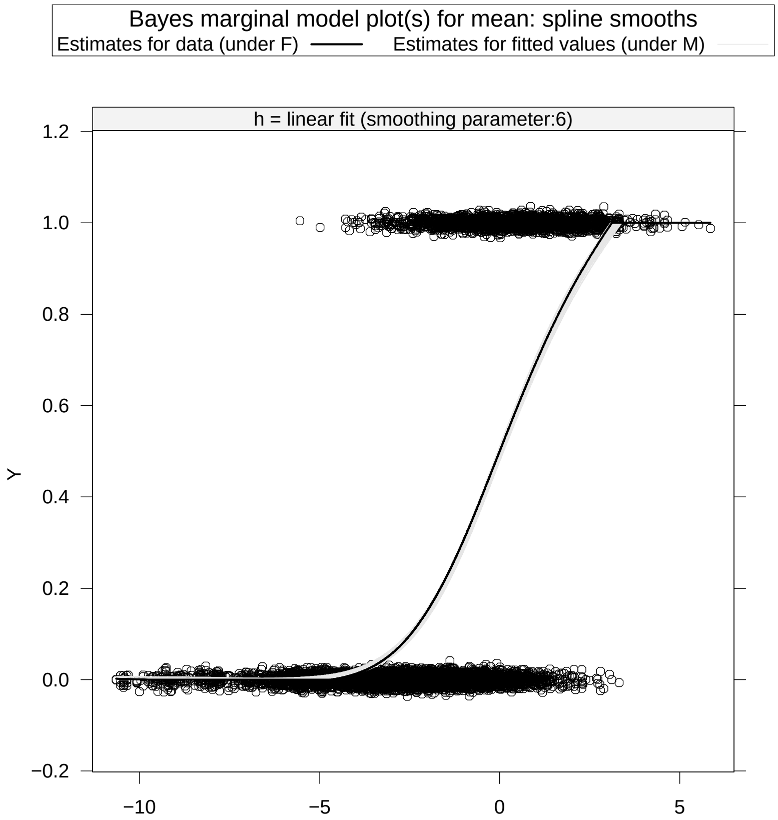
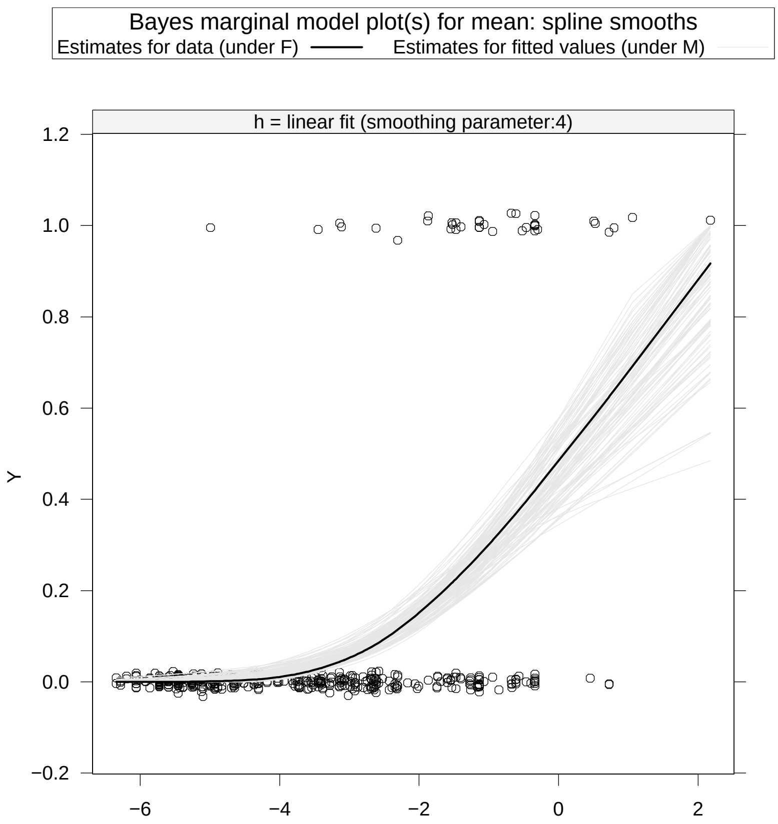


Fig. 1. Bayes marginal model plot (BMMP) for the multilevel model with $h = \boldsymbol{X}_i^T \boldsymbol{G}_j \hat{\boldsymbol{\eta}} + \boldsymbol{X}_i^T \hat{\boldsymbol{\alpha}}_j$. The data have been jittered to aid visualization of relative density and the smooths are smoothing splines with six effective degrees of freedom.



