Ann. Inst. Statist. Math.
Vol. 46, No. 3, 497–507 (1994)

APPROXIMATE BAYESIAN SHRINKAGE ESTIMATION

WANG-SHU LU

Department of Statistics, College of Arts and Science, University of Missouri-Columbia, 222 Math Sciences Bldg., Columbia, MO 65211, U.S.A.

(Received March 8, 1993; revised February 8, 1994)

Abstract. A Bayesian shrinkage estimate for the mean in the generalized linear empirical Bayes model is proposed. The posterior mean under the empirical Bayes model has a shrinkage pattern. The shrinkage factor is estimated by using a Bayesian method with the regression coefficients to be fixed at the maximum extended quasi-likelihood estimates. This approach develops a Bayesian shrinkage estimate of the mean which is numerically quite tractable. The method is illustrated with a data set, and the estimate is compared with an earlier one based on an empirical Bayes method. In a special case of the homogeneous model with exchangeable priors, the performance of the Bayesian estimate is illustrated by computer simulations. The simulation result shows as improvement of the Bayesian estimate over the empirical Bayes estimate in some situations.

Key words and phrases: Bayes estimate, empirical Bayes, extended quasi-likelihood, generalized linear model, relative saving loss, shrinkage estimate.

1. Introduction

A hierarchical modeling structure in an empirical Bayes (EB) version of the generalized linear model of McCullagh and Nelder (1989) is referred to as the generalized linear EB model by Lu and Morris (1994). This model, along with its descriptive and inferential perspectives, using the mean-variance structure in distributional expression is described as follows.

Assume the data $y_1, \dots, y_k$ are observed from the natural exponential family (NEF) with a quadratic variance function (Morris (1982)). The mean-variance structure (two entries in the square brackets) is given by

$$y_i \mid \mu_i \overset{\text{indep}}{\sim} \mathrm{NEF}\left[\mu_i, \frac{V(\mu_i)}{n_i}\right], \qquad i = 1, \dots, k \tag{1.1}$$

where $V(t) = v_2 t^2 + v_1 t + v_0$ with v_2, v_1, v_0 known is a quadratic variance function of the mean which characterizes the NEF, and $n_i > -v_2$ is a known convolution parameter which need not be an integer. In a one-way layout model, y_i can stand for the sample mean with the sample size n_i in the i-th group. When all $n_i \equiv n$, the model is said to be homogeneous. This NEF with the quadratic variance function includes most common distributions such as normal, Poisson, gamma, binomial etc.

Under the NEF in (1.1), the conjugate prior of μ_i is assumed to have a generalized linear regression form with the mean-variance structure:

$$\mu_i \mid \alpha, \beta \overset{\text{indep}}{\sim} \text{Conj}[\mu_{0i}, \alpha V(\mu_{0i})], \qquad i = 1, \ldots, k \tag{1.2}$$

where $\alpha = 1/(m - v_2)$ with $m > v_2$, not necessarily an integer, being the convolution parameter for the conjugate distribution, and β is a $p \times 1$ vector of regression coefficients ($p < k$) which is related to the prior mean μ_{0i} through $g(\mu_{0i}) = x_i^T \beta$. The function $g(\cdot)$ is a known one-one link function and x_i is a $p \times 1$ vector of regressors. For notational convenience, I still write μ_{0i} as the prior mean with understanding that $\mu_{0i} = g^{-1}(x_i^T \beta)$ is a function of β. When $p = 1$ with all $x_i \equiv 1$ (no regressor), equation (1.2) is called the exchangeable priors.

Our objective is to estimate the mean μ_i by means of Bayesian EB analysis.

Under the model of (1.1) and (1.2), the posterior distribution of μ_i, given the hyperparameters (α, β), can be shown to be

$$\mu_i \mid y_i, \alpha, \beta \overset{\text{indep}}{\sim} \text{Conj}\left[\mu_i^*, \frac{V(\mu_i^*)}{n_i + \alpha^{-1}}\right], \qquad i = 1, \ldots, k \tag{1.3}$$

where $\mu_i^* = E(\mu_i \mid y_i, \alpha, \beta) = (1 - b_i) y_i + b_i \mu_{0i}$ is the posterior mean which shrinks y_i towards μ_{0i} with the shrinkage factor $b_i = m/(m+n_i) = (1+v_2\alpha)/[1+(n_i+v_2)\alpha]$, a function of α only.

