SECOND ORDER BAYESIAN REVISION OF A GENERALISED LINEAR MODEL

Greg Taylor

Taylor Fry Consulting Actuaries
Level 8, 30 Clarence Street
Sydney NSW 2000
Australia

Professorial Associate,
Centre for Actuarial Studies
Faculty of Economics and Commerce
University of Melbourne
Parkville VIC 3052
Australia

Adjunct Professor in Actuarial Studies
Faculty of Commerce and Economics
University of New South Wales
Kensington NSW 2033
Australia

Phone: 61 2 9249 2901
Fax: 61 2 9249 2999
greg.taylor@taylorfry.com.au

May 2005

Table of Contents

Summary **1**

1. Introduction **2**

2. Notation and preliminary results **3**

3. Exponential dispersion family **5**

4. Generalised linear models **7**

5. Bayesian revision of a GLM **10**

6. Dynamic generalised linear models **28**

References **56**

Appendices

A **Taylor expansion of (5.9)**

B **Evaluation of** $E_{\beta}\left[b'\left(t_j\right)\right]$ **and** $Var_{\beta}\left[b'\left(t_j\right)\right]$

C **Bayesian revision of mean and covariance**

D **Derivatives of** φ

E **Bayesian revision of companion GLM**

Summary

It is well known that the exponential dispersion family (EDF) of univariate distributions is closed under Bayesian revision in the presence of natural conjugate priors. However, this is not the case for the general multivariate EDF.

This paper derives a second order approximation to the posterior likelihood of a naturally conjugated generalised linear model (GLM), i.e. multivariate EDF subject to a link function (Section 5.5). It is **not** the same as a normal approximation. It does, however, lead to **second order Bayes estimators** of parameters of the posterior.

The family of second order approximations is found to be closed under Bayesian revision. This generates a recursion for repeated Bayesian revision of the GLM with the acquisition of additional data.

The recursion simplifies greatly for a canonical link. The resulting structure is easily extended to a filter for estimation of the parameters of a dynamic generalised linear model (DGLM) (Section 6.2). The Kalman filter emerges as a special case.

A second type of link function, related to the canonical link, and with similar properties, is identified. This is called here the **companion canonical link**. For a given GLM with canonical link, the companion to that link generates a **companion GLM** (Section 4). The recursive form of the Bayesian revision of this GLM is also obtained (Section 5.5.3).

There is a perfect parallel between the development of the GLM recursion and its companion. A dictionary for translation between the two is given so that one is readily derived from the other (Table 5.1).

The companion canonical link also generates a **companion DGLM**. A filter for this is obtained (Section 6.3).

Keywords: Bayesian revision, companion canonical link, dynamic generalised linear model, exponential dispersion family, generalised linear model, Kalman filter.

1. Introduction

This paper is concerned with a **generalised linear model (GLM)** whose parameter vector is subject to a prior distribution, and with calculation of the posterior distribution of that parameter vector conditional on a vector of data. This is the framework of the **dynamic generalised linear model (DGLM)**.

A natural conjugate prior can be found, and the posterior calculated quite simply, in certain degenerate cases, eg for a 1-dimensional data vector and identity link in the GLM, which amounts to Bayesian revision of a univariate member of the **exponential dispersion family (EDF)** (Nelder and Verrall, 1997; Landsman and Makov, 1998).

An alternative degenerate case occurs when the natural parameter of the GLM is assumed subject to a natural conjugate prior. However, as pointed out by West, Harrison and Migon (1985), this imposes severe and unrealistic restrictions on the form of the prior.

In the general case the Bayesian revision of the GLM is analytically rather intractable. Various devices have been used to approximate the revision. For example, West, Harrison and Migon retain the natural conjugate prior on the natural parameter mentioned above and impose certain specific forms of parameter revision that are justified by analogy with dynamic linear models.

Fahrmeir and Kaufman (1991) and Fahrmeir (1992) focus on estimation of the **mode** of the posterior rather than entire distribution or its mean. Smith (1979) approached the problem from a decision theoretic standpoint. De Jong (1997) addresses the Bayesian revision problem by means of a **scan sampler**. An example is given in which this is applied to a DGLM, but one in which the prior is Gaussian.

More recent literature has approached Bayesian models of this type by means of Markov chain Monte Carlo (MCMC) simulation (Hastings, 1970; Smith and Roberts, 1993; Tierney, 1994). Examples in the actuarial literature appear in Scollnik (2001, 2002), Ntzoufras and Dellaportas (2002), de Alba (2002).

While the MCMC approach is flexible and practical, there remains some advantage in the use of closed form analytical expressions for Bayesian revision. The present paper pursues this objective, but does so in a manner different from the analytic approaches mentioned above.

