Empirical Bayes estimation for generalized linear models

by

Walter W. Piegorsch

and

George Casella

BU-1067-M

January 1990

Empirical Bayes estimation for generalized linear models

WALTER W. PIEGORSCH
Statistics and Biomathematics Branch, National Institute of Environmental Health Sciences

and

GEORGE CASELLA
Biometrics Unit, Cornell University

Empirical Bayes approaches are discussed for estimating the mean from an exponential family model. A generalized linear model is assumed, under which the mean, μ, is related to a linear combination of predictor variables via a so-called *link* function: $g(\mu) = \beta_1 x_1 + \cdots + \beta_p x_p$. $g(\cdot)$ defines the scale over which the systematic effects are modeled as additive. Bayes methods for this modeling scenario are considered that employ a conjugate exponential prior given by Albert (1988). These lead to (parametric) empirical Bayes estimators of the individual mean parameters. The conditional independence of the marginal distributions on the observations is utilized to construct a marginal likelihood for the hyperparameters of the model. (These include the predictor parameters, β_j, and a precision parameter, λ.) Maximum marginal likelihood estimators are calculated, and then substituted into the empirical Bayes estimates for the mean parameters. Empirical Bayes variance estimation is also discussed, emphasizing that simple substitution of the maximum marginal likelihood estimates for β_j and λ into the posterior variance for μ underestimates the actual posterior variance. Additional terms are required to correctly estimate the extra variability induced by estimation of β_j and λ. Specific details are provided for binomial and Poisson models, with added attention directed towards logistic and log-linear link specifications under these respective models. It is also noted that generalized linear models can be extended, using parametric families of link functions. These families embed specific links of interest — such as a logistic link — within their parametric structure. Empirical Bayes estimation under such extended model scenarios is addressed, with specific emphasis directed at the binomial model.

KEY WORDS: Binomial model; Exponential family; Extended link families; GLIM models; Logistic regression; Log-linear models; Non-linear regression; Poisson model; Shrinkage estimation

1. Introduction: Bayes approaches for generalized linear models

Generalized linear models (GLMs) are extensions of well-known (normal) linear models that allow for departures from both the normal distribution assumption and strict equality between the mean parameter, μ, and the linear predictor, η (McCullagh and Nelder 1983). The distribution of the response variables, Y_i, from a sample of size N is given an exponential family form:

$$f(y_i|\theta_i) = c(y_i,\varphi_i)\,exp\{\varphi_i[y_i\theta_i - b(\theta_i)]\} \tag{1.1}$$

where $c(y,\varphi)$ and $b(\theta)$ are known functions, and φ_i is a (set of) scale parameter(s) that will be assumed known. The unknown parameter θ_i is the *natural* parameter of the exponential family (Lehmann 1983, §1.4). It is related to the mean, $\mu_i = E[Y_i|\theta_i]$, via $\mu_i = b'(\theta_i)$.

Given a vector of predictor variables, $\mathbf{x}_i' = [x_{i1} \cdots x_{iP}]$, we define the linear predictor $\eta_i = \mathbf{x}_i'\beta$, where β is a P×1 vector of unknown regression parameters. The classical, non-Bayes approach to generalized linear modeling then *links* η_i to $\mu_i = b'(\theta_i)$ via a one-to-one function $g[b'(\theta_i)] = \mathbf{x}_i'\beta$. The inverse link function is denoted by $g^{-1}(\cdot)$.

This generalized linear structure allows for application of a wide variety of statistical methodologies for estimation of the mean response vector. For example, Bayes approaches have been discussed for linear model estimation under the exponential family in (1.1) (Albert 1988; West 1985), with specific applications noted to select members of the family such as binomial (Leonard 1972; Ramsey 1972; Santner and Duffey 1989, §5.4; Zellner and Rossi 1984), Poisson (Albert 1985; Brown and Farrell 1985), and of course the well-known normal distribution. The literature on Bayes estimation for the normal linear model is large; selected references of interest include Lindley and Smith (1972), Box and Tiao (1973, §2.7), Ramsey and Novick (1972), and Consonni and Dawid (1985).

A conjugate Bayes analysis for GLMs was recently given by Albert (1988). Denote the prior means of $b'(\theta_i)$ by m_i. Then, Albert suggested use of the prior exponential form

$$\pi(\theta_i|m_i,\lambda) = exp\{\lambda[m_i\theta_i - b(\theta_i)] + k(m_i,\lambda)\} \quad . \tag{1.2}$$

The hyper-parameter λ is assumed positive-valued. Under (1.2), one finds $var[b'(\theta_i)|m_i,\lambda] = E[b''(\theta_i)|m_i,\lambda]/\lambda$. To reflect the GLM, the prior means are assumed to satisfy $g(m_i) = \mathbf{x}_i'\beta$ (i=1,...N), where $g(\cdot)$ is the hypothesized link function. λ can then be employed as a precision parameter that reflects the degree of prior belief in the GLM. As Albert (1988) notes, sending $\lambda \rightarrow \infty$ concentrates μ_i about m_i in (1.2), collapsing the model, in effect, to a non-Bayesian, "classical" GLM.

In terms of the hierarchy of models assumed under this approach, we have the sampling distribution for the observations

$$f(y_i|\theta_i) \, ,$$

and the prior for the natural parameters

$$\pi(\theta_i|m_i,\lambda) \, .$$

Since we incorporate the GLM into the model via $g(m_i) = \mathbf{x}_i'\beta$, we can write the prior as

$$\pi(\theta_i|\beta,\lambda) \, .$$

The full Bayes approach also specifies a hyper-prior for the hyperparameters:

$$\pi(\beta,\lambda) \, .$$

For example, Albert (1988) chose the noninformative hyper-prior $\pi(\beta,\lambda) \propto (1+\lambda)^{-2}$ $(\lambda>0)$. He noted that this hierarchical model was therefore a generalization of the normal model from Lindley and Smith (1972). Albert also described computational approximations that allow for (unconditional) posterior inferences on the distribution of $\theta_i|\mathbf{y}$, and, in particular, on $\mu_i|\mathbf{y}$. Albert and Pepple (1989) described an exponential-mixture extension of this formulation, useful for modeling over-dispersion in GLMs.