When (α, β) are known, the Bayesian estimate of μ_i with respect to the squared error loss is μ_i^*, and the EB risk is $E(\mu_i^* - \mu_i)^2$, where the expectation is evaluated in terms of the joint distribution of (y_i, μ_i) with the mean-covariance structure:

$$(y_i, \mu_i) \mid \alpha, \beta \overset{\text{indep}}{\sim} [(\mu_{0i}, \mu_{0i}), \alpha V(\mu_{0i}) \Sigma_i], \qquad i = 1, \ldots, k \tag{1.4}$$

with $\Sigma_i = \begin{pmatrix} 1/(1 - b_i) & 1 \\ 1 & 1 \end{pmatrix}$. It is easy to show that $E(\mu_i^* - \mu_i)^2 = \alpha b_i V(\mu_{0i})$.

When (α, β) are unknown, an attempt to obtain an estimate for μ_i using EB and hierarchical Bayes methods is appropriate. Morris (1988) addresses the problem of evaluating the posterior mean of μ_i and determining its accuracy in a homogeneous modeling setting with the exchangeable priors. Albert (1988) contributes computational methods using a hierarchical Bayesian analysis to estimate μ_i and discusses the assessment of the goodness-of-fit of the generalized linear EB model. Lu and Morris (1994) propose an EB estimate of μ_i using the extended quasi-likelihood of Nelder and Pregibon (1987) to estimate the hyperparameters (α, β), which can be readily implemented via the GLIM software.

Among various methods of estimating (α, β), it has been noted that estimating β is more accurate than estimating α in most situations. In particular, when the true value of α is close to zero, the maximum likelihood estimate (MLE) of α could erroneously lead to a negative value if k is small, although one can usually force the estimate of α to be zero in this case. Thus a primary error in finally estimating μ_i through μ_i^* would be caused by misestimating α (hence the shrinkage factor b_i) rather than β (hence the μ_{0i}). For this reason, it would be possible to provide an improvement over the EB estimate by treating α as a random variable and estimating the shrinkage factor b_i by a Bayesian approach.

In order to concentrate on α only, the values of β are assumed to be known and are fixed at the maximum extended quasi-likelihood estimates (Lu and Morris (1994)). Since only α is unknown, the full posterior mean and posterior variance of μ_i are numerically quite tractable by means of only one-dimensional integrations. This method provides an approximate Bayesian estimate of μ_i, and its formulation is derived in Section 2. In Section 3, the method, along with comparing with the EB method, is illustrated for a traffic accident data example. For this data set, the present Bayesian estimate for the shrinkage factor corrects bias and is almost perfectly linearly related to that by Lu and Morris (1994). A special case of the homogeneous model with exchangeable priors is considered in Section 4. For the Poisson, the gamma, and the binomial of non-normal NEF distributions using certain parameter values, the present Bayesian estimate is compared with the Morris' (1988) EB estimate via computer simulation to evaluate their relative saving losses. The simulation result shows that the performance of the Bayesian estimate appears better, especially for small value of k, than the EB estimate for most cases, with an exception of the binomial model for large values of k.

2. Bayesian estimation

From equation (1.4), the marginal distribution of y_i is given by

$$y_i \mid \alpha, \beta \overset{\text{indep}}{\sim} [\mu_{0i}, \phi_i V(\mu_{0i})], \qquad i = 1, \ldots, k \tag{2.1}$$

with $\phi_i = \alpha/(1-b_i) = 1/n_i + (1 + v_2/n_i)\alpha$. The formulation of the likelihood for (α, β) based on the distribution (2.1) is generally quite complex and can be approximated by the extended quasi-likelihood of Nelder and Pregibon (1987). With the variance function $V(\cdot)$ and the dispersion parameter ϕ_i in the mean-variance structure (2.1), the extended quasi-likelihood for (α, β) is given by

$$Q(\alpha, \beta) = \prod_{i=1}^{k} \left[\phi_i^{-1/2} \exp\left(-\frac{d_i}{2\phi_i} \right) \right], \tag{2.2}$$

where $d_i \equiv d_i(\beta) = -2\int_{y_i}^{\mu_{0i}} \frac{y_i - t}{V(t)} dt$ is the deviance function of the NEF in the i-th component.

A noninformative uniform prior distribution on α at the third stage of the EB model is assumed: $\alpha \sim \text{Uniform}(0, u)$ with $u = 1/\max(0, -v_2)$. Here the value of u is chosen to ensure that no negative shrinkage amount b_i would occur when $v_2 < 0$.