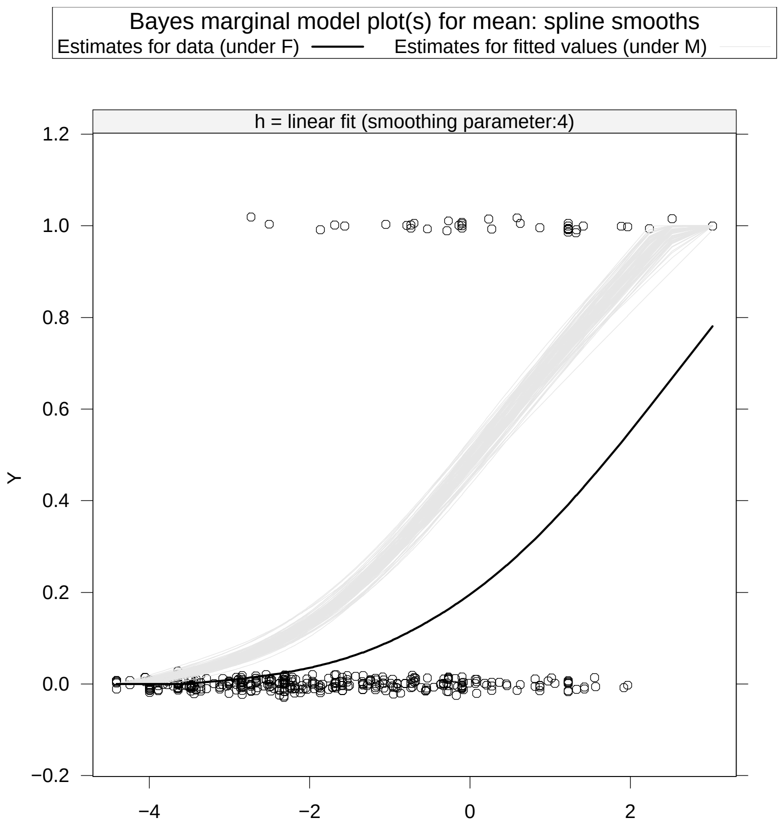


Fig. 2. BMMPs for county 13; $h = \boldsymbol{X}_i^T \boldsymbol{G}_j \hat{\boldsymbol{\eta}} + \boldsymbol{X}_i^T \hat{\boldsymbol{\alpha}}_j$ for multilevel model (upper) and $h = \boldsymbol{X}_i^T \boldsymbol{G}_j \hat{\boldsymbol{\eta}}$ for conventional model (lower). The smoothing splines have four effective degrees of freedom.

