JIRSS (2005)
Vol. 4, No. 1, pp 1-19

Empirical Bayes Estimators with Uncertainty Measures for NEF-QVF Populations

Michele Crescenzi[1], **Malay Ghosh**[2], **Tapabrata Maiti**[3]

[1]Universita La Sapienza, Roma, Italy.
[2]University of Florida.
[3]Iowa State University.

Abstract. The paper proposes empirical Bayes (EB) estimators for simultaneous estimation of means in the natural exponential family (NEF) with quadratic variance functions (QVF) models. Morris (1982, 1983a) characterized the NEF-QVF distributions which include among others the binomial, Poisson and normal distributions. In addition to the EB estimators, we provide approximations to the MSE's of these estimators. Our approach generalizes the findings of Prasad and Rao (1990) for the random effects model where only area specific direct estimators and covariates are available. The EB estimators are derived using the theory of optimal estimating functions as proposed by Godambe and Thompson (1989). This is in contrast to the approach of Morris (1988) who found some approximate EB estimators for this problem. Also, unlike Morris (1988), we allow unequal number of observations in different clusters in the derivation of the EB estimators. In finding approximations to the MSE's, we apply a bias-correction technique as proposed in Cox and Snell (1968). We

Key words and phrases: Estimating function, Mean squared error, toxoplasmosis data.

illustrate our methodology by reanalyzing the toxoplasmosis data of Efron (1978, 1986).

1 Introduction

Empirical Bayes (EB) methods have played a major role in the theory and practice of statistics for nearly two decades. These methods are most advantageous in the context of simultaneous estimation and prediction. In many instances, the direct estimators of the parameters in the different strata have low precision, primarily due to smallness of the sample sizes in the strata. EB estimators, on the other hand, gain in precision by "borrowing strength" across all the strata. The EB methods have been implemented very successfuly in many diverse areas, such as surveys, economics, insurance, medicine and others.

There exists now a huge EB literature built on normal models, popularized especially by Efron and Morris in a series of articles, Morris (1983b) is a good source of reference for many of these key papers. In contrast, the corresponding literature for the analysis of discrete data is relatively sparse. Beginning with Dempster and Tomberlin (1980), there are several attempts towards EB analysis of binomial data (see for example MacGibbon and Tomberlin, 1989; Farrell, MacGibbon and Tomberlin, 1997; Jiang and Lahiri, 2001), but a general systematic approach based on exponential family models which handles both discrete and continuous data is mostly lacking.

Sarkar and Ghosh (1998) initiated such an EB analysis for the natural exponential family (NEF) with quadratic variance functions (QVF) models. Morris (1982,1983a) characterized the NEF-QVF distributions. Its members are the binomial, Poisson, normal with known variance, gamma, negative binomial and the generalized hyperbolic secant distributions. The Sarkar-Ghosh procedure was partly ad hoc, especially in the context of estimating the dispersion parameters (to be explained in Section 2). More importantly, they did not provide any *measures of uncertainty* associated with their estimates.

The primary objective of this paper is to develop EB estimation procedures along with approximate measures of uncertainty for NEF-QVF models based on the theory of "optimal estimating functions" as proposed by Godambe and Thompson (1989). The stimulus for this work came from our association with the United States Bureau of Census. The Bureau has the Federal mandate to produce estimates of the proportion of poor school-age-children (chilren in the age-group 5-17) in alternate years for the different states, counties and school-districts of the United States. This forms a part of its Small Area Income and Poverty Estimation (SAIPE) project. The direct estimates, though reliable at the national level, are not so for lower levels of geography, such as states, counties and school-districts due to small sample sizes. EB methods are particularly valuable here because of their inherent ability to borrow strength.

The present article, however, is primarily methodological, where we formulate the EB procedure for NEF-QVF models and study their properties. The procedure seems to be quite versatile, and can be adapted in many different circumstances including small area estimation problems. The EB estimation procedure is developed in Section 2. In this section, we highlight also the difference between our approach and the one proposed in Morris (1988) and Sarkar and Ghosh (1998). Morris (1988) initiated EB estimation for NEF-QVF models. However, there are three main counts where we differ from Morris. First, our EB estimators are exact rather than approximate as in Morris. Second, unlike Morris, we can allow unequal number of observations in the different strata. Finally, while Morris provided approximations for the posterior means and variances, we provide, instead, in Section 3, approximations to the mean squared errors of the proposed EB estimators . Our results are intended to generalize those of Prasad and Rao (1990) for the random effects (often referred to as the Fay-Herriot) model. Other than Prasad and Rao (1990), related work in this area based primarily on the normal model is that of Lahiri and Rao (1996) and Datta and Lahiri (2001), although the former requires normality assumption only of the errors and not of the random effects. Section 4 contains an application of the proposed methodology for estimating the proportions of subjects for the disease toxoplasmosis in 34 cities of El Salvador. The main objective of this example is to reemphasize that the naive (plug in) estimator of the MSE associated with an EB estimator is typically an underestimate, and needs correction due to uncertainty in estimating the prior parameters. Indeed, a comparison of our corrected MSE's with the naive MSE's reveals clearly this fact. On the other hand, the corrected MSE's are still smaller than the variances of the direct estimates, namely, the sample proportions. In this way, we reap the benefits of borrowing strength. The present work also extends that of Sarkar and Ghosh (1998) who provided only the naive MSE's along with the EB estimators. Section 5 contains some concluding remarks. Some of the technical derivations are deferred to the Appendix.

2 EB Estimators

Let y_j denote the direct estimator of the j-th stratum mean ($j = 1, \ldots, k$). We assume the y_j to be independent with pdf's

$$f(y_j|\theta_j) = exp\{\xi_j[y_j\theta_j - \psi(\theta_j)]\}c(y_j, \xi_j), \tag{2.1}$$

where $\xi_j(> 0)$ are known constants. This is the one-parameter exponential family model. From McCullagh and Nelder (1989, p.28) we have

$$\begin{aligned} E(y_j|\theta_j) &= \psi'(\theta_j) = \mu_j \text{ (say)}; \\ Var(y_j|\theta_j) &= \psi''(\theta_j)/\xi_j = V(\mu_j) \text{ (say)}. \end{aligned} \tag{2.2}$$

Moreover, since $Var(y_j|\theta_j) > 0$, μ_j is strictly increasing in θ_j. For the NEF-QVF family of distributions, $V(\mu_j) = v_0 + v_1\mu_j + v_2\mu_j^2$, where v_0, v_1 and v_2 are not simultaneously zeroes. In particular, for the binomial model, $v_0 = 0$, $v_1 = 1$ and $v_2 = -1$. For the Poisson model, $v_0 = v_2 = 0$ and $v_1 = 1$. For the normal model with known variance, $v_0 = 1$ and $v_1 = v_2 = 0$.