Based on equation (2.2), the posterior density of α, given β, is

$$
\begin{aligned}
(2.3) \qquad p(\alpha \mid y) \propto Q(\alpha,\beta) &= \prod_{j=1}^{k}\left[\phi_j^{-1/2}\exp\left(-\frac{d_j}{2\phi_j}\right)\right] \\
&\propto \exp\left\{-\frac{1}{2}\sum_{j=1}^{k}\left[\left(b_j-\frac{v_2}{n_j+v_2}\right)(n_j+v_2)d_j\right]\right\} \\
&\quad \cdot\prod_{j=1}^{k}\left(b_j-\frac{v_2}{n_j+v_2}\right)^{1/2} \quad \text{for} \quad 0<\alpha<u,
\end{aligned}
$$

where $y=(y_1,\ldots,y_k)$ and b_j is the shrinkage factor defined in (1.3).

In terms of the density (2.3), the expectation of any function of α can be easily evaluated numerically since only one-dimensional integrations are involved.

As a result, the posterior mean of b_i is given by

$$
(2.4) \qquad \tilde{b}_i \equiv E(b_i \mid y) = \frac{v_2}{n_i+v_2} + w_i
$$

with

$$
w_i = \frac{\int_0^u \exp\{-\frac{1}{2}\sum_{j=1}^{k}[(b_j-\frac{v_2}{n_j+v_2})(n_j+v_2)d_j]\}\prod_{j=1}^{k}(b_j-\frac{v_2}{n_j+v_2})^{1/2+\delta_{ij}}d\alpha}{\int_0^u \exp\{-\frac{1}{2}\sum_{j=1}^{k}[(b_j-\frac{v_2}{n_j+v_2})(n_j+v_2)d_j]\}\prod_{j=1}^{k}(b_j-\frac{v_2}{n_j+v_2})^{1/2}d\alpha},
$$

where $\delta_{ij}=1$ or 0 as $i=j$ or $i\neq j$.

Note that the value of $\tilde{b}_i$ is always between $v_2/(n_i+v_2)$ and 1 if $v_2\geq 0$, and between 0 and 1 if $v_2<0$.

Since μ_i^* is a linear function of b_i, the posterior mean of μ_i (given β) then is

$$
(2.5) \qquad \tilde{\mu}_i = (1-\tilde{b}_i)y_i + \tilde{b}_i\mu_{0i}.
$$

Moreover, the posterior variance of μ_i can be formulated from equation (1.3) by integrating some functions of α based on the density (2.3), but I will omit performing this here.

In practice, the values of β must be pre-estimated in order to have the final formulation of the estimation in (2.5). Lu and Morris (1994) propose a feasible way to find estimates of β. Noting that the quasi-likelihood (2.2) follows a double exponential family (Efron (1986)) for the marginal distribution of y_i's, we can find the values of (α,β) to maximize $Q(\alpha,\beta)$ in (2.2) via a marco facility of the GLIM package in two NEF distributions. This process consists of the following two steps:

(a) Given α (so ϕ_i), make GLIM regression method applicable for estimating β with the link $g(\mu_{0i})=x_i^T\beta$ and the prior weights ϕ_i's by treating the data y_i's as a random sample from the NEF in (1.1), but with n_i in (1.1) replaced by $1/\phi_i$;

(b) Given β (so μ_{0i}), make GLIM regression method applicable for estimating α with the identity link $\phi_i=1/n_i+(1+v_2/n_i)\alpha$ by treating the data d_i's (because

μ_{0i} fixed) as a random sample from gamma distribution (NEF) with the scale value of 2.

Cycling between holding α fixed in (a) and holding β fixed in (b) leads to estimates of (α, β) which maximize $Q(\alpha, \beta)$ in (2.2). In this implementation, the final estimate of β is the MLE in the NEF with the link $g(\mu_{0i}) = x_i^T \beta$ and the prior weights ϕ_i's fixed at an estimated value of α. For more details about this procedure, see Lu and Morris (1994). In all computations regarding α as a random variable here, I will treat β to be fixed at such an estimated value. It is quite reasonable to do so, since the MLE of β in the NEF is not sensitive over a certain range of values of α. Indeed, estimating β is independent of α if the model is homogeneous. Thus when all n_i do not substantially differ from one to another, little would be affected by the value of α in estimating β by the above procedure.

Using a Bayesian method to estimate the shrinkage factor b_i (a function of α only) in such a way yields an approximate Bayesian estimate for μ_i. This develops a method that is based on the EB method in Lu and Morris (1994), and has actually the Bayesian EB manner, but I will refer to it as Bayesian for simplicity.