Conditional on λ and β — or, equivalently, on λ and

$$m_i = g^{-1}(\mathbf{x}_i'\beta),$$

the inverse link — the posterior distribution of θ_i is of the form

$$\pi(\theta_i|y_i,m_i,\lambda) = exp\left\{ k\left(\frac{y_i\varphi_i + m_i\lambda}{\varphi_i+\lambda}, \varphi_i+\lambda\right) + (\varphi_i+\lambda)\left[\frac{y_i\varphi_i + m_i\lambda}{\varphi_i+\lambda}\theta_i - b(\theta_i)\right]\right\} \qquad (1.3)$$

with posterior mean for $b'(\theta_i)$ given by

$$E[b'(\theta_i)|y_i,m_i,\lambda] = \frac{y_i\varphi_i + m_i\lambda}{\varphi_i + \lambda} \, . \qquad (1.4)$$

Notice that the posterior variance for $b'(\theta_i)$ is $var[b'(\theta_i)|y_i,m_i,\lambda] = E[b''(\theta_i)|y_i,m_i,\lambda]/(\lambda+\varphi_i)$. A Bayes estimator for μ_i is the posterior mean in (1.4):

$$\mu_i^B = \frac{y_i\varphi_i + m_i\lambda}{\varphi_i + \lambda} \qquad (i=1,\ldots,N) \qquad ,$$

conditional, of course, on λ and m_i (φ_i known). Also, the marginal distributions of $y_i|m_i,\lambda$ are conditionally independent, with

$$f(y_i|m_i,\lambda) = c(y_i\,,\varphi_i)\,exp\left\{ k(m_i\,,\lambda) - k\left(\frac{y_i\varphi_i + m_i\lambda}{\varphi_i+\lambda}\,,\varphi_i+\lambda\right)\right\} \qquad (1.5)$$

($i=1,\ldots,N$) (Albert 1988).

As an alternative to full-scale Bayes analyses for such GLMs, we will consider (parametric) empirical Bayes methods for estimating μ_i. The general approach using Albert's conjugate model is described in §2, along with a short review of empirical Bayes approaches for this generalized linear setting. Details for binomial and Poisson regression are given in §3, where an example using a binomial-logistic model is also presented. Some extensions are noted in §4.

2. Parametric empirical Bayes methods for GLMs

2.1 Review

The empirical Bayes (EB) paradigm is based on the concept that information about an unspecified (or less-than-fully-specified) prior distribution may be garnered from the marginal distribution of the Y_i. The approach dates back to Robbins (1955), who formalized what is commonly called "non-parametric" EB analysis. Robbins' approach is non-parametric in the sense that the entire prior distribution is left unspecified; see Maritz and Lwin (1989, §§1-2). Our emphasis will be directed towards *parametric* EB analysis (Morris 1983), where the form of the prior distribution is pre-specified, and complete specification via intermediate estimation of the prior parameters is of interest. Kass and Steffey (1989) refer to this structure as a conditionally independent hierarchical model.

Methods of EB analysis have been extensively discussed (see Casella (1985)), including applications to various distributional members of the exponential family. These include estimation of binomial probabilities (Albert 1984; Berry and Christensen 1979; Brier, Zacks and Marlow 1986), Poisson means (Hudson 1985), and, of course, normal means (Casella and Hwang 1983; Laird and Louis 1989). Martz and Waller (1982, §13) give a useful parametric EB overview for many of the exponential family members, and some other distributions, with applications to reliability analysis.

For analyses under a normal linear model, Singh (1985) discusses non-parametric EB estimation, while Nebebe and Stroud (1986) give parametric EB estimators. For specific EB applications under a normal linear model, see Rubin (1980), DuMouchel and Harris (1983), Strenio et al. (1983), Hui and Berger (1983), and Louis (1989). EB applications for a Poisson regression model are noted by O'Bryan (1979), Clayton and Kaldor (1987), and Hudson (1985).

For binomial linear modeling, Duffy and Santner (1989) discuss EB methodology directed at logistic regression, i.e., for the specific case of $g(\mu) = \mathit{log}\{\mu/(1-\mu)\}$, where μ is the probability of response under binomial sampling. Duffy and Santner effectively assume a multivariate normal prior on the regression parameters, $\beta \sim N(\mathbf{0},\sigma^2\mathbf{I})$ (in similar fashion to the full-scale Bayes approach taken by Zellner and Rossi (1984)), and then employ a form of the EM algorithm (Dempster, Laird and Rubin 1977) to achieve an EB estimate of σ^2; see also Santner and Duffy (1989, §5.4). This is similar to the approach taken by Laird (1978) for EB estimation of effects in two-way contingency tables.

2.2 Introducing a GLM into the EB format

Empirical Bayes estimators are often recognized as belonging to a class known as *shrinkage* estimators (Casella 1988). That is, they often may be interpreted as "shrinking" an estimate away from the observations or cell means, and towards some fitted sub-model. The exponential family discussed herein is a good example of this phenomenon: For the family in (1.1), and under the conjugate prior (1.2), we construct a (marginal) log-likelihood function associated with the parameters $\mathbf{m}' = [m_1 \cdots m_N]$ and λ. Incorporating the GLM gives $m_i = g^{-1}(\mathbf{x}_i'\beta)$. Maximizing this function with respect to β and λ yields marginal maximum likelihood (MML) estimates $\hat{\beta}$ and $\hat{\lambda}$, which are substituted into the Bayes estimator of μ_i:

$$\mu_i^{EB} = \frac{\varphi_i}{\varphi_i + \hat{\lambda}} y_i + \frac{\hat{\lambda}}{\varphi_i + \hat{\lambda}} g^{-1}(\mathbf{x}_i'\hat{\beta}) \quad , \tag{2.1}$$

($i=1,\ldots,N$). As $\hat{\lambda}$ increases away from zero, μ_i^{EB} "shrinks" away from y_i and towards the estimated prior mean $\hat{m}_i = g^{-1}(\mathbf{x}_i'\hat{\beta})$.

To find the MML estimates, we view the joint marginal density as a marginal likelihood,

$$L(\beta,\lambda) = \prod_{i=1}^{N} f(y_i \,|\, g_i^{-1},\lambda)$$

(notice the use of the conditional independence of the y_i). The individual marginal densities, $f(y_i|m_i,\lambda)$, are given in (1.5). The marginal log-likelihood becomes

$$\ell(\beta,\lambda) = \sum_{i=1}^{N} log\,\{c(y_i\,,\varphi_i)\} + \sum_{i=1}^{N} log \left\{ k(m_i,\lambda) - k\left(\frac{y_i\varphi_i + m_i\lambda}{\varphi_i + \lambda}, \varphi_i + \lambda\right)\right\} \quad (2.2)$$

The MML estimates maximize (2.2) with respect to β and λ. Kass and Steffey (1989) note that employing the MML estimates in (2.1) actually approximates, to order 1/N, the full Bayes estimator μ_i^B under an (improper) independent uniform hyper-prior on β and λ. Thus the (parametric) EB estimates possess interpretation as approximations to the corresponding Bayes estimates under vague prior information.