For θ_j we consider the conjugate prior

$$\pi(\theta_j) = exp\{\lambda[m_j\theta_j - \psi(\theta_j)]\}K(\lambda, m_j). \tag{2.3}$$

The prior mean and variance of μ_j defined in (2.2) are then given by (cf. Morris, 1983a) $E(\mu_j) = m_j$, $Var(\mu_j) = V(m_j)/(\lambda - v_2)$ provided $\lambda > v_2$. This requires $\lambda > 0$ for the Poisson and normal models (since $v_2 = 0$) and $\lambda > -1$ for the binomial model (since $v_2 = -1$).

It is easy to check from (2.1) and (2.3) that the posterior $\pi(\theta_j|y_j)$ also belongs to the NEF-QVF family with

$$\pi(\theta_j|y_j) \propto exp[(\xi_j y_j + \lambda m_j)\theta_j - (\xi_j + \lambda)\psi(\theta_j)], \tag{2.4}$$

which leads to

$$E(\mu_j|y_j) = (1 - B_j)y_j + B_j m_j, \; B_j = \lambda/(\lambda + \xi_j); \tag{2.5}$$

$$Var(\mu_j|y_j) = \frac{V[E(\mu_j|y_j)]}{\xi_j + \lambda - v_2}. \tag{2.6}$$

Also from (2.1) and (2.3), it follows that

$$\begin{aligned} E(y_j) &= EE(y_j|\mu_j) = E(\mu_j) = m_j; \\ Var(y_j) &= Var[E(y_j|\mu_j)] + E[Var(y_j|\mu_j)] \\ &= \frac{V(m_j)(\lambda+\xi_j)}{\xi_j(\lambda - v_2)}. \end{aligned} \tag{2.7}$$

Since $Var(y_j)$ is monotonically decreasing in λ, λ may be referred to the precision parameter. Also, the prior means m_j are modeled as $g(m_j) = \mathbf{x}_j^T\mathbf{b}$, where g is a strictly increasing function. (See for example, Albert (1988)). Here, $\mathbf{x}_j$ are the design vectors, and $\mathbf{b}$ is the regression parameter.

The usual Fay-Herriot (Fay and Herriot, 1979) normal random effects model is given by $y_j|\theta_j \overset{ind}{\sim} N(\theta_j, \sigma_j^2)$ and $\theta_j \overset{ind}{\sim} N(\mathbf{x}_j^T\mathbf{b}, A)$. This is a special case of (2.1) and (2.3) where g is the identity function, $v_0 = 1, v_1 = v_2 = 0, \xi_j = \sigma_j^{-2}$ and $\lambda = A^{-1}$. Also, for the binomial and Poisson cases, (2.1) and (2.3) together lead respectively to the beta-binomial and gamma-Poisson marginal models for the y_j's, models that are typically used to account for overdispersion.

Although (2.1) and (2.3) can be viewed together as a generalized mixed linear model, the present formulation is not quite the same as a standard GLMM. For instance, in the latter formulation with a canonical link, the θ_j's are modeled as $\theta_j = \mathbf{x}_j^T\mathbf{b} + u_j$ where the u_j's are iid $N(0, \sigma_u^2)$ (cf. Breslow and Clayton, 1993). Jiang and Lahiri (2001) have adopted such a formulation in the binomial case. In contrast, in our canonical formulation, $g = (\psi')^{-1}$ so that $E(\mu_j) = m_j = \psi'(\mathbf{x}_j^T\mathbf{b})$. Thus, in the binary case with success probabilities p_j, $\theta_j = logit(p_j)$ so that in the usual GLMM formulation, $E[logit(p_j)] = \mathbf{x}_j^T\mathbf{b}$, while in our formulation, $logit[E(p_j)] = \mathbf{x}_j^T\mathbf{b}$. For the Poisson($\lambda_j$) case, $\theta_j = log(\lambda_j)$ and the usual GLMM formulation yields, $E[log(\lambda_j)] = \mathbf{x}_j^T\mathbf{b}$ in contrast to $log[E(\lambda_j)] = \mathbf{x}_j^T\mathbf{b}$ which comes out of our formulation.

While there is no particular theoretical reason to prefer one formulation over the other, the present formulation has the pragmatic advantage in nonnormal cases of obtaining the Bayes estimators of the θ_j's in closed form rather than as integrals arising out of the other method. This will be particularly useful in obtaining the EB estimators, and then finding closed form approximations to their MSE's for very large datasets. We can provide a readily implementable software, which takes smaller amount of time to find all the EB estimates and the related root mean squared errors as opposed to other approaches. This is particularly valuable when the number of strata is in the order of thousands, for example, simultaneous estimation of poverty rates for all the counties of the United States.

For simplicity in the remainder of the paper, we will take $g \equiv (\psi')^{-1}$. In addition, we take $\xi_j = n_j$ which reflects the tacit assumption that y_j is the average of n_j iid random variables each having a distribution of the form (2.1) with $\xi_j = 1$.

We now develop the EB estimators of the θ_j's. This requires estimation of $\mathbf{b}$ and λ from the marginal distributions of the y_j's. We use the theory of optimal estimating functions (Godambe and Thompson, 1989) for *simultaneous* estimation of $\mathbf{b}$ and λ. Noting that $E(y_j) = m_j$ and $Var(y_j) = E(y_j - m_j)^2 = \phi_j V(m_j)$ where $\phi_j = (\lambda + n_j)/[n_j(\lambda - v_2)]$, following Godambe and Thompson (1989), we begin with the elementary estimating functions $\mathbf{g}_j = (g_{1j}, g_{2j})^T$, where $g_{1j} = y_j - m_j$ and $g_{2j} = (y_j - m_j)^2 - \phi_j V(m_j)$, $j = 1, \ldots, k$.