An entire Bayesian method could be further developed if the regression coefficients β are treated as random in the modeling setting, but finding the full Bayesian estimate of μ_i will involve a large number of integrals. In practice, the full Bayesian approach under our model has not been widely used due to difficulties in numerically evaluating substantial multi-integrations. The goal in this paper is to provide a methodology that is a hybrid of the Bayesian and the EB methods using the exist GLIM package and readily available software of numerical one-dimensional integration, that leads to an approximation of the full Bayesian estimate. Yet the accuracy of such a method or the variance attributed to the final estimate cannot be obtained until true values of the coefficients β or exactly full Bayesian computations are available. On the other hand, by departing from a full Bayesian analysis, it should be emphasized that our modeling setting has the EB sense in that the performance of any estimate is judged via the EB risk that is evaluated based on the joint distribution of (y_i, μ_i) in (1.4) with the hyperparameters (α, β) being fixed. Thus numerical comparisons of the present Bayesian method with other methods, for example with the EB method in this paper, can still be evaluated.

Davidian and Carroll (1988) point out that the maximum extended quasi-likelihood estimate of α is inconsistent in general. Therefore, using such a Bayesian method by treating α as a random variable in the estimation suggests a motivation for improving the EB method.

Working on the extended quasi-likelihood from (2.1) does not require the assumption of the conjugate prior distribution on μ_i. For any prior of μ_i with mean-variance specified in (1.2), the posterior mean μ_i^* in (1.3) is still optimal in the linear Bayesian sense that μ_i^* minimizes the squared error loss within the class of all linear functions of y_i (Robbins (1983)).

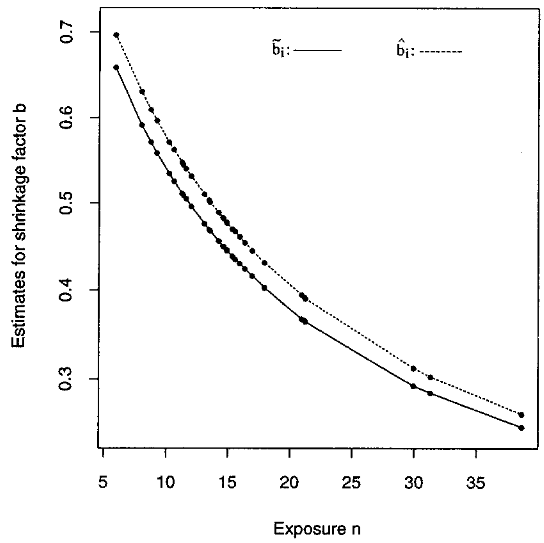

Fig. 1. Plots of the Bayesian estimate $\tilde{b}_i$ in equation (2.4) and the maximum quasi-likelihood estimate $\hat{b}_i$ in Lu and Morris (1994) for the shrinkage factor b_i vs. the exposure n_i for the traffic data example.

3. A data example

The Bayesian method will now be illustrated on a traffic accident data set previously analyzed by Lu and Morris (1994). The data consist of the observed traffic accident rate y_i with the exposure n_i for $k = 33$ signalized intersections in Tucson, Arizona for the two years (July 1981–June 1983). The observed rate y_i was assumed in terms of a Poisson distribution setting in the NEF of (1.1), with the exposure n_i as the "sample size". Also, the exchangeable gamma priors of (1.2) were assumed. The EB estimate of μ_i, the true accident rate in the i-th intersection, was found as $\hat{\mu}_i = (1 - \hat{b}_i)y_i + 0.984\hat{b}_i$, where $\hat{b}_i = 1/(1 + 0.0734n_i)$ is the estimate for the shrinkage factor and 0.984 is the estimate for the common prior mean μ_0 (see Lu and Morris (1994)).

The Bayesian estimate $\tilde{b}_i$ in terms of equation (2.4) is now computed, where the integrations are evaluated via the subroutine QDAGI in IMSL, and the corresponding Bayesian estimate $\tilde{\mu}_i$ in equation (2.5) is then obtained with $\mu_{0i} \equiv \mu_0 = 0.984$.