Estimation of the posterior standard deviation or, effectively, the posterior variance of μ_i is also possible under an EB format. It is inappropriate, however, to mimic the EB estimate for $\mu_i = b'(\theta_i)$ and simply substitute the MML estimates $\hat{m}_i$ and $\hat{\lambda}$ into $var[b'(\theta_i)|y_i,m_i,\lambda]$, since the simple substitution fails to take account of the variation in the MML estimators. Several authors have recognized this phenomenon — e.g., Casella (1988), Laird and Louis (1989), Kass and Steffey (1989) — and note that an additional term is required in the variance estimator to account for this increase in variability. That is, the variance, $var[b'(\theta_i)|y_i,m_i,\lambda]$, evaluated at $\hat{m}_i$ and $\hat{\lambda}$ is, formally,

$$var[b'(\theta_i)|y_i,\, m_i{=}\hat{m}_i,\, \lambda{=}\hat{\lambda}]\ ,$$

which may be appropriate as a point estimate of the posterior variance of $b'(\theta_i)$, but not a variance estimate of μ_i^{EB}. Since the true values of m_i and λ are unknown, the variance estimator of μ_i^{EB}, to be useful, must be unconditional on m_i and λ. To achieve such, a standard calculation shows

$$var[b'(\theta_i)|y] = E\{var[b'(\theta_i)|y_i,m_i,\lambda]\} + var\{E[b'(\theta_i)|y_i,m_i,\lambda]\}$$

where the outer expectation and variance are taken with respect to the joint distribution of m_i and λ (cf. Kass and Steffey (1989)). Thus $var[b'(\theta_i)|y_i,\, m_i{=}\hat{m}_i,\, \lambda{=}\hat{\lambda}]$ gives an estimate of only the first portion of $var[b'(\theta_i)|y]$; an additional term is required to avoid variance underestimation.

Following Kass and Steffey (1989), we construct the additional term for the variance estimator by assuming the existence of some hyper-prior for β and λ, say $\pi(\beta,\lambda)$, and note that the log-posterior likelihood with respect to these parameters has the form $log\{L(\beta,\lambda)\pi(\beta,\lambda)\}$. The corresponding inverse negative Hessian matrix is denoted by $\Sigma = \{\sigma_{jk}\}$. If we take $\pi(\beta,\lambda)$ to represent an independent, uniform hyper-prior, then Σ is simply the inverse negative Hessian matrix for the MML function, with

$$\Sigma^{-1} = \begin{bmatrix} \left\{-\frac{\partial^2 \ell}{\partial\beta_j \partial\beta_k}\right\} & \left\{-\frac{\partial^2 \ell}{\partial\beta_j \partial\lambda}\right\} \\ & -\frac{\partial^2 \ell}{\partial\lambda^2} \end{bmatrix} .$$

Denoting $\hat{\sigma}_{jk}$ as the (j,k)th element of Σ evaluated at the MML estimate, we obtain (Kass and Steffey 1989)

$$\text{var}[b'(\theta_i)\,|\,y] \approx \text{var}[b'(\theta_i)\,|\,y_i, \hat{m}_i, \hat{\lambda}] + \sum_j \sum_k \hat{\sigma}_{jk}\hat{\delta}_j\hat{\delta}_k \quad , \qquad (2.3)$$

where

$$\hat{\delta}_j = \frac{\partial}{\partial\beta_j} E[b'(\theta_i)|y_i, m_i, \lambda]\Big|_{\hat{\beta}\hat{\lambda}} \qquad (j=1,\ldots,P)$$

and

$$\hat{\delta}_{P+1} = \frac{\partial}{\partial\lambda} E[b'(\theta_i)|y_i, m_i, \lambda]\Big|_{\hat{\beta}\hat{\lambda}} \quad .$$

These latter quantities simplify to

$$\hat{\delta}_j = \frac{\hat{\lambda}}{\varphi_i + \hat{\lambda}} \frac{\partial m_i}{\partial\beta_j}\Big|_{\hat{\beta}\hat{\lambda}} \qquad (j=1,\ldots,P)$$

and $\hat{\delta}_{P+1} = \varphi_i(\hat{m}_i - y_i)/(\varphi_i + \hat{\lambda})^2$. Also, some simplification in $\text{var}[b'(\theta_i)|y_i,\hat{m}_i,\hat{\lambda}]$ may be available by recalling that $\text{var}[b'(\theta_i)|y_i,m_i,\lambda] = E[b''(\theta_i)|y_i,m_i,\lambda]/(\varphi_i + \lambda)$, i=1,...,N.

3. Distribution-specific examples

3.1 Binomial model

Suppose we observe proportions $y_i = w_i/\varphi_i$, where the φ_i are (known) positive integers and $W_i \sim$ (indep.) $b(\varphi_i,\mu_i)$, along with a set of predictor variables $x_i' = [x_1 \cdots x_p]$, i=1,...,N. The parameters of interest are the probabilities μ_i; the natural parameter is related to these probabilities via $\theta_i = log\{\mu_i/(1-\mu_i)\}$. Hence, $\mu_i = b(\theta_i) = log\{1 + exp(\theta_i)\}$ and $b'(\theta_i) = 1/\{1 + exp(-\theta_i)\}$. Also, for this density, $c(y,\varphi)$ is the binomial coefficient

$$\binom{\varphi}{\varphi y}.$$

We leave the form of the link function unspecified for now, but will examine a special case (a logistic link) in the data example below.

Under Albert's (1988) conjugate prior (1.2), we specify

$$k(m_i,\lambda) = \frac{log\left\{\Gamma(\lambda)\right\}}{log\left\{\Gamma(m_i\lambda)\right\} + log\left\{\Gamma(\lambda - m_i\lambda)\right\}} \tag{3.1}$$

where $\Gamma(\cdot)$ is the (complete) gamma function (Abramowitz and Stegun 1972, equ. 6.1.1). This corresponds to the familiar beta prior for μ_i over the unit interval: $\pi(\mu_i|a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\mu_i^{a-1}(1-\mu_i)^{b-1}$, with $a = m_i\lambda$ and $b = \lambda - m_i\lambda$. In its more general form on $\theta = log\{\mu/(1-\mu)\}$, this is

$$\pi(\theta|a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)}(1+e^{-\theta})^{1-a}(1+e^{\theta})^{1-b}\frac{e^{-\theta}}{(1+e^{-\theta})^2}.$$

This is a form of beta-logistic density, noted, e.g., by Leonard et al. (1989), since at $a=b=1$ we recover the standard logistic density $e^{-\theta}/(1+e^{-\theta})^2$.