Let

$$\begin{aligned}\mathbf{D}_j &= \begin{bmatrix} -E(\frac{\partial g_{1j}}{\partial \mathbf{b}}) & -E(\frac{\partial g_{2j}}{\partial \mathbf{b}}) \\ -E(\frac{\partial g_{1j}}{\partial \lambda}) & -E(\frac{\partial g_{2j}}{\partial \lambda}) \end{bmatrix} \\ &= \begin{bmatrix} V(m_j)\mathbf{x}_j & V(m_j)V'(m_j)\phi_j \mathbf{x}_j \\ \mathbf{0} & V(m_j)\frac{(n_j+v_2)}{n_j(\lambda - v_2)^2} \end{bmatrix}. \end{aligned} \tag{2.8}$$

Next, writing $\mu_{rj} = E(y_j - m_j)^r$ $(r = 1, 2, \ldots)$, let

$$\mathbf{V}_j = \begin{bmatrix} \mu_{2j} & \mu_{3j} \\ \mu_{3j} & \mu_{4j} - \mu_{2j}^2 \end{bmatrix} \tag{2.9}$$

(Godambe and Thompson (1989) have expressed $\mathbf{V}_j$ in terms of the cumulants rather than the central moments of y_j). Let $\boldsymbol{\eta} = (\mathbf{b}, \lambda)^T$ and

$$\mathbf{S}_k(\boldsymbol{\eta}) = \sum_{j=1}^{k} \mathbf{D}_j^T \mathbf{V}_j^{-1} \mathbf{g}_j. \tag{2.10}$$

Following Godambe and Thompson (1989), the optimal estimating equations are given by $\mathbf{S}_k(\boldsymbol{\eta}) = \mathbf{0}$ and its solution, say $\hat{\boldsymbol{\eta}}$ is used to estimate $\boldsymbol{\eta}$. Note that

$$\mathbf{V}_j^{-1} = \Delta_j^{-1} \begin{bmatrix} \mu_{4j} - \mu_{2j}^2 & -\mu_{3j} \\ -\mu_{3j} & \mu_{2j} \end{bmatrix},$$

where $\Delta_j = |\mathbf{V}_j| = \mu_{4j}\mu_{2j} - \mu_{3j}^2 - \mu_{2j}^3$. Thus, the estimating equations reduce to

$$\sum_{j=1}^{k} \Delta_j^{-1}[\{\mu_{4j} - \mu_{2j}^2 - \mu_{3j}\phi_j V'(m_j)\}g_{1j} + \{\mu_{2j}\phi_j V'(m_j) - \mu_{3j}\}g_{2j}]V(m_j)\mathbf{x}_j = \mathbf{0} \qquad (2.11)$$

and

$$\sum_{j=1}^{k} \Delta_j^{-1}[g_{2j}\mu_{2j}V(m_j) - \mu_{3j}g_{1j}]V(m_j)\frac{(n_j + v_2)}{n_j(\lambda - v_2)^2} = 0. \qquad (2.12)$$

Solving (2.11) and (2.12) simultaneously, one obtains $\hat{\boldsymbol{\eta}} = (\hat{\mathbf{b}}, \hat{\lambda})^T$. The proposed EB estimator of μ_j is then given by

$$\hat{\mu}_j^{EB} = (1 - \hat{B}_j)y_j + \hat{B}_j\psi'(\mathbf{x}_j^T\hat{\mathbf{b}}), \qquad (2.13)$$

where $\hat{B}_j = \hat{\lambda}/(\hat{\lambda} + n_j)$.

The equations (2.11) and (2.12) can only be numerically obtained. We accomplish this by the Nelder-Meade algorithm. Specifically, we use the *optim* function built in R to solve these equations. The R code can be found from the authors.

In order to find $\hat{\boldsymbol{\eta}}$, we first need to find $\mathbf{V}_j$. We have seen already that $\mu_{2j} = \phi_j V(m_j)$. The following theorem provides expressions for μ_{3j} and μ_{4j}.

Theorem 2.1 . *(a)* $\mu_{3j} = \frac{V(m_j)V'(m_j)(\lambda+n_j)(\lambda+2n_j)}{n_j^2(\lambda-v_2)(\lambda-2v_2)}$ *provided the denominator is not zero;*
(b) Let $d = v_1^2 - 4v_0v_2$ *(cf. Morris, 1983a). Then*

$$\begin{aligned}
\mu_{4j} &= \frac{d + 3V(m_j)(n_j + 2v_2)}{n_j^3}V(m_j)\\
&+ \frac{dv_2 + 6n_j^2V(m_j) + (7n_j + 6v_2)(d/dm_j)(V(m_j)V'(m_j))}{n_j^3}\\
&\times E(\mu_j - m_j)^2\\
&+ \frac{6(n_j + v_2)(n_j + 2v_2)V'(m_j)}{n_j^2}E(\mu_j - m_j)^3\\
&+ \frac{(n_j + v_2)(n_j + 2v_2)(n_j + 3v_2)}{n_j^3}E(\mu_j - m_j)^4.
\end{aligned}$$

The proof of this theorem involves heavy algebra, and is omitted. The details can be found from the authors. We have noted already that $E(\mu_j - m_j)^2 = Var(\mu_j) = V(m_j)/(\lambda - v_2)$. The expressions for $E(\mu_j - m_j)^3$ and $E(\mu_j - m_j)^4$ are found from Morris (1983a, Theorem 5.3), and, in our notations, are given by

$$E(\mu_j - m_j)^3 = 2V(m_j)V'(m_j)/[(\lambda - v_2)(\lambda - 2v_2)]$$

and

$$E(\mu_j - m_j)^4 = [3(\lambda + 6v_2)V^2(m_j) + 6dV(m_j)]/[(\lambda - v_2)(\lambda - 2v_2)(\lambda - 3v_2)]$$

provided $\lambda > max(v_2, 3v_2)$.

The expressions given in Theorem 2.1 simplify somewhat for the binomial and Poisson cases. In particular for the binomial case (where $v_0 = 0, v_1 = 1, v_2 = -1$),

$$\mu_{2j} = \frac{m_j(1-m_j)(\lambda+n_j)}{n_j(\lambda+1)}; \tag{2.14}$$

$$\mu_{3j} = \frac{m_j(1-m_j)(1-2m_j)(\lambda+n_j)(\lambda+2n_j)}{n_j^2(\lambda+1)(\lambda+2)}; \tag{2.15}$$

$$\begin{aligned}\mu_{4j} = {} & \frac{3(n_j-1)(n_j-2)(n_j-3)m_j(1-m_j)[2(1-3m_j+3m_j^2)+\lambda m_j(1-m_j)]}{n_j^3(\lambda+1)(\lambda+2)(\lambda+3)} \\ & + \frac{12(n_j-1)(n_j-2)m_j(1-m_j)(1-2m_j)^2}{n_j^3(\lambda+1)(\lambda+2)} \\ & + \frac{6(n_j-1)(n_j-6)m_j^2(1-m_j)^2}{n_j^3(\lambda+1)} \\ & + \frac{m_j(1-m_j)}{n_j^3(\lambda+1)} \\ & \times \{7(n_j-1)+(\lambda+1)[1+3(n_j-2)(\lambda+1)m_j(1-m_j)]\}.\end{aligned} \tag{2.16}$$