Figure 1 gives plots of the shrinkage estimates $\tilde{b}_i$ and $\hat{b}_i$ against the exposure n_i. It is seen that both shrinkage estimates decrease in the exposure but $\tilde{b}_i$ is slightly smaller than $\hat{b}_i$. The correlation between $\tilde{b}_i$ and $\hat{b}_i$ is 1.000 which indicates an almost perfectly linear relationship between them. The regression line of $\tilde{b}_i$ on $\hat{b}_i$ with no intercept is found as $\tilde{b}_i = 0.935\hat{b}_i$. Lu and Morris (1994) note the bias of the estimate $\hat{b}_i$ and suggest a modification by multiplying $\hat{b}_i$ by a factor $C = (k - 3)/k$. For this data set, $C = 30/33 \approx 0.909$ (comparing with the regression slope 0.935). Although the modification corrects for bias somehow, it is inappropriate when k is small because the modified shrinkage estimate is bounded above by C. Figure 2 gives a scale plot of the present Bayesian estimate $\tilde{\mu}_i$ against

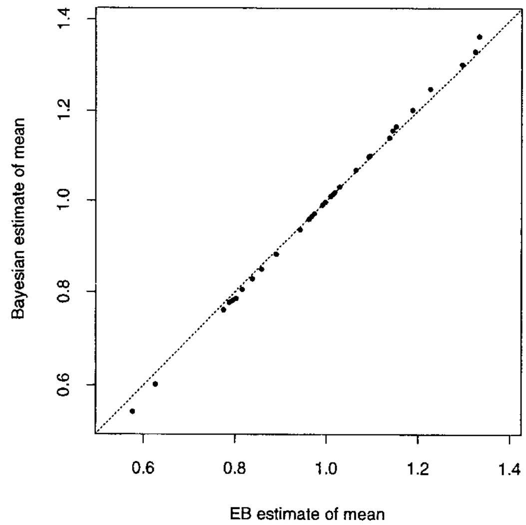

Fig. 2. Scale plot of the Bayesian estimate $\tilde{\mu}_i$ in equation (2.5) vs. the EB estimate $\hat{\mu}_i$ in Lu and Morris (1994) for the traffic data example.

the EB estimate $\hat{\mu}_i$ by Lu and Morris. Since $\hat{\mu}_i$ shrinks y_i towards μ_0 more than $\tilde{\mu}_i$ does, the slope of the plotted points tends to be slightly greater than 1. These two estimates of μ_i are almost identical for those intersections in which y_i is close to the prior mean μ_0. Moreover, the average of $\tilde{\mu}_i$ is almost identical to $\mu_0 = 0.984$ that is precisely the average of $\hat{\mu}_i$ (Lu and Morris (1994)), and is close to $\bar{y} = 0.981$.

4. Homogeneous model with exchangeable priors

A special case of the homogeneous model with exchangeable priors in which $n_i \equiv n$ and $\mu_{0i} \equiv \mu_0$ for all i is considered in this section. Morris (1988) proposes an EB estimate of μ_i by

$$\hat{\mu}_i = (1 - \hat{b})y_i + \hat{b}\bar{y}, \tag{4.1}$$

where $\hat{b} = \frac{v_2}{n+v_2}\frac{k-1}{k} + \frac{n}{n+v_2}\min(\frac{(k-3)V(\bar{y})}{nS}, \frac{k-3}{k-1})$ with $\bar{y} = \sum y_i/k$ and $S = \sum(y_i - \bar{y})^2$, which is a modified version to the method of moments for estimating b.

The Bayesian estimate for the shrinkage factor b (free of i) by (2.4) becomes

$$\tilde{b} = \frac{v_2}{n + v_2} + w, \tag{4.2}$$

where, upon changing of integral variable by $t = b - v_2/(n + v_2)$,

$$w = \frac{\int_q^r t^{k/2-1} \exp\left[-\left(\frac{n+v_2}{2}\right) Dt\right] dt}{\int_q^r t^{k/2-2} \exp\left[-\left(\frac{n+v_2}{2}\right) Dt\right] dt}$$

with $q = \max(0, -v_2/(n+v_2))$, $r = n/(n+v_2)$, and $D = \sum d_i$, the total deviance of the NEF.

The value of w in (4.2) can be expressed in terms of chi-square distributions as

$$w = \left[\frac{G_k(nD) - G_k(cD)}{G_{k-2}(nD) - G_{k-2}(cD)}\right]\left[\frac{k-2}{(n+v_2)D}\right],$$

where $c = \max(0, -v_2)$ and $G_\gamma(\cdot)$ denotes the chi-square distribution function with degrees of freedom γ.