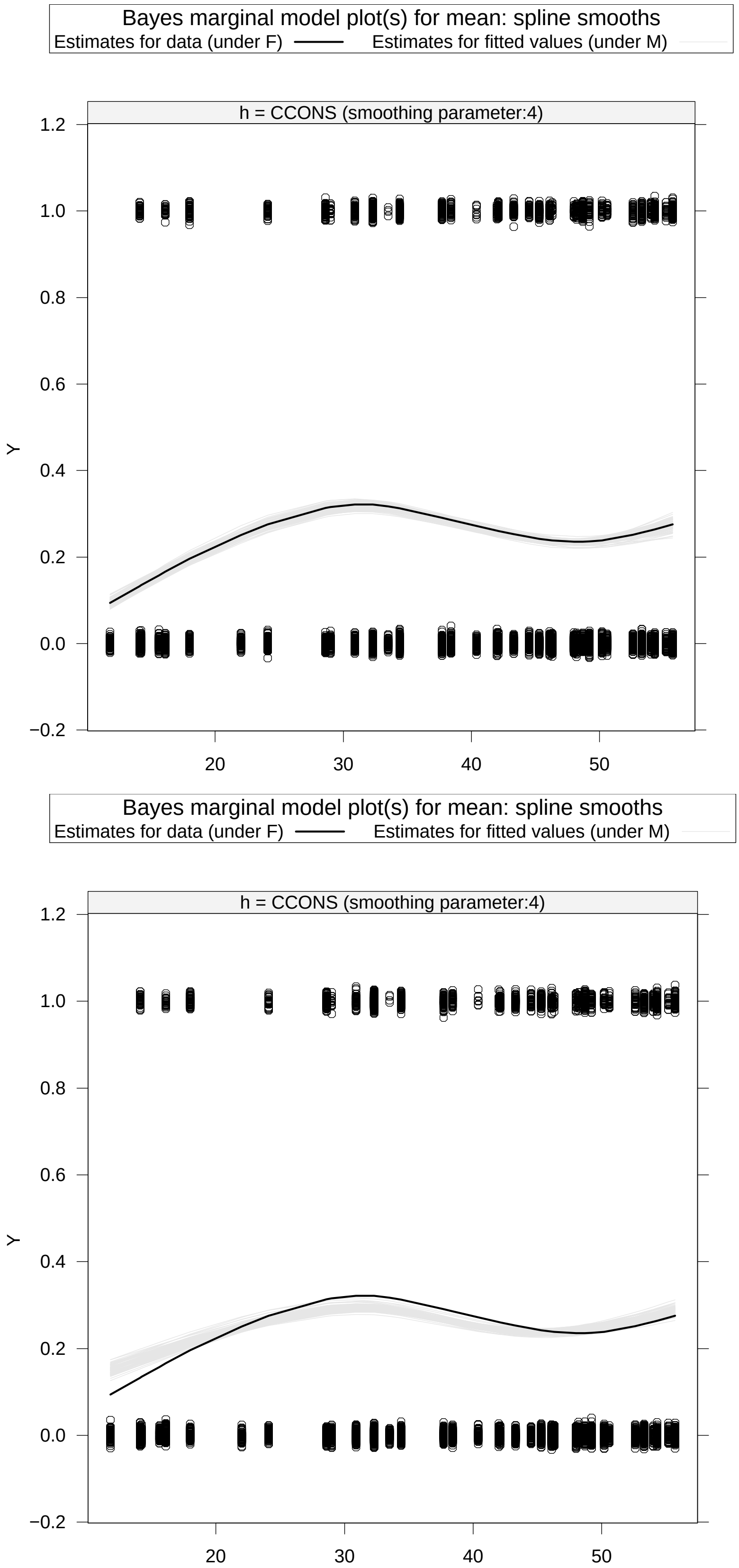


Fig. 3. BMMPs with $h = \text{C}\textsc{cons}$ for multilevel model (upper) and conventional model (lower). The smoothing splines have four effective degrees of freedom.