It adopts a family of priors $\mathtt{F}$ on the GLM parameter vector (not the natural parameter) that is a natural generalisation of natural conjugate prior for the univariate case mentioned above. In general, these priors are not natural conjugate, and so the posterior does not lie within $\mathtt{F}$. It can, however, be approximated by the member of $\mathtt{F}$ with the same first and second moments.

In certain cases, this approximate posterior is in a form convenient for it to function as a prior for a further Bayesian revision. Repeated revision can then take place as additional items of data are received. This is of particular use in the estimation of DGLM parameters.

2. Notation and preliminary results

When it is helpful, the dimensions of vectors and matrices will be written beneath the symbols representing them, eg $\underset{m \times q}{X}$.

If f: $\mathcal{R} \rightarrow \mathcal{R}$, and $x = (x_1,...,x_m) \in \mathcal{R}^m$, then $f(x)$ will be understood as the vector

$$\underset{m \times 1}{f(x)} = \left(f(x_1),...,f(x_m)\right)^T \tag{2.1}$$

ie scalar functions will be understood to operate component-wise on vectors.

Composition of functions will be indicated by the operator symbol $\circ$. Thus, for scalar functions f and g

$$f \circ g(x) = f\left(g(x)\right). \tag{2.2}$$

For any vector $x = (x_1,...,x_m)^T$, define

$$DIAG \underset{m \times 1}{x} = \underset{m \times m}{diag(x_1,...,x_m)} \tag{2.3}$$

The inverse operation is denoted VEC, ie

$$VEC\, \underset{m \times m}{diag(x_1,...,x_m)} = \underset{m \times 1}{x} = (x_1,...,x_m)^T.$$

For $g_i : \mathcal{R}^q \rightarrow \mathcal{R}, (Dg_i)(t)$ denotes the **gradient** of g_i evaluated at $t \in \mathcal{R}^q$, ie

$$(Dg_i)(t) = \left[\partial g_i(t)/\partial t_1,...,\partial g_i(t)/\partial t_q\right]. \tag{2.4}$$

For $g = (g_1,...,g_m)^T : \mathcal{R}^q \rightarrow \mathcal{R}^m, (Dg)(t)$ denotes the **Jacobian matrix** evaluated at t, ie:

$$\underset{m \times q}{(Dg)(t)} = \begin{bmatrix} (Dg_1)(t) \\ \vdots \\ (Dg_m)(t) \end{bmatrix}. \tag{2.5}$$

Also, $(D^2 g_i)(t)$ denotes the q x q **Hessian matrix** with (j,k) element $\partial^2 g_i(t)/\partial t_j \partial t_k$. Let $(D^2 g)(t)$ denote the **stacked Hessian matrix**:

$$\underset{mq \times q}{(D^2 g)(t)} = \begin{bmatrix} (D^2 g_1)(t) \\ \vdots \\ (D^2 g_m)(t) \end{bmatrix}. \tag{2.6}$$

For m x q and n x r matrices A and B, with (j,k) elements a_{jk} and b_{jk} respectively, the Kronecker product $A \otimes B$ is defined as the following partitioned matrix:

$$\underset{mn \times qr}{A \otimes B} = \begin{bmatrix} a_{11}B & a_{12}B & & \\ a_{21}B & a_{22}B & & \\ & & \ddots & \\ & & & a_{mn}B \end{bmatrix}. \tag{2.7}$$

A subscript associated with an expectation operator indicates that the expectation is taken with respect to the variate appearing in the subscript, ie

$$E_\beta\left[f(\beta)\right] = \int f(\beta)\,dF(\beta) \tag{2.8}$$

where β is a random variable with measure F.

Matrix identities

Occasionally, the following results will be useful. Let A,B be matrices such that AB is square and $1+AB$ is non-singular. Then

$$(1+AB)^{-1}A = A(1+BA)^{-1} \tag{2.9}$$

$$B(1+AB)^{-1}A = (1+BA)^{-1}BA. \tag{2.10}$$

Taylor expansions

The Taylor expansion of f: $\mathcal{R}^q \rightarrow \mathcal{R}$ about μ, truncated to second order, is

$$f(x) = f(\mu) + \underset{1 \times q}{Df(\mu)}(x-\mu) + \tfrac{1}{2}(x-\mu)^T \underset{q \times q}{Df(\mu)}(x-\mu). \tag{2.11}$$

The corresponding expansion of f: $\mathcal{R}^q \rightarrow \mathcal{R}^m$ is

$$\underset{m \times 1}{f(x)} = \underset{m \times 1}{f(\mu)} + \underset{m \times q}{Df(\mu)}\underset{q \times 1}{(x-\mu)} + \tfrac{1}{2}\begin{bmatrix} (x-\mu)^T & D^2 f_1(\mu) & (x-\mu) \\ & \vdots & \\ \underset{1 \times q}{(x-\mu)^T} & \underset{q \times q}{D^2 f_1(\mu)} & \underset{q \times 1}{(x-\mu)} \end{bmatrix}$$

$$= f(\mu) + Df(\mu)(x-\mu) + \tfrac{1}{2}\left[\underset{q \times 1}{(x-\mu)} \otimes \underset{m \times m}{1}\right]^T \underset{mq \times q}{D^2 f(\mu)}(x-\mu). \tag{2.12}$$