Under the homogeneous model, the value of μ_{0i} in (2.5) can be obtained independently of α. In particular, when $\mu_{0i} \equiv \mu_0$ (exchangeable priors), the maximum quasi-likelihood estimate of μ_0, based on equation (2.2), is $\bar{y}$. Using $\bar{y}$ as an estimated value of μ_0 can follow a Bayesian framework, as noted by Morris (1988). In fact, if a noninformative uniform prior on β with the entire space as its support is assumed, the posterior density of β based on equation (2.2) for any given value of α is

$$p(\beta \mid y) = C \exp\left[\frac{1-b}{\alpha}\sum_{i=1}^{k}\int_{y_i}^{\mu_0}\frac{y_i - t}{V(t)}dt\right]$$

for some constant C. If we furthermore use a canonical link function $g(\cdot)$ of the NEF (recall that $g(\mu_0) = \beta$), we have that $g'(\mu_0) = V^{-1}(\mu_0)$ and hence $dp(\beta \mid y) = \frac{k(1-b)}{\alpha}(\bar{y} - \mu_0)p(\beta \mid y)d\beta$. Therefore, $\int(\bar{y} - \mu_0)p(\beta \mid y)d\beta = 0$ or $E(\mu_0 \mid y) = \bar{y}$. That is, $\bar{y}$ is the posterior mean of μ_0.

Thus under the homogeneous model with exchangeable priors, the Bayesian estimate of μ_i in (2.5) becomes

$$\tilde{\mu}_i = (1 - \tilde{b})y_i + \tilde{b}\bar{y} \tag{4.3}$$

with $\tilde{b}$ given by (4.2).

Lu and Morris (1994) suggest an EB estimate similar to (4.3) with $\tilde{b}$ being replaced by $\min(\frac{v_2}{n+v_2} + \frac{k-3}{(n+v_2)D}, 1)$. The performance of the Bayesian estimate (4.3) for the normal model is explored in Lu (1993).

Here computer simulations are applied to some non-normal NEF to compare the relative saving loss (RSL) for the two estimates (4.1) and (4.3). Analogous to that defined by Efron and Morris (1973) in the normal case, the RSL of the estimate $\hat{\mu}_i$ (similar for the estimate $\tilde{\mu}_i$) is defined by

$$\mathrm{RSL}(\hat{\mu}_i) = \frac{(\text{EB risk of } \hat{\mu}_i) - (\text{EB risk of } \mu_i^*)}{(\text{EB risk of } y_i) - (\text{EB risk of } \mu_i^*)}, \tag{4.4}$$

where EB risk of $\hat{\mu}_i$ is $E(\hat{\mu}_i - \mu_i)^2$, EB risk of μ_i^* is $E(\mu_i^* - \mu_i)^2 = \alpha b V(\mu_0)$ by (1.4), and EB risk of y_i is $E(y_i - \mu_i)^2 = \alpha b V(\mu_0)/(1-b)$ by simple calculations. According to definition (4.4), $\mathrm{RSL}(\hat{\mu}_i)$ measures the proportion of available improvement using $\hat{\mu}_i$ over the frequentist estimate y_i.

The computer simulations are conducted for the Poisson ($V(t) = t$), the gamma ($V(t) = t^2$), and the binomial ($V(t) = t(1-t)$) distributions of the NEF (the corresponding conjugate priors are the gamma, the inverse gamma and the beta distributions, respectively), with fixed values $n = 10$ and $\mu_0 = 0.5$. Under the exchangeable priors setting, the value μ_0 is the common prior mean. The corresponding regression coefficient β (intercept) can be obtained if a link function is specified. For example, when the canonical link function of the NEF is chosen, we have that for $\mu_0 = 0.5$, the corresponding value of $\beta = \ln(\mu_0) = -0.693$ in the Poisson case, $\beta = -1/\mu_0 = -2$ in the gamma case, and $\beta = \ln[\mu_0/(1-\mu_0)] = 0$ in the binomial case. The values of α are chosen so that the corresponding shrinkage factors are from $b = 0.05$ to $b = 0.95$ with an increment of 0.05 ($b = 0.05$ is excluded in the gamma case because in this case b must be greater than $v_2/(n+v_2) = 1/(10+1) = 0.09$). For each parameter setting in the three models, the computation is performed for $k = 5, 10, 20, 40$ with simulation sizes $M = 2000, 1000, 500, 250$, respectively. Such a selection of the simulation sizes gives total $k \times M = 10{,}000$ samples in computing the RSL for each case because equation (4.4) is independent of $i = 1, \ldots, k$ in the underlying model. The simulated results of $\mathrm{RSL}(\tilde{\mu}_i)$ and $\mathrm{RSL}(\hat{\mu}_i)$ are displayed in Figs. 3(a), (b), and (c).