The expressions are even simpler for the Poisson case. Here, $v_0 = v_2 = 0$ and $v_1 = 1$. Then,

$$\mu_{2j} = \frac{m_j(\lambda+n_j)}{n_j\lambda};$$

$$\mu_{3j} = \frac{m_j(\lambda+n_j)(\lambda+2n_j)}{(n_j\lambda)^2};$$

$$\mu_{4j} = \frac{m_j(\lambda+n_j)(\lambda^2+6\lambda n_j+6n_j^2)}{(n_j\lambda)^3} + \frac{3m_j^2(\lambda+n_j)^2}{(n_j\lambda)^2}.$$

Remark 2.1. The optimal estimating equation approach adopted here is different from the currently available procedures for estimation of $\mathbf{b}$ and λ even in the normal case. Back to the Fay-Herriot (1979) formulation, marginally $y_j \stackrel{ind}{\sim} N(\mathbf{x}_j^T\mathbf{b}, A+\sigma_j^2)$. Now we take $\lambda = A^{-1}$ and $\xi_j = \sigma_j^{-2}$ (rather than n_j). Accordingly, the elementary estimating functions g_{1j} and g_{2j} are given respectively by $g_{1j} = y_j - \mathbf{x}_j^T\mathbf{b}$, $g_{2j} = (y_j - \mathbf{x}_j^T\mathbf{b})^2 - (A+\sigma_j^2)$. We write now

$$\mathbf{D}_j^T = \begin{bmatrix} -E(\frac{\partial g_{1j}}{\partial \mathbf{b}}) & -E(\frac{\partial g_{2j}}{\partial \mathbf{b}}) \\ -E(\frac{\partial g_{1j}}{\partial A}) & -E(\frac{\partial g_{2j}}{\partial A}) \end{bmatrix} = \begin{bmatrix} \mathbf{x}_j & \mathbf{0} \\ 0 & 1 \end{bmatrix}$$

(We are differentiating with respect to A rather than A^{-1}).

Also,

$$\mathbf{V}_j = \begin{bmatrix} A+\sigma_j^2 & 0 \\ 0 & 2(A+\sigma_j^2)^2 \end{bmatrix}.$$

This leads to the estimating equations

$$\textstyle\sum_{j=1}^k (A+\sigma_j^2)^{-1}(y_j - \mathbf{x}_j^T\mathbf{b})\mathbf{x}_j = \mathbf{0}; \tag{2.17}$$

$$\textstyle\sum_{j=1}^k (A+\sigma_j^2)^{-2}[(y_j - \mathbf{x}_j^T\mathbf{b})^2 - (A+\sigma_j^2)] = \mathbf{0}$$

$$\Leftrightarrow \textstyle\sum_{j=1}^k (A+\sigma_j^2)^{-2}(y_j - \mathbf{x}_j^T\mathbf{b})^2 = \sum_{j=1}^k (A+\sigma_j^2)^{-1} \tag{2.18}$$

The set of estimating equations (2.17) is the same as that of Fay and Herriot (1979) and Morris (1983b). Indeed, if A were known, this provides the optimal (weighted least squares) estimator of $\mathbf{b}$. However, (2.18) is different from the equations of Fay and Herriot (1979), Morris (1983b) or Prasad and Rao (1990) which were built on a method of moments approach based on squared residuals. Sarkar and Ghosh (1998) set up an ANOVA equation for estimating λ instead of (2.18), but as mentioned in the introduction, this was more or less an ad hoc procedure.

Remark 2.2. We may notice that in the very special case of normal models with balanced data, $y_j|\theta_j \stackrel{ind}{\sim} N(\theta_j, \sigma^2)$, ($\sigma^2$ known), and $\theta_j \stackrel{ind}{\sim} N(\mathbf{x}_j^T\mathbf{b}, A)$. Here the estimating equations (2.17) and (2.18) simplify to $\sum_{j=1}^k (y_j - \mathbf{x}_j^T\mathbf{b})\mathbf{x}_j = \mathbf{0}$ and $\sum_{j=1}^k (y_j - \mathbf{x}_j^T\mathbf{b})^2 = k(A+\sigma_j^2)$. Now $\mathbf{b}$ is estimated by $\hat{\mathbf{b}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$, where $\mathbf{y} = (y_1, \ldots, y_k)^T$, and $\mathbf{X}^T = (\mathbf{x}_1, \ldots, \mathbf{x}_k)$. Then A is estimated by $\hat{A} = max(0, k^{-1}\sum_{j=1}^k (y_j -$

$\mathbf{x}_j^T\mathbf{b})^2 - \sigma^2)$. The resulting EB estimator of θ_j is similar to Lindley's modification of the James-Stein estimator.

Remark 2.3. One may also raise the issue of finding MLE's of $\mathbf{b}$ and λ from the marginal distribution of the y_j's. A closed form marginal likelihood for $\mathbf{b}$ and λ for the general NEF-QVF family does not seem feasible. Datta and Lahiri (2001) have proposed ML and REML methods for estimation in the normal case. The calculation of ML seems quite difficult even for the beta-binomial case due to the presence of a many digamma functions involved in the derivatives of the likelihood. This difficulty will be more pronounced in the next section when we seek approximations to the MSE's of the EB estimators.

Remark 2.4. Morris (1988) found approximate EB's for NEF-QVF models. He considered only the situation when each stratum contains the same number of observations. More importantly, he approximated the marginal distributions (after integrating out the parameters) of the observations with NEF-QVF likelihoods and conjugate priors by another NEF-QVF likelihood. This amounts, for example, to approximating a beta-binomial by a binomial, or a gamma-Poisson by a Poisson. The approximation is *exact* only for the normal distribution. Also, rather than finding estimates of the parameters of the conjugate prior, Morris put some prior on these parameters, similar to, but not the same as Jeffreys' prior. He then found approximations to the resulting posterior means and variances of the parameters of interest. Thus, Morris's approach is primarily Bayesian, whereas ours is primarily frequentist. Indeed, our mean squared error calculation is based on an overdispersed exponential family model. Also, it is not clear how to extend Morris's Bayesian approximation to the unbalanced case where the different strata contain unequal number of observations.