Taking expectations on both sides of (2.12) gives

$$Ef\left(X\right)=f\left(\mu\right)+\tfrac{1}{2}\begin{bmatrix}Tr\left[Var\left[X\right]D^2 f_1\left(\mu\right)\right]\\ \vdots \\ Tr\left[Var\left[X\right]D^2 f_m\left(\mu\right)\right]\end{bmatrix}, \qquad (2.13)$$

where Tr denotes the trace of its argument.

A special case arises when $q = m$ and f has the meaning assigned in (2.1) for f: $\mathcal{R} \to \mathcal{R}$. Then $D^2 f_i\left(\mu\right)$ has all elements zero except $f''\left(\mu_i\right)$ in the (i,i) position, and (2.13) reduces to

$$\underset{m\times 1}{Ef\left(X\right)}=\underset{m\times 1}{f\left(\mu\right)}+\tfrac{1}{2}\left[PRIN\,\underset{m\times m}{Var\left[X\right]}\right]\underset{m\times 1}{f''\left(\mu\right)} \qquad (2.14)$$

where, for an m x m matrix A, $PRIN\,A$ denotes the matrix obtained by setting all elements off the principal diagonal to zero.

Taking into account that $Ef\left(X\right)=f\left(\mu\right)$ to first order, by (2.14), equation (2.12) gives the following second order approximation:

$$\begin{aligned}Var\,f\left(X\right)&=E\left[Df\left(\mu\right)\left(X-\mu\right)\left(X-\mu\right)^T\left[Df\left(\mu\right)\right]^T\right]\\&=Df\left(\mu\right)Var\left[X\right]\left[Df\left(\mu\right)\right]^T\end{aligned} \qquad (2.15)$$

3. Exponential dispersion family

The **exponential dispersion family** was introduced by Nelder and Wedderburn (1972) and is treated in depth by McCullagh and Nelder (1989).

Consider a real valued random variable Y, depending on a **location parameter** θ and a **scale parameter** λ, and subject to the (quasi-) log-likelihood:

$$L\left(y;\theta,\lambda\right)=\lambda\left[y\theta-b\left(\theta\right)\right]+k\left(\lambda,y\right) \qquad (3.1)$$

for suitable functions b (called the **cumulant function**) and k.

The EDF is the family of such random variables.

The member of the family represented by (3.1) can be denoted more explicitly as $L\left(y;\theta,\lambda,b,k\right)$. This notation will be used occasionally, though the briefer form will be preferred where the context makes the choice of b and k clear.

The moments of Y are obtained by repeated differentiation of the equation $\int \exp L\left(y;\theta,\lambda\right)dy=1$ with respect to θ, giving

$$E\left[Y\mid\theta,\lambda\right]=b'\left(\theta\right) \qquad (3.2)$$

$$Var[Y \mid \theta,\lambda] = b''(\theta)/\lambda. \tag{3.3}$$

Denote $b'(\theta)$ by $\mu(\theta)$ whence, provided that $\mu(\cdot)$ is one-one,

$$Var[Y \mid \theta,\lambda] = V(\mu(\theta))/\lambda \tag{3.4}$$

for some function $V(\cdot)$ called the **variance function**.

Many applications of the EDF restrict the form of the variance function thus:

$$V(\mu) = \mu^p \tag{3.5}$$

for some constant *p*. This will be referred to subsequently as the case of the **power variance function**.

It may be checked, by reference to (3.2) and (3.3), that (3.5) corresponds to

$$b(\theta) = (2-p)^{-1}\left[(1-p)\theta\right]^{(2-p)/(1-p)} \tag{3.6}$$

where $p \in (-\infty,+\infty)$ and (3.6) contains the limiting cases

$$b(\theta) = \exp\theta \text{ for } p = 1 \tag{3.7}$$
$$= -\log(-\theta) \text{ for } p = 2. \tag{3.8}$$

In the case of the power variance function, substitution of (3.6) in (3.2) gives

$$\mu(\theta) = \left[(1-p)\theta\right]^{1/(1-p)} \tag{3.9}$$

$$\theta = \mu^{1-p}/(1-p) \tag{3.10}$$

with a limiting case for $p = 1$.