From Fig. 3 we can see that both $\tilde{\mu}_i$ and $\hat{\mu}_i$ improve y_i (except for $k = 5$ and small values of b in the binomial model for the EB estimate $\hat{\mu}_i$), and the improvement is more significant as k increases. The case where the EB estimate $\hat{\mu}_i$ provides a poor performance for $k = 5$ in the binomial model should be noted. This can be justified that when $v_2 = -1$ (binomial NEF) and k is small, the shrinkage estimate $\hat{b}$ in the EB estimate (4.1) tends to be negative if the true value of b is small (hence S is typically large). In Fig. 3(c) for the binomial model, the simulations yield, for $k = 5$, $\mathrm{RSL}(\hat{\mu}_i) = 1.87, 1.10, 0.91$ for $b = 0.05, 0.10, 0.15$, respectively, which are not plotted there.

In comparing the two estimates, some facts are observed. For the Poisson and the gamma models, the performance of $\tilde{\mu}_i$ is better than that of $\hat{\mu}_i$ except for extremely small or large value of b. For the binomial model, $\tilde{\mu}_i$ is better than $\hat{\mu}_i$ only when $k \leq 10$ and b is not large, whereas $\hat{\mu}_i$ becomes better than $\tilde{\mu}_i$ when k increases. For all three models, the Bayesian estimate $\tilde{\mu}_i$ provides a much better performance than the EB estimate $\hat{\mu}_i$ when k is small because $\hat{\mu}_i$ is based on the method of moments for estimating the shrinkage factor b which requires large value of k for its efficiency. When k increases, $\mathrm{RSL}(\tilde{\mu}_i)$ and $\mathrm{RSL}(\hat{\mu}_i)$ both tend to decrease and are close to each other. This is expected since the two estimates $\tilde{\mu}_i$ and $\hat{\mu}_i$ will approach the true Bayesian estimate μ_i^* ($\mathrm{RSL}(\mu_i^*) = 0$) as $k \to \infty$.

From the simulation results, we may conclude that the present Bayesian method provides an improvement over the EB method in estimating μ_i for some models, especially when k is small. The Morris' EB estimate $\hat{\mu}_i$ has a simple formulation and its performance is quite satisfactory when k is large, but it is only applicable to the homogeneous model with exchangeable priors. The Bayesian method proposed in this paper is more applicable without the homogeneous and exchangeable restrictions.