3 Mean Squared Error Approximation

First we find an asymptotic (up to the order k^{-1}) expansion of MSE of $\hat{\mu}_j^{EB}$'s. We calculate

$$E(\mu_j - \hat{\mu}_j^{EB})^2 \quad = \quad E\{\mu_j - [(1 - B_j)y_j + B_j m_j]$$

$$
\begin{aligned}
&+[(1-B_j)y_j+B_jm_j-(1-\hat{B}_j)y_j-\hat{B}_j\hat{m}_j]\}^2\\
=\;& E\{\mu_j-[(1-B_j)y_j+B_jm_j]\}^2\\
&+E[(1-B_j)y_j+B_jm_j-(1-\hat{B}_j)y_j-\hat{B}_j\hat{m}_j]^2\\
=\;& T_{1j}(\boldsymbol{\eta})+T_{2j}(\boldsymbol{\eta}) \text{ (say).} \qquad (3.1)
\end{aligned}
$$

First we calculate

$$
\begin{aligned}
T_{1j}(\boldsymbol{\eta}) &= E[(1-B_j)(y_j-\mu_j)-B_j(\mu_j-m_j)]^2\\
&= (1-B_j)^2[V(\mu_j)/n_j]+B_j^2V(m_j)/(\lambda-v_2)\\
&= (1-B_j)^2\frac{V(m_j)+v_2V(m_j)/(\lambda-v_2)}{n_j}+B_j^2V(m_j)/(\lambda-v_2)\\
&= \frac{n_j^2\lambda V(m_j)}{(\lambda+n_j)^2(\lambda-v_2)n_j}+\frac{\lambda^2}{(\lambda+n_j)^2}\frac{V(m_j)}{\lambda-v_2}=\frac{\lambda}{(\lambda+n_j)}\frac{V(m_j)}{\lambda-v_2}\\
&= V(m_j)h_j(\lambda), \qquad (3.2)
\end{aligned}
$$

where $h_j(\lambda)=\lambda(\lambda+n_j)^{-1}(\lambda-v_2)^{-1}$.

In order to evaluate $T_{2j}(\boldsymbol{\eta})$, first we write $q(\boldsymbol{\eta},y_j)=(1-B_j)y_j+B_jm_j$. Then, by one term Taylor expansion,

$$
\begin{aligned}
T_{2j}(\boldsymbol{\eta}) &= E[q(\hat{\boldsymbol{\eta}},y_j)-q(\boldsymbol{\eta},y_j)]^2 \doteq E[(\frac{\partial q}{\partial\boldsymbol{\eta}})^T(\hat{\boldsymbol{\eta}}-\boldsymbol{\eta})]^2\\
&= E[tr\{(\frac{\partial q}{\partial\boldsymbol{\eta}})(\frac{\partial q}{\partial\boldsymbol{\eta}})^T(\hat{\boldsymbol{\eta}}-\boldsymbol{\eta})(\hat{\boldsymbol{\eta}}-\boldsymbol{\eta})^T\}]\\
&= trE[(\frac{\partial q}{\partial\boldsymbol{\eta}})(\frac{\partial q}{\partial\boldsymbol{\eta}})^T(\hat{\boldsymbol{\eta}}-\boldsymbol{\eta})(\hat{\boldsymbol{\eta}}-\boldsymbol{\eta})^T]. \qquad (3.3)
\end{aligned}
$$

But $\frac{\partial q}{\partial\mathbf{b}}=B_jV(m_j)\mathbf{x}_j$ and $\frac{\partial q}{\partial\lambda}=-\frac{\partial B_j}{\partial\lambda}(y_j-m_j)=n_j(y_j-m_j)/(\lambda+n_j)^2$. Hence, from (3.3),

$$
\begin{aligned}
&T_{2j}(\boldsymbol{\eta}) \doteq\\
&trE[\begin{pmatrix} B_j^2V(m_j)\mathbf{x}_j\mathbf{x}_j^T & \frac{n_jB_j}{(\lambda+n_j)^2}(y_j-m_j)V(m_j)\mathbf{x}_j\\ \frac{n_jB_j}{(\lambda+n_j)^2}(y_j-m_j)V(m_j)\mathbf{x}_j^T & \frac{n_j^2}{(\lambda+n_j)^4}(y_j-m_j)^2\end{pmatrix}\\
&\times(\hat{\boldsymbol{\eta}}-\boldsymbol{\eta})(\hat{\boldsymbol{\eta}}-\boldsymbol{\eta})^T]. \qquad (3.4)
\end{aligned}
$$

We will show that the right hand side of (3.4) can be approximated by

$$
tr\left[\begin{pmatrix} B_j^2V(m_j)\mathbf{x}_j\mathbf{x}_j^T & \mathbf{0}\\ \mathbf{0} & \frac{n_jV(m_j)}{(\lambda+n_j)^3(\lambda-v_2)}\end{pmatrix}\mathbf{U}_k^{-1}\right], \qquad (3.5)
$$

where $\mathbf{U}_k = \sum_{j=1}^{k} \mathbf{D}_j^T \mathbf{V}_j^{-1} \mathbf{D}_j$. The derivation of (3.5) is given in the Appendix.

Since, $\mathbf{U}_k = O(k)$, $\mathbf{U}_k^{-1} = O(k^{-1})$, we estimate $\mathbf{T}_{2j}(\boldsymbol{\eta})$ by

$$T_{2j}(\hat{\boldsymbol{\eta}}) = tr\left[\begin{pmatrix} \hat{B}_j^2 V(\hat{m}_j)\mathbf{x}_j\mathbf{x}_j^T & \mathbf{0} \\ \mathbf{0} & \frac{n_j V(\hat{m}_j)}{(\hat{\lambda}+n_j)^3(\hat{\lambda}-v_2)} \end{pmatrix} \hat{\mathbf{U}}_k^{-1}\right], \qquad (3.6)$$

where $\hat{\mathbf{U}}_k = \sum_{j=1}^{k} \mathbf{D}_j^T(\hat{\boldsymbol{\eta}}) \mathbf{V}_j^{-1}(\hat{\boldsymbol{\eta}}) \mathbf{D}_j(\hat{\boldsymbol{\eta}})$. This approximation is accurate up to $O(k^{-1})$.

Next to handle $\mathbf{T}_{1j}(\boldsymbol{\eta})$, writing the right hand side of (3.2) as $v_j(\boldsymbol{\eta})$, by a two step Taylor approximation,

$$v_j(\hat{\boldsymbol{\eta}}) \doteq v_j(\boldsymbol{\eta}) + (\frac{\partial v_j}{\partial \boldsymbol{\eta}})^T(\hat{\boldsymbol{\eta}}-\boldsymbol{\eta}) + \frac{1}{2}(\hat{\boldsymbol{\eta}}-\boldsymbol{\eta})^T(\frac{\partial v_j}{\partial \boldsymbol{\eta}})(\frac{\partial v_j}{\partial \boldsymbol{\eta}})^T(\hat{\boldsymbol{\eta}}-\boldsymbol{\eta}). \qquad (3.7)$$

Thus,

$$E[v_j(\hat{\boldsymbol{\eta}})] \doteq$$
$$v_j(\boldsymbol{\eta}) + (\frac{\partial v_j}{\partial \boldsymbol{\eta}})^T E(\hat{\boldsymbol{\eta}}-\boldsymbol{\eta}) + \frac{1}{2} tr[(\frac{\partial^2 v_j}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^T} E(\hat{\boldsymbol{\eta}}-\boldsymbol{\eta})(\hat{\boldsymbol{\eta}}-\boldsymbol{\eta})^T]. \qquad (3.8)$$