The power variance case includes normal $(p=0)$, Poisson $(p=1)$, gamma $(p=2)$ and inverse Gaussian $(p=3)$ distributions. The most common distribution **not** included in it is binomial, for which

$$b(\theta) = \log\left(1+e^{\theta}\right) \tag{3.11}$$

$$V(\mu) = \mu(1-\mu). \tag{3.12}$$

4. Generalised linear models

Let $Y=\left(Y_1,\ldots,Y_m\right)^T$ be a vector valued random variable with each of the stochastically independent components Y_i a member of the EDF $L\left(y_i;\theta_i,\lambda_i,b,k\right)$. Thus, all components have a common cumulant function, but are allowed different location and scale parameters.

Suppose further that the location parameter vector θ can be expressed in terms of a linear transformation of a further parameter vector $\beta=\left(\beta_1,\ldots,\beta_q\right)^T, q\le m$:

$$\underset{m\times 1}{\mu\left(\theta\right)}=h^{-1}\left(\underset{m\times q}{X}\;\underset{q\times 1}{\beta}\right) \qquad (4.1)$$

where $\mu\left(\theta\right)$ is the *m*-vector of values $\mu\left(\theta_i\right)$, X is an *m* x *q* **design matrix**, *h* is a strictly monotone function with range $\left(-\infty,+\infty\right)$ called the **link function**, and the vector equation (4.1) is interpreted in accordance with (2.1).

The quantity $X\beta$ in (4.1) is called the **linear response**.

Relation (4.1), together with the EDF specification of error structure, are said to constitute a **generalised linear model** for *Y*.

By (3.1), and the independence of the Y_i, the log-likelihood of *Y* takes the form

$$L\left(y;\beta,\Lambda\right)=y^T\Lambda\theta-1^T\Lambda b\left(\theta\right)+\sum_{i=1}^{m}k\left(\lambda_i,y_i\right) \qquad (4.2)$$

where $\Lambda=diag\left(\lambda_1,\ldots,\lambda_m\right)$ and 1 is the *m*-vector with all entries unity.

The GLM represented by (4.2) is defined by *b*, *h*, Λ and *X*. Denote it by $\mathcal{G}\left(b,h,\Lambda,X\right)$.

By (4.1) and definition of $\mu\left(\theta\right)$,

$$\theta=\left(b'\right)^{-1}\circ h^{-1}\left(X\beta\right) \qquad (4.3)$$

whence (4.2) takes the form (henceforth terms not involving the natural parameter θ will be omitted)

$$L\left(y;\beta,\Lambda\right)=y^T\Lambda\left(b'\right)^{-1}\circ h^{-1}\left(X\beta\right)-1^T\Lambda b\circ\left(b'\right)^{-1}\circ h^{-1}\left(X\beta\right). \qquad (4.4)$$

A link function of particular significance is the **canonical link** $h=\left(b'\right)^{-1}$. This reduces (4.3) to

$$\theta = X\beta. \tag{4.5}$$

That is, the location parameter vector is simply equal to the linear response.

In the case of the canonical link, the log-likelihood (4.2) may be expressed in the form

$$L(y;\beta,\Lambda) = y^T \Lambda X\beta - 1^T \Lambda b(X\beta) + \sum_{i=1}^{m} k(\lambda_i, y_i). \tag{4.6}$$

Though the canonical link function is widely used, note that for power variance function (3.5), it takes the form (see 3.10)

$$h(\mu) = \mu^{1-p}/(1-p) \tag{4.7}$$

which does not satisfy the requirement of range $(-\infty,+\infty)$ for p other than integral and even. As a result, it can lead to undefined values of $b(X\beta)$ in (4.6).

It will prove useful to define a **companion** to a link function h as

$$h^* = h \circ b' \circ b \circ (b')^{-1} \tag{4.8}$$

whence (4.3) is replaced by

$$\theta = (b')^{-1} \circ (h^*)^{-1} (X\beta) = b^{-1} \circ (b')^{-1} \circ h^{-1} (X\beta). \tag{4.9}$$

By (4.8) and the definition of the canonical link, the **companion canonical link** is

$$h^* = b \circ (b')^{-1}. \tag{4.10}$$

Equivalently,

$$1/(h^*)^{-1} = (b^{-1})'. \tag{4.11}$$

For the power variance function (3.5), substitution of (3.6) in (4.10) yields

$$h^*(\mu) = \mu^{2-p}/(2-p). \tag{4.12}$$

If h, h^* are written as h_p, h_p^* to make the dependence on p explicit, then it follows from (4.7) and (4.12) that

$$h_p^* = h_{p-1}. \tag{4.13}$$

Example 4.1: gamma error, companion canonical link. The gamma cumulant function is given by (3.8). Then $h^*(\mu) = \log \mu$.