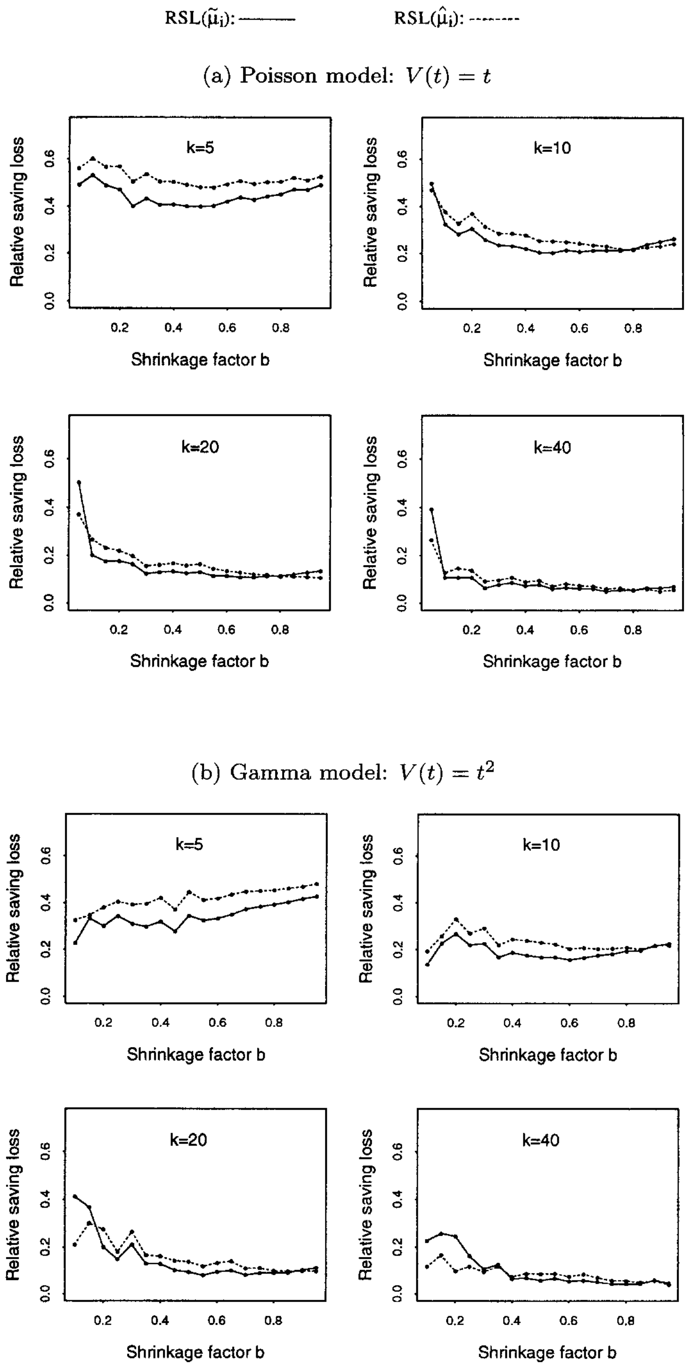

Fig. 3. Simulation results of the relative saving loss; plots of RSL($\tilde{\mu}_i$) for the Bayesian estimate (4.3) and RSL($\hat{\mu}_i$) for the EB estimate (4.1) vs. the shrinkage factor b with $n = 10$, $\mu_0 = 0.5$ for $k = 5, 10, 20, 40$, based on 10,000 samples for the homogeneous model with exchangeable priors of some non-normal NEF.

(c) Binomial model: $V(t) = t(1-t)$

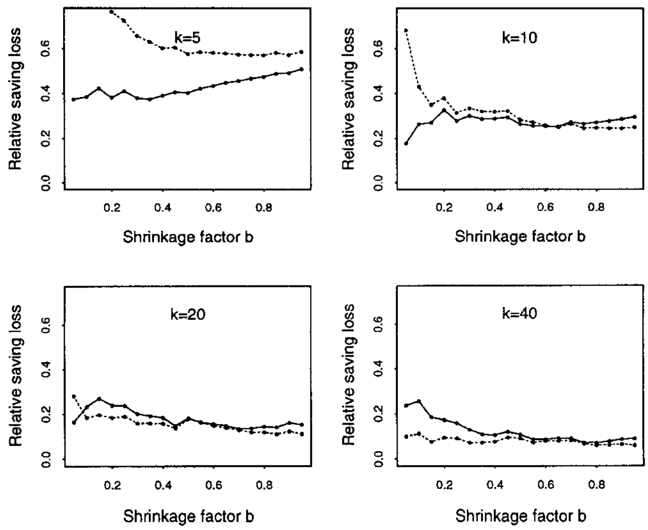

Fig. 3. (continued).

REFERENCES

Albert, J. H. (1988). Computational methods using a Bayesian hierarchical generalized linear model, *J. Amer. Statist. Assoc.*, **83**, 1037–1044.

Davidian, M. and Carroll, R. J. (1988). A note on extended quasi-likelihood, *J. Roy. Statist. Soc. Ser. B*, **50**, 74–82.

Efron, B. (1986). Double exponential families and their use in generalized linear regression, *J. Amer. Statist. Assoc.*, **81**, 709–721.

Efron, B. and Morris, C. N. (1973). Stein's estimation rule and its competitors—an empirical Bayes approach, *J. Amer. Statist. Assoc.*, **68**, 117–130.

Lu, W. (1993). Estimations in the normal regression empirical Bayes model, *Comm. Statist. A—Theory Methods*, **22**, 1773–1794.

Lu, W. and Morris, C. N. (1994). Estimation in the generalized linear empirical Bayes model using the extended quasi-likelihood, *Comm. Statist. A—Theory Methods*, **23**, 661–687.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, 2nd ed., Chapman & Hall, London.

Morris, C. N. (1982). Natural exponential families with quadratic variance functions, *Ann. Statist.*, **10**, 65–80.

Morris, C. N. (1988). Determining the accuracy of Bayesian empirical Bayes estimators in the familiar exponential families, *Statistical Decision Theory and Related Topics*, IV, Vol. 1, Proceedings of the 4th Purdue Symposium on Statistical Decision Theory and Related Topics (eds. S. S. Gupta and J. O. Berger), Springer, New York.

Nelder, J. A. and Pregibon, D. (1987). An extended quasi-likelihood function, *Biometrika*, **74**, 221–232.

Robbins, H. (1983). Some thoughts on empirical Bayes estimation, *Ann. Statist.*, **11**, 713–723.