At $a = m_i\lambda$ and $b = \lambda - m_i\lambda$, and under $m_i = g^{-1}(x_i'\beta)$, we construct the (marginal) likelihood function $\ell(\beta,\lambda)$ based on (2.2). This is found to have gradient elements

$$\frac{\partial \ell}{\partial \beta_j} = \lambda \sum_{i=1}^{N} \frac{\partial m_i}{\partial \beta_j}\left\{\psi(\varphi_i y_i + m_i\lambda) - \psi(\varphi_i - \varphi_i y_i + \lambda - m_i\lambda) - \psi(m_i\lambda) + \psi(\lambda - m_i\lambda)\right\}$$

(j=1,...,P) and

$$\frac{\partial \ell}{\partial \lambda} = N\psi(\lambda) + \sum_{i=1}^{N} \Big\{ m_i \big[\psi(\varphi_i y_i + m_i\lambda) - \psi(m_i\lambda)\big]$$

$$+ (1-m_i)\big[\psi(\varphi_i - \varphi_i y_i + \lambda - m_i\lambda) - \psi(\lambda - m_i\lambda)\big]\Big\} \quad .$$

The function $\psi(\cdot)$ is the di-gamma function: $\psi(\omega) = \partial log\Gamma(\omega)/\partial\omega$. Since this function satisfies

$$\psi(\omega + n) = \psi(\omega) + \textstyle\sum_{\nu=1}^{n} (\omega + \nu - 1)^{-1}$$

for integer-values of n (Abramowitz and Stegun 1972, equ. 6.3.6), the binomial's marginal likelihood gradients simplify to

$$\frac{\partial \ell}{\partial \beta_j} = \lambda \sum_{i=1}^{N} \frac{\partial m_i}{\partial \beta_j} \left\{ \sum_{\nu=1}^{w_i} \frac{1}{m_i\lambda + \nu - 1} - \sum_{\nu=1}^{\varphi_i - w_i} \frac{1}{m_i\lambda + \nu - 1} \right\}$$

$$\frac{\partial \ell}{\partial \lambda} = \sum_{i=1}^{N} \left\{ \sum_{\nu=1}^{w_i} \frac{m_i}{m_i\lambda + \nu - 1} - \sum_{\nu=1}^{\varphi_i} \frac{1}{\lambda + \nu - 1} + \sum_{\nu=1}^{\varphi_i - w_i} \frac{1 - m_i}{\lambda - m_i\lambda + \nu - 1} \right\} \tag{3.2}$$

where $w_i = y_i\varphi_i$, i=1,...,N; j=1,...,P. Specification of the link function $g(\cdot)$ allows for computation of m_i and $\partial m_i/\partial\beta_j$ in (3.2).

Solving $\partial\ell/\partial\beta_j = 0$ (j=1,...,P), $\partial\ell/\partial\lambda = 0$ yields MML estimates $\hat{\beta}$ and $\hat{\lambda}$. (In any such calculations below, we employed a gradient search algorithm, such as given by Fletcher (1970).) These estimates are then employed in (2.1).

Variance estimation is achieved via (2.3); specific forms for the second derivatives of ℓ that make up the elements of Σ^{-1} are given in the Appendix (§A.1). These quantities are evaluated at $\hat{\beta}$, $\hat{\lambda}$ for use in $\hat{\Sigma} = \{\hat{\sigma}_{jk}\}$. Also, $b'(\theta) = 1/(1 + e^{-\theta})$, $b''(\theta) = e^{-\theta}/(1+e^{-\theta})^2$, so $b''(\theta) = b'(\theta) - \{b'(\theta)\}^2$. Thus, equating $\mathrm{var}[b'(\theta_i)|y_i,m_i,\lambda] = E[b''(\theta_i)|y_i,m_i,\lambda]/(\varphi_i + \lambda) = E[(b'(\theta_i) - \{b'(\theta)\}^2)|y_i,m_i,\lambda]/(\varphi_i + \lambda)$ with the well-known relation $\mathrm{var}[b'(\theta_i)|y_i,m_i,\lambda] = E[\{b'(\theta)\}^2|y_i,m_i,\lambda] - E^2[b'(\theta)|y_i,m_i,\lambda]$ yields an expression for $\mathrm{var}[b'(\theta_i)|y_i,m_i,\lambda]$ in terms of only $E[b'(\theta)|y_i,m_i,\lambda]$ and $\varphi_i+\lambda$. Substituting the MML estimates into this expression gives

$$\mathrm{var}[b'(\theta)|y_i,\hat{m}_i,\hat{\lambda}] = \frac{E[b'(\theta)|y_i,\hat{m}_i,\hat{\lambda}]}{\varphi_i + \hat{\lambda} + 1}\left(1 - E[b'(\theta)|y_i,\hat{m}_i,\hat{\lambda}]\right)$$

or simply $\mathrm{var}[b'(\theta_i)|y_i,\hat{m}_i,\hat{\lambda}] = \mu_i^{EB}(1-\mu_i^{EB})/(\varphi_i+\hat{\lambda}+1)$. This is employed in the variance estimator based on (2.3):

$$v_i^{EB} = \frac{\mu_i^{EB}(1-\mu_i^{EB})}{\varphi_i + \hat{\lambda} + 1} + \sum_j \sum_k \hat{\sigma}_{jk} \hat{\delta}_j \hat{\delta}_k \quad .$$

3.2 Empirical Bayes logistic regression

Specification of a logistic link function for $g(\cdot)$ under the binomial model results in a form of EB logistic regression. Thus, for

$$m_i = (1 + exp\{-x_i'\beta\})^{-1}$$

we find $\partial m_i/\partial\beta_j = x_{ij} m_i^2 exp\{-x_i'\beta\}$ for use in (3.2). The EB estimates of μ_i can be written as

$$\mu_i^{EB} = \frac{\varphi_i}{\varphi_i + \hat{\lambda}} y_i + \frac{\hat{\lambda}}{(\varphi_i + \hat{\lambda})\left(1 + exp\left\{-x_i'\hat{\beta}\right\}\right)}$$

(i=1,...,N). Variance estimates, v_i^{EB}, are based on, e.g.,

$$\frac{\partial^2 m_i}{\partial\beta_j \partial\beta_k} = x_{ij} x_{ik} m_i^3 exp\{-x_i'\beta\}\,(exp\{-x_i'\beta\} - 1)\ ,$$

for use in (A.1). Also, $\hat{\delta}_j = \hat{\lambda} x_{ij} (\hat{m}_i)^2 exp\{-x_i'\hat{\beta}\}/(\varphi_i + \hat{\lambda})$ (j=1,...,P) is necessary for the computation of v_i^{EB}.