□

The log-likelihood L^* of Y corresponding to L in (4.4) is obtained by using (4.9) to substitute for θ in (4.2), thus

$$\begin{aligned} L^*(y;\beta,\Lambda) &= y^T \Lambda \left(b^{-1} \circ (b')^{-1} \circ h^{-1}\right)(X\beta) \\ &\quad -1^T \Lambda b\left(\left(b^{-1} \circ (b')^{-1} \circ h^{-1}\right)(X\beta)\right). \end{aligned} \tag{4.14}$$

This may be written more concisely, by means of (4.3), as

$$L^*(y;\beta,\Lambda) = y^T \Lambda b^{-1}\left(\theta^*\right) - 1^T \Lambda \theta^*$$

in which the roles of the vectors Λy and $\Lambda 1$ in L are reversed, the cumulant function b is replaced by b^{-1}, and the natural parameter is taken as $\theta^* = b(\theta)$.

Note, however, that the cumulant function of the GLM is not changed. It is still b.

When L^* is considered in isolation, just as a conditional log-likelihood of y, rather than in conjunction with L, the natural parameter may as well be written as just θ, in which case

$$L^*(y;\beta,\Lambda) = y^T \Lambda b^{-1}(\theta) - 1^T \Lambda \theta. \tag{4.15}$$

It is also possible to define GLMs in companion pairs. For given GLM $\mathcal{G}(b,h,\Lambda,X)$, define its **companion GLM**

$$\mathcal{G}^*(b,h,\Lambda,X) = \mathcal{G}\left(b,h^*,\Lambda,X\right) \tag{4.16}$$

with log-likelihood L^* as set out in (4.15).

Note that this may be written in the alternative form

$$L^*(y;\beta,\Lambda) = 1^T \Lambda(-\theta) - y^T \Lambda b^*(-\theta) \tag{4.17}$$

if b^* is defined as

$$b^*(\theta) = -b^{-1}(-\theta). \tag{4.18}$$

Differentiating through (4.18) and comparing the result with (4.11) gives

$$1/\left(h^*\right)^{-1}(-\mu) = (d/d\mu)\, b^*(\mu). \tag{4.19}$$

Comparison of (4.17) with (4.2) shows that $\mathcal{G} \to \mathcal{G}^*$ corresponds to the following replacements:

$$b \to b^*, h \to h^*, \theta \to -\theta, \Lambda y \leftrightarrow \Lambda 1. \qquad (4.20)$$

Example 4.2: Companion canonical pair with power variance function

Suppose GLM $\mathcal{G}$ has power variance function, so that $b(\cdot)$ is given by (3.6), characterised by *p*, and *h* is the canonical link. Then it may be checked, using (4.7) and (4.13), that in parallel with (4.13)

$$b_p^* = b_{p^*} \text{ with } p^* = 3 - p. \qquad (4.21)$$

Example 4.3: Normal error term. Consider the GLM $\mathcal{G}$ with normal error term and canonical link. This is the case $p = 0$. By (4.21) and (4.13), the companion GLM $\mathcal{G}^*$ has cumulant function corresponding to $p^* = 3$, and link corresponding to $p^* = -1$. For this case, $b(\theta) = \tfrac{1}{2}\theta^2, b^*(\theta) = -(-2\theta)^{\frac{1}{2}}, h(\mu) = \mu, h^*(\mu) = \tfrac{1}{2}\mu^2$. The companion likelihood is inverse Gaussian The likelihoods are given by (4.6) and (4.14):

$$L(y;\beta,\Lambda) = y^T \Lambda X\beta - \tfrac{1}{2} 1^T \Lambda (X\beta)^2 \qquad (4.22)$$

$$L^*(y;-\beta,\Lambda) = 1^T \Lambda X\beta + y^T \Lambda (-2X\beta)^{\frac{1}{2}}. \qquad (4.23)$$

Example 4.4: Gamma error term. Consider the GLM $\mathcal{G}$ with gamma error term and canonical link. This is the case $p = 2$. By (4.21) and (4.13), the companion GLM $\mathcal{G}^*$ has both cumulant function and link corresponding to $p^* = 1$, which is the Poisson canonical link. For this case $b(\theta) = -\log(-\theta), b^*(\theta) = \exp\theta, h(\mu) = -1/\mu, \quad h^*(\mu) = \log\mu$. The likelihoods are:

$$L(y;\beta,\Lambda) = y^T \Lambda X\beta + 1^T \Lambda \log(-X\beta) \qquad (4.24)$$

$$L^*(y;-\beta,\Lambda) = 1^T \Lambda X\beta - y^T \Lambda \exp(-X\beta). \qquad (4.25)$$

5. Bayesian revision of a GLM

5.1 General development

Consider the GLM $\mathcal{G}(b,h,\Lambda,X)$ with log-likelihood (4.2). Because of (4.3), it would be possible to express $\mathcal{G}$ in terms of β rather than θ. Equivalently, express $\mathcal{G}$ in terms of $t(\beta)$, where the function *t* is defined as

$$t = (b')^{-1} \circ h^{-1} \circ M \qquad (5.1)$$

M being an as yet unspecified q x q non-singular matrix.