Noting that

$$\frac{\partial v_j}{\partial \boldsymbol{\eta}} = \begin{pmatrix} \frac{\partial v_j}{\partial \mathbf{b}} \\ \frac{\partial v_j}{\partial \lambda} \end{pmatrix} = \begin{pmatrix} h_j(\lambda)V(m_j)V'(m_j)\mathbf{x}_j \\ h_j'(\lambda)V(m_j) \end{pmatrix}$$

and

$$\frac{\partial^2 v_j}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^T} =$$
$$\begin{bmatrix} \{2v_2 V(m_j)(V'(m_j))^2\} V(m_j) h_i(\lambda)\mathbf{x}_j\mathbf{x}_j^T & V(m_j)(V'(m_j))h_j'(\lambda)\mathbf{x}_j \\ V(m_j)(V'(m_j))h_j'(\lambda)\mathbf{x}_j^T & V(m_j)h_j''(\lambda) \end{bmatrix},$$

one has

$$tr[(\frac{\partial^2 v_j}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^T} E(\hat{\boldsymbol{\eta}}-\boldsymbol{\eta})(\hat{\boldsymbol{\eta}}-\boldsymbol{\eta})^T] \doteq \qquad (3.9)$$
$$tr\left[\begin{pmatrix} h_j^2(\lambda)V^2(m_j)[V'(m_j)]^2\mathbf{x}_j\mathbf{x}_j^T & h_j(\lambda)n_j'(\lambda)V^2(m_j)V'(m_j)\mathbf{x}_j \\ h_j(\lambda)n_j'(\lambda)V^2(m_j)V'(m_j)\mathbf{x}_j^T & [h_j'(\lambda)]^2 V^2(m_j) \end{pmatrix} \mathbf{U}_k^{-1}\right]$$

which is $O(k^{-1})$, and is estimated by

$$tr\left[\begin{pmatrix} h_j^2(\hat{\lambda})V^2(\hat{m}_j)[V'(\hat{m}_j)]^2\mathbf{x}_j\mathbf{x}_j^T & h_j(\hat{\lambda})h_j'(\hat{\lambda})V^2(\hat{m}_j)V'(\hat{m}_j)\mathbf{x}_j \\ h_j(\hat{\lambda})h_j'(\hat{\lambda})V^2(\hat{m}_j)V'(\hat{m}_j)\mathbf{x}_j^T & [h_j'(\hat{\lambda})]^2 V^2(\hat{m}_j) \end{pmatrix} \hat{\mathbf{U}}_k^{-1}\right]$$
$$= \mathbf{T}_{3j}(\hat{\boldsymbol{\eta}}) \text{ (say)} \qquad (3.10)$$

Finally we approximate $E(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})$ (up to $O(k^{-1})$) by the method of Cox and Snell (1968). We will write

$$E(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}) = T_4(\boldsymbol{\eta}), \tag{3.11}$$

where $T_4(\boldsymbol{\eta})$ will be derived in the appendix. We will also show that $T_4(\boldsymbol{\eta})$ is $O(k^{-1})$.

Thus from (3.8), (3.10) and (3.11), we estimate $v_j(\boldsymbol{\eta})$ by

$$\begin{aligned} & v_j(\hat{\boldsymbol{\eta}}) - \left[\begin{array}{cc} n_j(\hat{\lambda})V(\hat{m}_j)V'(\hat{m}_j)\mathbf{x}_j^T & n_j'(\hat{\lambda})V(\hat{m}_j) \end{array}\right] \mathbf{T}_4(\hat{\boldsymbol{\eta}}) - \frac{1}{2}\mathbf{T}_{3j}(\hat{\boldsymbol{\eta}}) \\ & = T_{5j}(\hat{\boldsymbol{\eta}}) \text{ (say)} \end{aligned} \tag{3.12}$$

The final approximate estimator of the MSE $E(\hat{\mu}_j^{EB} - \mu_j)^2$ is thus $T_{5j}(\hat{\boldsymbol{\eta}}) + T_{2j}(\hat{\boldsymbol{\eta}})$, where $T_{2j}(\hat{\boldsymbol{\eta}})$ is given in (3.6).

4 An Example

Efron (1978, 1986) considered an example showing the proportion of subjects testing positive for the disease toxoplasmosis in 34 cities of El Salvador. Efron's prime objective was to assess the effect of rainfall on the proportions positive. Efron (1978) met the objective by fitting a logistic regression to the data, and subsequently (1986), by fitting an overdispersed generalized linear model.

Our objective, however, is more modest. Since the sample sizes in the local areas are mostly small, it is anticipated that the EB estimates will have greater precision than the usual estimates, namely, the sample proportions. This is substantiated in the following table which provides the sample sizes (n_j), the the sample proportions ($\hat{p}_j$), the EB estimates ($\hat{p}_j^{EB}$), the estimated standard errors of the sample proportions, namely, $\mathrm{SE}(\hat{p}_j) = [\hat{p}_j(1-\hat{p}_j)/n_j]^{1/2}$, the naive standard errors of the EB estimators, namely, $\mathrm{MSE}(\hat{p}_j^{EB}) = [V((1-\hat{B}_j)y_j + \hat{B}_j\hat{m}_j)/(n_j+\hat{\lambda}+1)]^{1/2}$, and the approximate root mean squared errors of the EB estimators as derived in Section 3, denoted by $\mathrm{RMSE}(\hat{p}_j^{EB})$. We have considered the model $m_j = b_0 + b_1x_j + b_2x_j^2 + b_3x_j^3,\ j = 1, 2, \cdots, 34$, where x_j denotes the amount of rainfall in the jth city. The cubic equation is analogous to what is considered in Efron (1986).