As an illustrative example, consider the toxicological data given by Dunnick et al. (1988). Those authors reported on nasal toxicity in rodents after inhalation exposure to the solvent stabilizer 1,2-epoxybutane. For instance, in male rats the observed rates of nasal inflammation at the three dose exposures x_1=0 (control), x_2=200, and x_3=400 ppm were seen to be y_1=0.18, y_2=0.72, and y_3=0.84, with $\varphi_i = 50\ \forall i$. Modeling the dose response via the logistic model as detailed above, and employing a simple linear predictor in dose — $\eta_i = \beta_1 + \beta_2 x_i$ (P=2) — yields the MML estimates $\hat{\lambda}$=102.22 and $\hat{\beta}$=[-1.18 0.0082]′. The estimated prior means are $(1 + exp\{-[\hat{\beta}_1 + \hat{\beta}_2 x_i]\})^{-1}$ = 0.236, 0.613, and 0.891 (i=1,2,3, respectively), with associated EB estimates

x_i	0	200	400
μ_i^{EB}	0.216	0.649	0.875
v_i^{EB}	3.01×10^{-3}	3.06×10^{-3}	1.66×10^{-3}.

3.3 Poisson model

An additional model of interest for regression analysis of discrete data involves the Poisson distribution: $Y_i \sim$ Poisson(u_i), i=1,...,N. Some Bayes approaches for fitting Poisson regressions were noted above; classical methods for fitting Poisson regression are discussed, e.g., in Frome et al. (1973), Frome (1983), and Lawless (1987).

Within the exponential family as given in (1.1), the scale parameters under the Poisson model are all set to unity. The function $c(y_i,\varphi_i)$ is simply $c(y_i,\varphi_i) = 1/y_i!$, while $b(\theta_i) = exp\{\theta_i\}$.

To employ an EB approach for estimation of the μ_i under a Poisson linear model, we consider the prior specification $k(m_i,\lambda) = m_i\lambda\, log\lambda - log\{\Gamma(m_i\lambda)\}$ into the conjugate prior (1.2). This corresponds to the familiar gamma prior on μ_i:

$$\pi(\mu_i\,|m_i\,,\lambda) = \frac{\lambda^{m_i\lambda}\mu_i^{m_i\lambda-1} exp\left\{-\lambda\mu_i\right\}}{\Gamma(m_i\,\lambda)} \quad .$$

Thus, given a link specification for $m_i=g^{-1}(x_i'\beta)$, we calculate EB estimates using the MML estimates, as above. The marginal log-likelihood function can be shown to take the form

$$\ell(\beta,\lambda) = \textstyle\sum_{i=1}^{N}\left\{ log(1/y_i!)\} + \lambda m_i log\lambda - log\Gamma(\lambda m_i) - (y_i+\lambda m_i) log\Gamma(\lambda+1) + log\Gamma(y_i+\lambda m_i) \right\} .$$

Low-order derivatives of this Poisson (marginal) likelihood (i.e., gradient and negative Hessian elements) are given in §A.2. Recall that the negative Hessian elements are evaluated at the MML estimates in constructing $\hat{\Sigma} = \{\hat{\sigma}_{jk}\}$ for calculation of v_i^{EB}; see (2.3). We also require

$$\hat{\delta}_j = \frac{\hat{\lambda}}{\hat{\lambda}+1}\frac{\partial m_i}{\partial\beta_j}\bigg|_{\hat{\beta}}$$

(j=1,...,P) and $\hat{\delta}_{P+1} = (\hat{m}_i - y_i)/(\hat{\lambda} + 1)^2$. Since $b'(\theta_i) = b''(\theta_i) = exp\{\theta_i\}$, we have $var[b'(\theta_i)|y_i,m_i,\lambda] = E[b''(\theta_i)|y_i,m_i,\lambda]/(\lambda+1) = E[b'(\theta_i)|y_i,m_i,\lambda]/(\lambda+1)$. Hence we employ $var[b'(\theta_i)|y_i,\hat{m}_i,\hat{\lambda}] = \mu_i^{EB}/(\hat{\lambda}+1)$ (i=1,..,N) for use in

$$v_i^{EB} = \frac{\mu_i^{EB}}{\hat{\lambda}+1} + \sum_j\sum_k \hat{\sigma}_{jk}\hat{\delta}_j\hat{\delta}_k \quad .$$

A specific (and common) choice for the link is a log-linear characterization. This is $g(\mu) = log\mu$, with $m_i = exp\{x_i'\beta\}$, so that $\partial m_i/\partial\beta_j = x_{ij}exp\{x_i'\beta\} = x_{ij}m_i$ and $\partial^2 m_i/\partial\beta_j\partial\beta_k = x_{ij}x_{ik}m_i$.

4. Extensions

4.1 Fitting extended parametric link families

The specification of the link function $g(\cdot)$ in the GLM can be *extended* into a larger family, say with the function $g(\mu;\gamma)$. The (possibly vector-valued) parameter γ identifies the parametric family of interest. Common choices for this extended link employ basic forms from which more complex functions are constructed. For instance, much of the activity in constructing extended link families has been directed at the binomial model (Morgan 1988); therein, one can embed the popular logistic link into a larger family in assorted ways, such as

$$g(\mu;\gamma) = log\left\{\frac{\mu + (1-\mu)\gamma}{1-\mu}\right\} \quad . \tag{4.1}$$

The logistic link is recovered in (4.1) at $\gamma=0$, while $\gamma=1$ corresponds to a complementary-log model (also called a "one-hit" model). Whittemore (1983) employs (4.1) in comparing logistic versus exponential sub-models for carcinogenicity data. Zellner and Rossi (1984) discuss a similar issue — model selection between a logit and probit fit — employing instead a posterior odds ratio in their selection process; also see Smith and Spiegelhalter (1980).