It follows that

$$\beta = \left(M^{-1} \circ h \circ b'\right)(t) \qquad (5.2)$$

where $t(\beta)$ is denoted just t, and so, by (4.3),

$$\theta = (g \circ t)(\beta) \qquad (5.3)$$

with

$$g = (b')^{-1} \circ h^{-1} \circ XM^{-1} \circ h \circ b'. \qquad (5.4)$$

Then the log-likelihood (4.2) may be expressed in the form

$$L(y;\beta,\Lambda) = y^T \Lambda (g \circ t)(\beta) - 1^T \Lambda (b \circ g \circ t)(\beta). \qquad (5.5)$$

As a prior log-likelihood, select

$$\pi(\beta) = \underset{1 \text{x} q}{w_0^T} \; \underset{q \text{x} 1}{t(\beta)} - \underset{1 \text{x} q}{n_0^T} \; \underset{q \text{x} 1}{(b \circ t)(\beta)}. \qquad (5.6)$$

Appendix B gives a simple proof that $E_\beta\left[b'(t_j)\right] = w_{0j} / n_{0j}$.

Therefore

$$E_\beta\left[b'(t)\right] = N_0^{-1} w_0$$

where $N_0 = DIAG\, n_0$.

By (5.1), this may be written as

$$E_\beta\left[h^{-1}(M\beta)\right] = N_0^{-1} w_0. \qquad (5.7)$$

Note that, by (5.1)

$$t\left(M^{-1}\beta\right) = \left[(b')^{-1} \circ h^{-1}\right](\beta). \qquad (5.8)$$

Then

$$\pi\left(M^{-1}\beta\right) = \sum_{j=1}^{q} \left\{ w_{0j} \left[(b')^{-1} \circ h^{-1}\right](\beta_j) - n_{0j} \left[b \circ (b')^{-1} \circ h^{-1}\right](\beta_j) \right\}$$

and the prior consists of **independent** priors on the separate components $\beta_1,...,\beta_q$. The matrix M may therefore be used to introduce a prior correlation structure on β.

The posterior log-likelihood implied by (5.5) and (5.6) is

$$L(\beta \mid y,\Lambda) = \left[w_0^T t(\beta) + y^T \Lambda (g \circ t)(\beta)\right] - \left[n_0^T (b \circ t)(\beta) + 1^T \Lambda (b \circ g \circ t)(\beta)\right]. \tag{5.9}$$

In choosing a prior $\pi(\beta)$, one might have considered defining $t(\beta)$ as $M(b')^{-1} \circ h^{-1}(\beta)$, instead of as in (5.1). However, this definition would, in general, produce components of $t(\beta)$ of differing signs, whereas the domain of b will often be of only one sign (see e.g. (3.8)). The definition (5.1) achieves this constancy of sign in the same way as expression (4.3) achieves it for θ, also an argument of b (see (3.1)).

Example 5.1: Univariate case. Consider the case $m = q = 1$, so that $X = M = 1$ without loss of generality. Then, by (5.4), $g = 1$, and comparison of (5.9) with (5.6) shows that the latter family is closed under Bayesian revision with

$$(w_0, n_0) \mapsto (w_0 + \Lambda y, n_0 + \Lambda).$$

This is a well known result for the EDF (Landsman and Makov, 1998; Nelder and Verrall, 1997).

□

5.2 Canonical link

The above likelihoods simplify somewhat in the case of a canonical link. Then

$$t(\beta) = M\beta, \qquad g = XM^{-1}. \tag{5.10}$$

The log-likelihoods (5.5), (5.6) and (5.9) become

$$L(y;\beta,\Lambda) = y^T \Lambda X\beta - 1^T \Lambda b(X\beta) \tag{5.11}$$

$$\pi(\beta) = w_0^T M\beta - n_0^T b(M\beta) \tag{5.12}$$

$$L(\beta \mid y,\Lambda) = \left(w_0^T M + y^T \Lambda X\right)\beta - \left[n_0^T b(M\beta) + 1^T \Lambda b(X\beta)\right]. \tag{5.13}$$

5.3 Companion canonical link

Similar simplifications occur in the case of the companion canonical link (4.10). Substitution of this in (5.1) and (5.4) yields

$$t(\beta) = b^{-1}(M\beta), \qquad g = b^{-1} \circ XM^{-1} \circ b. \tag{5.14}$$