City(j)	n_j	$\hat{p}_j$	$\hat{p}_j^{EB}$	SE($\hat{p}_j$)	MSE($\hat{p}_j^{EB}$)	RMSE($\hat{p}_j^{EB}$)
1	4	.500	.394	.250	.143	.165
2	5	.200	.389	.179	.144	.167
3	2	1.000	.513	.000	.162	.189
4	8	.250	.256	.153	.114	.127
5	6	.500	.469	.204	.140	.160
6	24	.292	.353	.093	.075	.078
7	30	.500	.511	.091	.075	.082
8	1	.000	.501	.000	.193	.231
9	1	.000	.338	.000	.197	.233
10	9	.444	.501	.166	.120	.137
11	12	.167	.191	.134	.114	.130
12	11	.727	.690	.134	.114	.130
13	51	.471	.462	.070	.058	.059
14	82	.561	.569	.055	.042	.042
15	43	.535	.525	.076	.064	.066
16	13	.615	.530	.135	.094	.101
17	6	.167	.327	.152	.140	.161
18	10	.300	.346	.145	.119	.133
19	10	.300	.374	.145	.118	.132
20	5	.600	.445	.219	.129	.147
21	19	.368	.436	.111	.080	.083
22	10	.800	.643	.126	.104	.113
23	1	.000	.215	.000	.177	.207
24	22	.182	.268	.082	.077	.080
25	11	.545	.468	.150	.101	.109
26	54	.611	.586	.066	.051	.051
27	8	.625	.672	.171	.109	.131
28	1	.000	.474	.000	.196	.260
29	77	.532	.518	.057	.043	.043
30	16	.437	.414	.124	.093	.100
31	13	.692	.713	.128	.090	.096
32	75	.707	.685	.053	.045	.045
33	10	.300	.322	.145	.115	.128
34	37	.622	.595	.080	.067	.073

Table 1: The estimates and the standard errors.

Table 1 exhibits several interesting features. First, the naive standard errors grossly underestimate the precisions of the estimates, especially for the strata where the sample sizes are very small. Naturally, these are the strata where the RMSE($\hat{p}_j^{EB}$) correct these naive standard errors the most. For example, in City 3, the correction is of the order 16% over the naive estimate, whereas in cities 13, 14, 26, 29 and 32, there is no correction at all. Also, there are several strata where the sample sizes are 1, leading necessarily to zero standard errors of the classical MLE's. The EB RMSE's provide more credible measures of uncertainty in these cases. For other strata, RMSE($\hat{p}_j^{EB}$) are always smaller than SE($\hat{p}_j$)'s, and sometimes are substantially smaller when the strata sizes are very small (but not 1 or $\hat{p}_j \neq 1$). This is evidenced, for example, in city 1, where the improvement is of the order 33%. Also, the MLE's for the population proportions are shrunk most for strata with small sample sizes rather than those with moderate or large sample sizes.

5 Concluding Remarks

The paper develops EB estimators for simultaneous estimation in NEF-QVF populations. The EB estimators are obtained by employing the theory of optimal estimating functions as proposed in Godambe and Thompson (1989). We provide also closed form approximate MSE's in a vein similar to Prasad and Rao (1990). An example illustrates the applicability and merits of the proposed method. The method bears tremendous potential in the context of small area estimation where the number of local areas is very large, but sample sizes within these areas is very small.

6 Appendix

Derivation of (3.5):

The first step is to obtain the asymptotic distribution of $\hat{\boldsymbol{\eta}}$. Once again, by one step Taylor expansion,

$$\mathbf{0} = \mathbf{S}_k(\hat{\boldsymbol{\eta}}) \doteq \mathbf{S}_k(\boldsymbol{\eta}) + \left(\frac{\partial \mathbf{S}_k(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}}\right)^T (\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}).$$

Thus,

$$\hat{\boldsymbol{\eta}} - \boldsymbol{\eta} \doteq \left[\left(-\frac{\partial \mathbf{S}_k(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}}\right)^T\right]^{-1} \mathbf{S}_k(\boldsymbol{\eta}). \tag{6.1}$$

We may note that $V(\mathbf{S}_k(\boldsymbol{\eta})) = \sum_{j=1}^k \mathbf{D}_j^T \mathbf{V}_j^{-1} \mathbf{D}_j = \mathbf{U}_k$, and under some simple regularity conditions, $\mathbf{U}_k^{-\frac{1}{2}} \mathbf{S}_k(\boldsymbol{\eta})$ is asymptotically $N_k(\mathbf{0}, \mathbf{I}_k)$ by the central limit theorem. Thus, $\mathbf{U}_k^{-\frac{1}{2}} \mathbf{S}_k(\boldsymbol{\eta}) \doteq O_p(1)$. Now, from (6.1), writing

$$\hat{\boldsymbol{\eta}} - \boldsymbol{\eta} \doteq \left[\left(-\frac{\partial \mathbf{S}_k(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}}\right)^T\right]^{-1} \mathbf{U}_k^{\frac{1}{2}} \mathbf{U}_k^{-\frac{1}{2}} \mathbf{S}_k(\boldsymbol{\eta}),$$

it follows that

$$\mathbf{U}_k^{\frac{1}{2}}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}) \doteq \mathbf{U}_k^{\frac{1}{2}} \left[\left(-\frac{\partial \mathbf{S}_k(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}}\right)^T\right]^{-1} \mathbf{U}_k^{\frac{1}{2}} \mathbf{U}_k^{-\frac{1}{2}} \mathbf{S}_k(\boldsymbol{\eta}), \tag{6.2}$$

But

$$\left(-\frac{\partial \mathbf{S}_k(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}}\right)^T = -[\sum_{j=1}^k (\frac{\partial}{\partial \boldsymbol{\eta}} \mathbf{D}_j^T \mathbf{V}_j^{-1}) \mathbf{g}_j + \sum_{j=1}^k \mathbf{D}_j^T \mathbf{V}_j^{-1} (\frac{\partial \mathbf{g}_j}{\partial \boldsymbol{\eta}})^T].$$

Hence,

$$E\left[\left(-\frac{\partial \mathbf{S}_k(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}}\right)^T\right] = \sum_{j=1}^k \mathbf{D}_j^T \mathbf{V}_j^{-1} \mathbf{D}_j = \mathbf{U}_k. \tag{6.3}$$

It follows now from (6.2) and (6.3) that $\mathbf{U}_k^{\frac{1}{2}}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})$ has the same asymptotic distribution as $\mathbf{U}_k^{-\frac{1}{2}} \mathbf{S}_k(\boldsymbol{\eta})$ which is $N_k(\mathbf{0}, \mathbf{I}_k)$. Accordingly, $\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}$ is asymptotically normal with mean vector $\mathbf{0}$ and variance-covariance matrix $\mathbf{U}_k^{-1}$. Thus, $E[(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})^T] \doteq \mathbf{U}_k^{-1}$. Since $\mathbf{U}_k^{-1}$ is $O(k^{-1})$, from (3.4),

$$T_{2j}(\boldsymbol{\eta}) \doteq$$

$$trE\left[\begin{pmatrix} B_j^2 V(m_j) \mathbf{x}_j \mathbf{x}_j^T & \frac{n_j B_j V(m_j)}{(\lambda+n_j)^2}(y_j - m_j)\mathbf{x}_j \\ \frac{n_j B_j V(m_j)}{(\lambda+n_j)^2}(y_j - m_j)\mathbf{x}_j^T & \frac{n_j^2}{(\lambda+n_j)^4}(y_j - m_j)^2 \end{pmatrix} \mathbf{U}_k^{-1}\right]$$

which immediately yields (3.5).