A parametric EB approach to estimation of μ_i under an extended link follows in similar fashion to that noted in §2, above. Construct the (extended) marginal likelihood

$$\ell(\beta,\lambda,\gamma) = \sum_{i=1}^{N}\left\{ log\,\{c(y_i,\varphi_i)\} + k[g^{-1}(x_i'\beta;\gamma)\,,\lambda] - k\left[\frac{y_i\varphi_i + \lambda g^{-1}(x_i'\beta;\gamma)}{\varphi_i + \lambda}\,,\,\varphi_i + \lambda\right]\right\}$$

and find MML estimates $\hat{\lambda}$, $\hat{\beta}$, and $\hat{\gamma}$ by maximizing $\ell(\beta,\lambda,\gamma)$. We also write $\hat{m}_i = g^{-1}(x_i'\hat{\beta};\hat{\gamma})$. The EB estimates are then computed using an obvious extension of (2.1).

Variance estimation is accomplished as above: implicitly assuming an independent vague (uniform) hyper-prior for γ under the Kass-Steffey (1989) formulation, we form the negative inverse Hessian matrix for the P+2 parameters β, λ, and γ, and then evaluate the individual elements at the MML estimates. We also require

$$\hat{\delta}_{P+2} = \left.\frac{\partial E[b'(\theta)|y_i, m_i, \lambda]}{\partial\gamma}\right|_{\hat{\beta}\hat{\lambda}\hat{\gamma}} = \left.\frac{\hat{\lambda}}{\varphi_i + \hat{\lambda}}\frac{\partial m_i}{\partial\gamma}\right|_{\hat{\beta}\hat{\lambda}\hat{\gamma}}$$

for use in (2.3).

4.2 Example: Binomial model (continued)

Continuing our development of the binomial model, we can construct an extended likelihood under a beta prior for μ_i in the same form as seen in §3.1, except that m_i is now a function of the γ-parameter: $m_i = g^{-1}(x_i'\beta;\gamma)$. Thus, equations (3.2) and (A.1) retain validity. Some additional low-order derivatives required for MML estimation of γ and of the v_i^{EB} are given in §A.3.

As an extension of the logistic example discussed in §3.2, we employed Whittemore's (1983) sub-exponential family in (4.1) to the rodent nasal toxicity data discussed therein. Specifically, we took

$$m_i = \begin{cases} \dfrac{exp\left\{x_i'\beta\right\} - \gamma}{1 + exp\left\{x_i'\beta\right\} - \gamma} & \text{if } \gamma < exp\left\{x_i'\beta\right\} \\ 0 & \text{otherwise} \end{cases} \tag{4.2}$$

Applied to the male rat nasal inflammation data, this model achieves MML estimates of $\hat{\lambda}$ = 129.25, $\hat{\beta}$ = [2.72 7.06×10^{-4}]′, and $\hat{\gamma}$ = 15.02. These yield the following extended EB estimates:

x_i	0	200	400
μ_i^{EB}	0.182	0.718	0.839
v_i^{EB}	2.92×10^{-3}	2.68×10^{-3}	1.63×10^{-3}

If desired, greater accuracy in the estimation process may be attainable by moving to second-order formulae for posterior moments of $b'(\theta)$, as given by Kass and Steffey (1989). This is typically achieved under greater computational expense, however.

Appendix

A.1. Marginal likelihood derivatives for binomial model

Variance estimation under (2.3) for the binomial model from §3.1 requires the elements of the negative Hessian matrix, Σ^{-1}. These include

$$-\frac{\partial^2 \ell}{\partial\beta_j \partial\beta_k} = \lambda \sum_{i=1}^{N} \left\{ \lambda \frac{\partial m_i}{\partial\beta_j}\frac{\partial m_i}{\partial\beta_k} \left[\sum_{v=1}^{w_i} \frac{1}{(m_i\lambda + v - 1)^2} + \sum_{v=1}^{\varphi_i - w_i} \frac{1}{(m_i\lambda + v - 1)^2} \right] \right.$$

$$\left. - \frac{\partial^2 m_i}{\partial\beta_j \partial\beta_k} \left[\sum_{v=1}^{w_i} \frac{1}{m_i\lambda + v - 1} - \sum_{v=1}^{\varphi_i - w_i} \frac{1}{\lambda - m_i\lambda + v - 1} \right] \right\}$$

$$-\frac{\partial^2 \ell}{\partial\beta_j \partial\lambda} = \sum_{i=1}^{N} \frac{\partial m_i}{\partial\beta_j} \left\{ \sum_{v=1}^{\varphi_i - w_i} \frac{v-1}{(\lambda - m_i\lambda + v - 1)^2} - \sum_{v=1}^{w_i} \frac{v-1}{(m_i\lambda + v - 1)^2} \right\} \qquad \text{(A.1)}$$

$$-\frac{\partial^2 \ell}{\partial\lambda^2} = \sum_{i=1}^{N} \left\{ \sum_{v=1}^{w_i} \left(\frac{m_i}{m_i\lambda + v - 1} \right)^2 - \sum_{v=1}^{\varphi_i} \frac{1}{(\lambda + v - 1)^2} + \sum_{v=1}^{\varphi_i - w_i} \left(\frac{1 - m_i}{\lambda - m_i\lambda + v - 1} \right)^2 \right\}.$$

These quantities are evaluated at $\hat{\beta}$, $\hat{\lambda}$ for use in $\hat{\Sigma} = \{\hat{\sigma}_{jk}\}$.