The log-likelihoods (5.5), (5.6) and (5.9) become

$$L\left(y;\beta,\Lambda\right)=y^T\Lambda b^{-1}\left(X\beta\right)-1^T\Lambda X\beta \tag{5.15}$$

$$\pi\left(\beta\right)=w_0^T b^{-1}\left(M\beta\right)-n_0^T M\beta \tag{5.16}$$

$$L\left(\beta\,|\,y,\Lambda\right)=-\left(n_0^T M+1^T\Lambda X\right)\beta+\left[w_0^T b^{-1}\left(M\beta\right)+y^T\Lambda b^{-1}\left(X\beta\right)\right]. \tag{5.17}$$

5.4 Companion GLM

Consider the GLM $\mathcal{G}\left(b,h,\Lambda,X\right)$ and its companion $\mathcal{G}^*$ with log-likelihood (4.14). In parallel with (5.1) and (5.4), define

$$t^*=\left(b'\right)^{-1}\circ\left(h^*\right)^{-1}\circ M \tag{5.18}$$

$$g^*=\left(b'\right)^{-1}\circ\left(h^*\right)^{-1}\circ XM^{-1}\circ h^*\circ b'. \tag{5.19}$$

Then, in parallel with (5.5),

$$L^*\left(y;\beta,\Lambda\right)=y^T\Lambda\left(g^*\circ t^*\right)\left(\beta\right)-1^T\Lambda\left(b\circ g^*\circ t^*\right)\left(\beta\right). \tag{5.20}$$

This result can also be checked by noting that, by definition (4.8),

$$\left(b'\right)^{-1}\circ\left(h^*\right)^{-1}=b^{-1}\circ\left(b'\right)^{-1}\circ h^{-1} \tag{5.21}$$

and using this to show that

$$g^*\circ t^*=b^{-1}\circ g\circ t \tag{5.22}$$

$$\theta=\left(b\circ g^*\circ t^*\right)\left(\beta\right)\quad\text{[by (5.3)]} \tag{5.23}$$

which may be substituted in (4.15) to yield (5.20).

5.5 Second order approximation

5.5.1 Revision of likelihood

The objective here is to approximate the revised log-likelihood (5.9) by an expression of the form

$$w_1^T Pt\left(\beta\right)-n_1^T b\circ Pt\left(\beta\right) \tag{5.24}$$

where P is a linear transformation; w_1, n_1 are vectors corresponding to n_0, w_0; and the approximation is accurate to second order in $t\left(\beta\right)$ about $t_0=t\left(\beta_0\right)$ where β_0 is an arbitrarily selected value of β.

This is done by taking the Taylor expansion of each member of (5.9). The results, from Appendix A, are substituted in (5.9), yielding

$$\begin{aligned} L(\beta \mid y,\Lambda) = (t-t_0)^T a - \tfrac{1}{2}(t-t_0)^T \Big\{ N_0 B_0 + (Dg)^T G\Lambda(Dg) \\ - \Big[\Lambda\big(y - h^{-1}(X\beta_0)\big) \otimes \underset{q\times q}{1}\Big]^T (D^2 g)\Big\}(t-t_0) \end{aligned} \tag{5.25}$$

where 0-th order terms in *t* (which do not depend on β) have been omitted, and

$$a = \big[w_0 - N_0 h^{-1}(M\beta_0)\big] + (Dg)^T \Lambda\big[y - h^{-1}(X\beta_0)\big] \tag{5.26}$$

$$B_0 = DIAG\, b'' \circ (b')^{-1} \circ h^{-1}(M\beta_0) \tag{5.27}$$

$$H_{(-1)} = DIAG\, h^{-1}(X\beta_0) \tag{5.28}$$

$$G = DIAG\big(b'' \circ g(t_0)\big) = DIAG\big[b'' \circ (b')^{-1} \circ h^{-1}(X\beta_0)\big] \tag{5.29}$$

and $D^2 g$ is as defined in Section 2. These quantities, as well as b', b'', are evaluated at $t = t_0$.

At this point, it is convenient to fix the arbitrary parameter β_0 by setting

$$h^{-1}(M\beta_0) = E_\beta\big[h^{-1}(M\beta)\big]. \tag{5.30}$$

Then (5.7) reduces the first member of *a* in (5.26) to zero, giving

$$a = (Dg)^T \Lambda\big[y - h^{-1}(X\beta_0)\big]. \tag{5.31}$$

Now diagonalise the matrix within braces in (5.25):

$$P^T D P = N_0 B_0 + (Dg)^T G\Lambda(Dg) - \Big[\Lambda\big(y - h^{-1}(X\beta_0)\big) \otimes \underset{q\times q}{1}\Big]^T (D^2 g) \tag{5.32}$$

where *D* is diagonal and *P* orthogonal.