Derivation of (3.11)

Let $\mathbf{U}_k^{-1} = ((U_k^{rs}))$. We first need a few notations stemming out of $\mathbf{D}_j^T \mathbf{V}_j^{-1}$.

Let $e_{1j} = \Delta_j^{-1}[\mu_{4j} - \mu_{2j}^2 - \mu_{3j}\phi_j V(m_j)]$, $e_{2j} = \Delta_j^{-1}[\mu_{2j}\phi_j V'(m_j) - \mu_{3j}]$, $e_{3j} = -\Delta_j^{-1}\mu_{3j}V(m_j)$ and $e_{4j} = \Delta_j^{-1}\mu_{2j}V(m_j)$, where we may recall that $V(m_j) = v_0 + v_1 m_j + v_2 m_j^2$ and $\phi_j = (n_j + \lambda)/[n_j(\lambda - v_2)]$. We now write $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_{p+1})^T$ and

$$\mathbf{S}_k(\boldsymbol{\eta}) \equiv \mathbf{S}_k = (s_{k1}, \ldots, s_{kp}, s_{k,p+1})^T,$$

where

$$s_{kr} = \sum_{j=1}^{k}(e_{1j}g_{1j} + e_{2j}g_{2j})V(m_j)x_{jr}$$

and

$$s_{k,p+1} = -\sum_{j=1}^{k}(e_{3j}g_{1j} + e_{4j}g_{2j})\frac{\partial \phi_j}{\partial \lambda}.$$

We will also find it convenient to write $\lambda = b_{p+1}$. Then following Cox and Snell (1968), let $J_{t,rs} = Cov(U_k^{st}, \frac{\partial s_{kr}}{\partial b_s})$, and $K_{rtu} = E(\frac{\partial^2 s_{kr}}{\partial b_s \partial b_t})$, $(r, s, t = 1, \ldots, p+1)$. Now writing $\mathbf{J}_{k(r)} = ((J_{t,rs}))$ and $\mathbf{K}_{k(r)} = ((E(\frac{\partial^2 s_{kr}}{\partial b_s \partial b_t})))$, by the Cox-Snell formula $(r, s, t = 1, \ldots, p+1)$

$$\begin{aligned} E(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}) &= \frac{1}{2}\mathbf{U}_k^{-1}\left[2\begin{pmatrix} tr(\mathbf{U}_k^{-1}\mathbf{J}_{k(r)}) \\ tr(\mathbf{U}_k^{-1}\mathbf{J}_{k(p+1)}) \end{pmatrix} + \begin{pmatrix} tr(\mathbf{U}_k^{-1}\mathbf{K}_{k(r)}) \\ tr(\mathbf{U}_k^{-1}\mathbf{K}_{k(p+1)}) \end{pmatrix}\right] \\ &= \mathbf{T}_4(\boldsymbol{\eta}) \text{ (say)}. \end{aligned} \tag{6.4}$$

Finding the elements of $\mathbf{T}_4(\boldsymbol{\eta})$ requires heavy algebra, and can be found from the authors. We omit the details. We note also that $\mathbf{T}_4(\boldsymbol{\eta}) = O(k^{-1})$ since $\mathbf{U}_k^{-1}$ is $O(k^{-1})$ and its multiplier is $O(1)$.

Acknowledgements

This research was partially supported by NSF Grant SES-9911485 and SES-0317589.

References

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

Breslow, N. E. and Clayton D. G. (1993), Approximate inference in generalized linear mixed models. Journal of the American Statistical Association, **88**, 9-25.

Cox, D. R. and Snell E. J. (1968), A general definition of residuals. Journal of the Royal Statistical Society, Series B, **30**, 248-275.

Datta, G. S. and Lahiri, P. (2001), A unified measure of uncertainty of estimated best linear unbiased predictor in small-area estimation problems. Statistica Sinica, **10**, 613-627.

Dempster, A. P. and Tomberlin, T. J. (1980), The analysis of census undercount from a post-enumeration survey. Proceedings of the Conference on Census Undercount, pp 88-94.

Efron, B. (1978), Regression and ANOVA with zero-one data: measures of residual variation. Journal of the American Statistical Association, **73**, 113-121.

Efron, B. (1986), Double exponential families and their use in generalized linear regression. Journal of the American Statistical Association, **81**, 709-721.

Farrell, P. J., MacGibbon, B., and Tomberlin, T. J. (1997), Empirical Bayes estimates of small area proportion in multistage designs. Statistica Sinica, **7**, 1065-1083.

Fay, R. E. and Herriot, R. A. (1979), Estimates of income from small places: an application of James-Stein procedures to census data. Journal of the American Statistical Association, **74**, 269-277.

Ghosh, M., Natarajan, K., Stroud, T. W. F., and Carlin, B. P. (1998), Generalized linear models for small-area estimation. Journal of the American Statistical Association, **93**, 273-282.

Godambe, V. P. and Thompson, M. E. (1989), An extension of quasi-likelihood estimation. (with discussion), Journal of Statistical Planning and Inference, **22**, 137-172.

Jiang, J. and Lahiri, P. (2001), Empirical best prediction for small area inference with binary data. Annals of the Institute of Statistical Mathematics, **53**, 217-243.

MacGibbon, B. and Tomberlin, T. J. (1989), Small area estimation of proportions via empirical Bayes methods. Survey Methodology, **15**, 237-252.

McCullagh, P. and Nelder, J. A. (1989), Generalized Linear Models. 2nd edition, London: Chapman and Hall.

Morris, C. (1982), Natural exponential families with quadratic variance functions. The Annals of Statistics, **10**, 65-80.

Morris, C. (1983a), Natural exponential families with quadratic variance functions: statistical theory. The Annals of Statistics, **11**, 515-529.

Morris, C. (1983b), Parametric empirical Bayes inference: theory and applications. (with discussion), Journal of the American Statistical Association, **78**, 47-65.

Morris, C. (1988), Determining the accuracy of Bayesian empirical Bayes estimates in the familiar exponential families. Statistical Decision Theory and Related Topics IV, V1. Eds. S. S. Gupta and J. Berger. Springer-Verlag, New York, pp 251-263.

Prasad, N. G. N. and Rao, J. N. K. (1990), The estimation of the mean squared error of small-area estimators. Journal of the American Statistical Association, **85**, 163-171.

Sarkar, S. and Ghosh, M. (1998), Empirical Bayes estimation of local area means for NEF-QVF superpopulations. Sankhyā, Series B, **60**, 464-487.