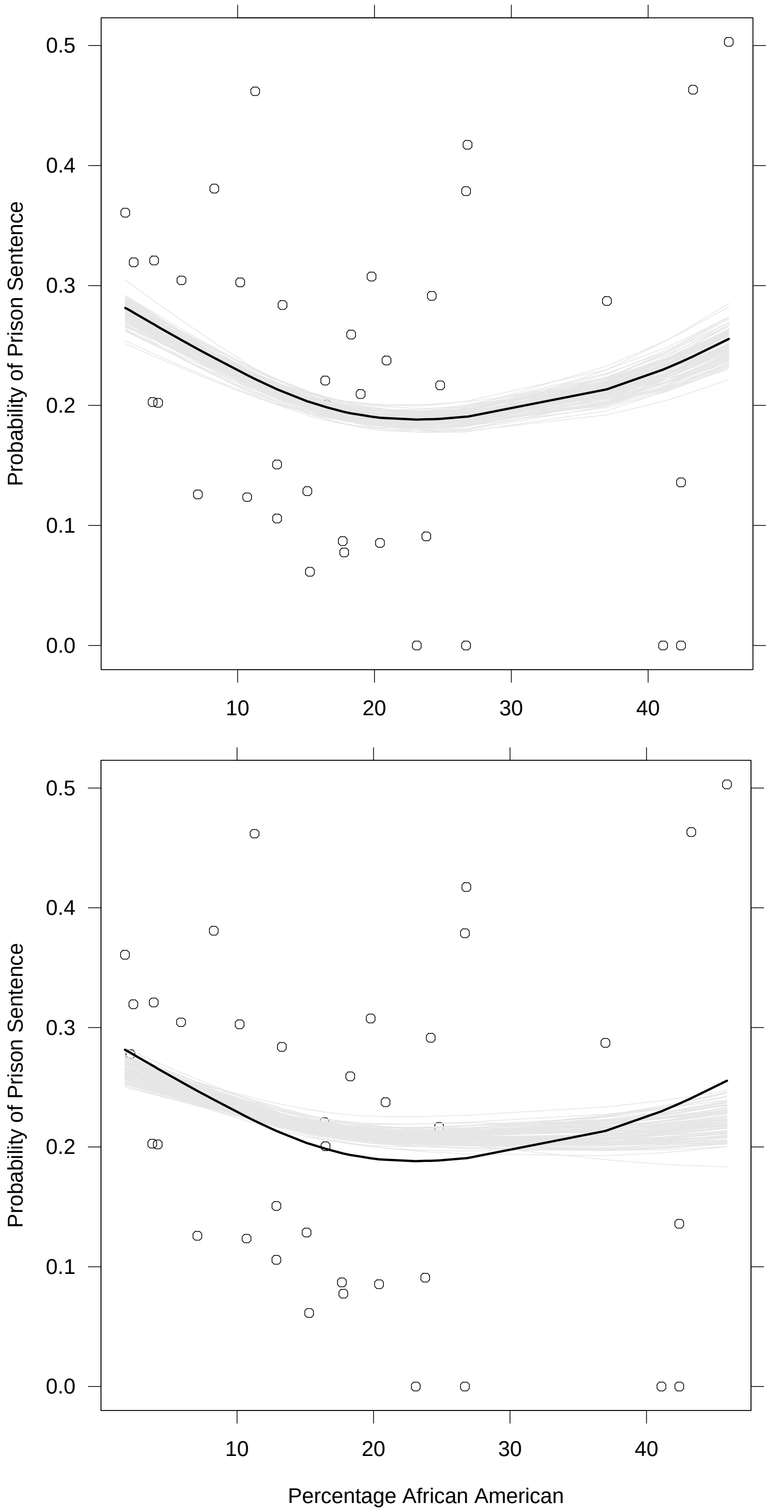


Fig. 4. County-level BMMPs with $h = \text{CBLPCT}$ for multilevel model (upper) and conventional (lower). The smoothing splines have four effective degrees of freedom.

References

[1] Britt, C. L. (2000). Social context and racial disparities in punishment decisions. *Justice Quarterly 17*, 707–732.

[2] Cook, R. D. and S. Weisberg (1997). Graphics for assessing the adequacy of regression models. *Journal of the American Statistical Association 92*, 490–499.

[3] Dixon, J. (1995). The organizational context of criminal sentencing. *American Journal of Sociology 100*, 1157–1198.

[4] 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.

[5] Greenberg, D. and V. West (2001). State prison populations and their growth, 1971–1991. *Criminology 39*, 615–654.

[6] Harrison, P. M. and A. J. Beck (2002). *Prisoners in 2001 (Bureau of Justice Statistics Bulletin)*. Washington, DC: Bureau of Justice Statistics.

[7] McCarthy, S. R. (1990). A micro-level analysis of social structure and social control: Intrastate use of jail and prison confinement. *Justice Quarterly 7*, 326–340.

[8] McGarrell, E. F. (1993). Institutional theory and the stability of a conflict model of the incarceration rate. *Justice Quarterly 10*, 7–28.

[9] Mears, D. P. (1998). The sociology of sentencing: Reconceptualizing decision-making processes and outcomes. *Law and Society Review 32*, 667–724.

[10] Mears, D. P. and S. H. Field (2000). Theorizing sanctioning in a criminalized juvenile court. *Criminology 38*, 983–1020.

[11] Michalowski, R. J. and M. A. Pearson (1990). Punishment and social structure at the state level: A cross-sectional comparison of 1970 and 1980. *Journal of Research in Crime and Delinquency 27*, 52–78.

[12] Myers, M. A. and S. M. Talarico (1987). *The Social Contexts of Criminal Sentencing*. New York: Springer-Verlag.

[13] Pardoe, I. and R. D. Cook (2002). A graphical method for assessing the fit of a logistic regression model. *The American Statistician 56(4)*, 263–272.

[14] Pardoe, I., R. R. Weidner, and R. Frase (2003). Sentencing convicted felons in the United States: A Bayesian analysis using multilevel covariates. Technical report, Charles H. Lundquist College of Business, University of Oregon.

[15] Rasbash, J., W. Browne, H. Goldstein, M. Yang, I. Plewis, M. Healy, G. Woodhouse, D. Draper, I. Longford, and T. Lewis (2000). *A User's Guide to MLwiN* (2nd ed.). London: Institute of Education.

[16] Raudenbush, S. W., A. S. Bryk, Y. F. Cheong, and R. Congdon (2001). *HLM 5: Hierarchical Linear and Nonlinear Modeling* (2nd ed.). Chicago: Scientific Software International.

[17] Rountree, P. W., K. C. Land, and T. D. Miethe (1994). Macro-micro integration in the study of victimization: A hierarchical logistic model analysis across Seattle neighborhoods. *Criminology 32*, 387–414.

[18] Spiegelhalter, D. J., A. Thomas, N. G. Best, and D. Lunn (2003). *WinBUGS Version 1.4 User Manual*. Cambridge, UK: MRC Biostatistics Unit.

[19] Steffensmeier, D., J. Kramer, and C. Streifel (1993). Gender and imprisonment decisions. *Criminology 31*, 411–446.

[20] Weidner, R. R. and R. Frase (2001). A county-level comparison of the propensity to sentence felons to prison. *International Journal of Comparative Criminology 1*, 1–22.

[21] Wong, G. Y. and W. M. Mason (1985). The hierarchical logistic regression model for multilevel analysis. *Journal of the American Statistical Association 80*, 513–524.

[22] Wooldredge, J. (1996). A state-level analysis of sentencing policies and inmate crowding in state prisons. *Crime and Delinquency 42*, 456–466.

[23] Wooldredge, J., T. Griffin, and T. Pratt (2001). Considering hierarchical models for research on inmate behavior: Predicting misconduct with multilevel data. *Justice Quarterly 18*, 203–231.