A.2. Marginal likelihood derivatives for Poisson model

For the (marginal) likelihood under the Poisson model from §3.2, the gradient elements are

$$\frac{\partial \ell}{\partial\beta_j} = \lambda \sum_{i=1}^{N} \frac{\partial m_i}{\partial\beta_j} \left\{ \log\lambda + \sum_{v=1}^{y_i} \frac{1}{m_i\lambda + v - 1} \right\} \qquad \text{(j=1,...P)}$$

$$\frac{\partial \ell}{\partial\lambda} = \frac{N}{N+1}\bar{y} + \sum_{i=1}^{N} m_i \left\{ \frac{1}{1+\lambda} + \log\left\{ \frac{\lambda}{1+\lambda} \right\} + \sum_{v=1}^{y_i} \frac{1}{m_i\lambda + v - 1} \right\}.$$

Also, the negative Hessian elements involve

$$-\frac{\partial^2 \ell}{\partial\beta_j \partial\beta_k} = \lambda \sum_{i=1}^{N} \left\{ \lambda \frac{\partial m_i}{\partial\beta_j}\frac{\partial m_i}{\partial\beta_k} \sum_{v=1}^{y_i} \frac{1}{(m_i\lambda + v - 1)^2} \right.$$

$$\left. - \frac{\partial^2 m_i}{\partial\beta_j \partial\beta_k} \left[\log\left\{ \frac{\lambda}{\lambda+1} \right\} + \sum_{v=1}^{y_i} \frac{1}{m_i\lambda + v - 1} \right] \right\},$$

$$-\frac{\partial^2 \ell}{\partial\beta_j\partial\lambda} = \sum_{i=1}^{N}\frac{\partial m_i}{\partial\beta_j}\left\{\lambda\sum_{v=1}^{y_i}\frac{1}{(m_i\lambda+v-1)^2} - \sum_{v=1}^{y_i}\frac{1}{m_i\lambda+v-1} - log\left\{\frac{\lambda}{\lambda+1}\right\} - \frac{1}{\lambda+1}\right\}$$

and

$$-\frac{\partial^2 \ell}{\partial\lambda^2} = \frac{N}{(\lambda+1)^2}\bar{y} + \sum_{i=1}^{N} m_i\left\{m_i\sum_{v=1}^{y_i}\frac{1}{(m_i\lambda+v-1)^2} - \frac{1}{\lambda(\lambda+1)^2}\right\} .$$

As above, these latter quantities are evaluated at $\hat{\beta}$, $\hat{\lambda}$ for use in $\hat{\Sigma} = \{\hat{\sigma}_{jk}\}$, which is in turn necessary for variance estimation under (2.3).

A.3. Marginal likelihood derivatives for extended binomial model

For the (marginal) likelihood under the extended binomial model from §4.2, the additional gradient element for the γ-parameter is

$$\frac{\partial\ell}{\partial\gamma} = \sum_{i=1}^{N}\lambda\frac{\partial m_i}{\partial\gamma}D_1(y_i,\lambda) \quad .$$

The negative second derivatives include

$$-\frac{\partial^2\ell}{\partial\gamma^2} = \lambda\sum_{i=1}^{N}\left\{\lambda\left(\frac{\partial m_i}{\partial\gamma}\right)^2 D_2(y_i,\lambda) - \frac{\partial^2 m_i}{\partial\gamma^2}D_1(y_i,\lambda)\right\}$$

$$-\frac{\partial^2\ell}{\partial\beta_j\partial\gamma} = \lambda\sum_{i=1}^{N}\left\{\lambda\frac{\partial m_i}{\partial\beta_j}\frac{\partial m_i}{\partial\gamma}D_2(y_i,\lambda) - \frac{\partial^2 m_i}{\partial\beta_j\partial\gamma}D_1(y_i,\lambda)\right\} \tag{A.2}$$

$$-\frac{\partial^2\ell}{\partial\lambda\partial\gamma} = \sum_{i=1}^{N}\frac{\partial m_i}{\partial\gamma}\left\{\lambda\left[m_i D_2(y_i,\lambda) - \sum_{v=1}^{\varphi_i-w_i}\left(m_i\lambda+v-1\right)^{-2}\right] - D_1(y_i,\lambda)\right\}$$

where, for notation's sake, we set

$$D_1(y_i,\lambda) = \sum_{v=1}^{w_i}(m_i\lambda+v-1)^{-1} - \sum_{v=1}^{\varphi_i-w_i}(\lambda-m_i\lambda+v-1)^{-1},$$

$$D_2(y_i,\lambda) = \sum_{v=1}^{w_i}(m_i\lambda+v-1)^{-2} + \sum_{v=1}^{\varphi_i-w_i}(\lambda-m_i\lambda+v-1)^{-2},$$

and $w_i = y_i\varphi_i$. These are evaluated at the MML estimates for construction of v_i^{EB}. All of these quantities depend, of course, on the specific functional form chosen for the extended link family. For example, under Whittemore's extended family (4.2),

$$\frac{\partial m_i}{\partial\beta_j} = x_{ij}exp\{\mathbf{x}_i'\beta\}/(1+exp\{\mathbf{x}_i'\beta\}-\gamma)^2 \ ,$$

$$\frac{\partial^2 m_i}{\partial\beta_j\,\partial\beta_k} = \frac{x_{ij}\,x_{ik}\,exp\left\{x_i'\beta\right\}\left(1 - exp\left\{x_i'\beta\right\} - \gamma\right)}{\left(1 + exp\left\{x_i'\beta\right\} - \gamma\right)^3} \quad ,$$

$$\frac{\partial m_i}{\partial\gamma} = -(1 + exp\{x_i'\beta\} - \gamma)^{-2} \quad ,$$

for use in (2.3), $\partial\ell/\partial\gamma$ and (A.2), above.

References

Abramowitz, M. and Stegun, I. A. (1972), *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables* (10th ed.), New York: Wiley Interscience.

Albert, J. H. (1984), "Empirical Bayes estimation of a set of binomial probabilities," *Journal of Statistical Computation and Simulation*, 20, 129-144.

Albert, J. H. (1985), "Simultaneous estimation of Poisson means under exchangeable and independence models," *Journal of Statistical Computation and Simulation*, 23, 1-14.

Albert, J. H. (1988), "Computational methods using a Bayesian hierarchical generalized linear model," *Journal of the American Statistical Association*, 83, 1037-1044.

Albert, J. H. and Pepple, P. A. (1989), "A Bayesian approach to some overdispersion models," *Canadian Journal of Statistics*, 17, 333-344.

Berry, D. A. and Christensen, R. (1979), "Empirical Bayes estimation of a binomial parameter via mixtures of Dirichlet processes," *Annals of Statistics*, 7, 558-568.

Box, G. E. P. and Tiao, G. C. (1973), *Bayesian Inference in Statistical Analysis*, Reading, Mass.: Addison-Wesley.

Brier, S. S., Zacks, S. and Marlow, W. H. (1986), "An application of empirical Bayes techniques to the simultaneous estimation of many probabilities," *Naval Research Logistics Quarterly*, 33, 77-90.

Brown, L. D. and Farrell, R. H. (1985), "All admissible linear estimators of a multivariate Poisson mean," *Annals of Statistics*, 13, 282-294.

Casella, G. (1985), "An introduction to empirical Bayes data analysis," *American Statistician*, 39, 83-87.

Casella, G. (1988), "Empirical Bayes methods — A tutorial," Technical Report, number 88-18, Department of Statistics, Purdue University.

Casella, G. and Hwang, J. T. (1983), "Empirical Bayes confidence sets for the mean of a multivariate normal distribution," *Journal of the American Statitical Association*, 78, 688-697.

Clayton, D. and Kaldor, J. (1987), "Empirical Bayes estimates of age-standardized relative risks for use in disease mapping," *Biometrics*, 43, 671-681.

Consonni, G. and Dawid, A. P. (1985), "Invariant normal Bayesian linear models and experiment designs," in *Bayesian Statistics* (Vol. 2), ed. Bernardo, J. M., DeGroot, M. H., Lindley, D. V. and Smith, A. F. M., Amsterdam: Elsevier Science Publishers B.V. (North-Holland), pp. 629-644.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977), "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society, series B*, 39, 1-38.

Duffey, D. E. and Santner, T. J. (1989), "On the small sample properties of norm-restricted maximum likelihood estimators for logistic regression models," *Communications in Statistics — Theory and Methods*, 18, 959-980.

DuMouchel, W. M. and Harris, J. E. (1983), "Bayes methods for combining the results of cancer studies in humans and other species (with discussion)," *Journal of the American Statistical Association*, 77, 293-313.

Dunnick, J. K., Eustis, S. L., Piegorsch, W. W. and Miller, R. A. (1988), "Respiratory tract lesions in F344/N rats and B6C3F1 mice after exposure to 1,2-Epoxybutane," *Toxicology*, 50, 69-82.

Fletcher, R. (1970), "A new approach to variable metric algorithms," *The Computer Journal*, 13, 317-322.

Frome, E. L. (1983), "The analysis of rates using Poisson regression models," *Biometrics*, 39, 665-674.

Frome, E. L., Kutner, M. H. and Beauchamp, J. J. (1973), "Regression analysis of Poisson-distributed data," *Journal of the American Statistical Association*, 68, 935-940.

Hudson, N. M. (1985), "Adaptive estimators for simultaneous estimation of Poisson means," *Annals of Statistics*, 13, 246-261.

Hui, S. L. and Berger, J. O. (1983), "Empirical Bayes estimation of rates in longitudinal studies," *Journal of the American Statistical Association*, 78, 753-760.

Kass, R. E. and Steffey, D. (1989), "Approximate Bayesian inference in conditionally independent hierarchical models (parametric empirical Bayes methods)," *Journal of the American Statistical Association*, 84, 717-726.

Laird, N. (1978), "Empirical Bayes methods for two-way contingency tables," *Biometrika*, 65, 581-590.

Laird, N. M. and Louis, T. A. (1989), "Empirical Bayes confidence intervals for a series of related experiments," *Biometrics*, 47, 481-495.

Lawless, J. F. (1987), "Regression methods for Poisson process data," *Journal of the American Statistical Association*, 82, 808-815.

Lehmann, E. L. (1983), *Theory of Point Estimation*, New York: Wiley.

Leonard, T. (1972), "Bayesian methods for binomial data," *Biometrika*, 59, 581-589.

Leonard, T., Hui, J. S. J. and Tsui, K.-W. (1989), "Bayesian marginal inference," *Journal of the American Statistical Association*, 84, 1051-1058.

Lindley, D. V. and Smith, A. F. M. (1972), "Bayes estimates for the linear model (with discussion)," *Journal of the Royal Statistical Society, series B*, 34, 1-41.

Louis, T. A. (1989), "Analysis of a random sample of two year carcinogen bioassays from the NCI data base," *Fundamental and Applied Toxicology*, 12, 222-231.

Maritz, J. S. and Lwin, T. (1989), *Empirical Bayes Methods* (2nd ed.), London: Chapman and Hall.

Martz, H. and Waller, J. (1982), *Bayesian Reliability Analysis*, New York: Wiley.

McCullagh, P. and Nelder, J. A. (1983), *Generalized Linear Models*, London: Chapman and Hall.

Morgan, B. J. T. (1988), "Extended models for quantal response data," *Statistica Neerlandica*, 42, 253-272.

Morris, C. (1983), "Parametric empirical Bayes inference: Theory and practice," *Journal of the American Statistical Association*, 78, 47-65.

Nebebe, F. and Stroud, T. W. F. (1986), "Bayes and emperical Bayes shrinkage estimation of regression coefficients," *Canadian Journal of Statistics*, 14, 267-280.

O'Bryan, T. E. (1979), "Rates of convergence in a modified empirical Bayes estimation problem involving Poisson distributions," *Communications in Statistics — Theory and Methods*, 8, 167-174.

Ramsey, F. L. (1972), "A Bayesian approach to bioassay," *Biometrics*, 28, 841-858.

Robbins, H. (1955), "An empirical Bayes approach to statistics," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 1, Berkeley: University of California Press, pp. 157-164.

Rubin, D. B. (1980), "Using empirical Bayes techniques in the Law School Validation Studies," *Journal of the American Statistical Association*, 75, 801-816.

Santner, T. J. and Duffey, D. E. (1989), *The Statistical Analysis of Discrete Data*, New York: Springer-Verlag.

Singh, R. S. (1985), "Empirical Bayes estimation in a multiple linear regression model," *Annals of the Institute of Statistical Mathematics, Part A*, 37, 71-86.

Smith, A. F. M. and Spiegelhalter, D. J. (1980), "Bayes factors and choice criteria for linear models," *Journal of the Royal Statistical Society, series B*, 42, 213-220.

Strenio, J. F., Weisberg, H. I. and Bryk, A. S. (1983), "Empirical Bayes estimation of individual growth-curve parameters and their relationship to covariates," *Biometrics*, 39, 71-86.

West, M. (1985), "Generalized linear models: Scale parameters, outlier accomodation and prior distributions," in *Bayesian Statistics* (Vol. 2), ed. Bernardo, J. M., DeGroot, M. H., Lindley, D. V. and Smith, A. F. M., Amsterdam: Elsevier Science Publishers B.V. (North-Holland), pp. 531-558.

Whittemore, A. S. (1983), "Transformations to linearity in binary regression," *SIAM Journal of Applied Mathematics*, 43, 703-710.

Zellner, A. and Rossi, P. E. (1984), "Bayesian analysis of dichotomous quantal response models," *Journal of Econometrics*, 25, 365-393.