Then

$$\begin{aligned} L(\beta \mid y,\Lambda) &= (t-t_0)^T a - \tfrac{1}{2}(t-t_0)^T P^T D P (t-t_0) \\ &= (u-u_0)^T P a - \tfrac{1}{2}(u-u_0)^T D (u-u_0) \end{aligned} \tag{5.33}$$

where

$$u = Pt, \quad u_0 = Pt_0. \tag{5.34}$$

Define

$$N_1 = D\left[DIAG\, b''(u_0)\right]^{-1}. \tag{5.35}$$

Then (5.33) can be rewritten as

$$\begin{aligned} L(\beta \mid y, \Lambda) &= (u-u_0)^T \left\{Pa + \left[DIAG\, b'(u_0)\right]n_1\right\} \\ &\quad -(u-u_0)^T \left[DIAG\, b'(u_0)\right]n_1 - \tfrac{1}{2}(u-u_0)^T N_1 \left[DIAG\, b''(u_0)\right](u-u_0) \\ &= w_1^T u - n_1^T b(u) \end{aligned} \tag{5.36}$$

where

$$\begin{aligned} w_1 &= Pa + N_1 b'(u_0) \\ &= P(Dg)^T \Lambda\left[y - h^{-1}(X\beta_0)\right] + N_1 b'(u_0) \end{aligned} \tag{5.37}$$

$$n_1 = VEC\, N_1 \tag{5.38}$$

$b(u)$ has been substituted for its second order Taylor expansion, and the term $w_1^T u_0$ has been omitted.

Thus Bayesian revision of the prior $\pi(\beta)$ in (5.6) is effected, to second order accuracy, by the substitutions

$$w_0 \to w_1, \qquad n_0 \to n_1, \qquad t(\beta) \to Pt(\beta). \tag{5.39}$$

Note that, by (5.10), in the case of the canonical link the last of these substitutions is the same as $M \to PM$.

The second order Bayesian revision of (5.6) thus maintains that parametric form subject to a rotation of the parameter vector, and the revision (5.6) $\to$ (5.36) is a recursion.

This is not so for links other than canonical because $Pt(\beta) = P \circ (b')^{-1} \circ h^{-1}(M\beta)$, which is not of the form $(b')^{-1} \circ h^{-1}(Q\beta)$ for some Q.

The revision is second order in the sense that the function $b(\cdot)$ appearing in the prior is reproduced in the posterior up to a second order term involving $b''(\cdot)$. In view of (3.3), and the fact that the variance in (3.3) varies inversely with sample size (through λ), posterior estimators must be **second order Bayes** in the sense of Levit (1980).

These have been applied in an actuarial context by Landsman (2002, 2004, and other papers referred to there). Landsman (2004) in particular gives some examples of second order Bayes estimators that return highly accurate results in the single-dimensional case.

Note also that, while second order approximation here amounts to matching the first two orders of the log-likelihood, it is **not** a normal approximation. This can be seen in (5.36), which is not normal.

5.5.2 Revision of mean and covariance

Recall the prior (5.6), expressed in terms of the parameter $t = t(\beta)$:

$$\pi(t) = w_0^T t - n_0^T b(t). \tag{5.40}$$

By (5.1),

$$b'(t) = h^{-1}(M\beta). \tag{5.41}$$

The mean and covariance of $b'(t)$ are given by (5.7) and (C.8):

$$E_\beta\left[h^{-1}(M\beta)\right] = E_\beta\left[b'(t)\right] = N_0^{-1} w_0 \tag{5.42}$$

$$Var_\beta\left[h^{-1}(M\beta)\right] = \left(N_0 - A_0''\right)^{-1} A_0 \tag{5.43}$$

where

$$A_0 = DIAG\ \alpha(v_0) \tag{5.44}$$

$$A_0'' = DIAG\ \alpha''(v_0) \tag{5.45}$$

$$v_0 = h^{-1}(M\beta_0) \tag{5.46}$$

$$\alpha = b'' \circ (b')^{-1} \tag{5.47}$$

Appendix C also shows that

$$E[Y \mid \beta] = h^{-1}(X\beta) \tag{5.48}$$

and so unconditional moments of Y are given by $E[Y] = E_\beta\left[h^{-1}(X\beta)\right]$ and $V_\beta\left[h^{-1}(X\beta)\right]$. The quantities are calculated in Appendices C.2 and C.3, where $h^{-1}(X\beta)$ is written in the form

$$h^{-1}(X\beta) = \phi(v) \tag{5.49}$$

with

$$\phi = h^{-1} \circ XM^{-1} \circ h. \tag{5.50}$$

It is shown that

$$E[Y_i] = \phi_i(v_0) + \tfrac{1}{2}\left\{H'_{(-1)} X M^{-1} H''\right\}_i (N_0 - A''_0)^{-1} A_0 1 + \tfrac{1}{2}\left\{\left[H''_{(-1)}\right]^{\frac{1}{2}} X M^{-1} H'\right\}_i (N_0 - A''_0)^{-1} A_0 \left\{\left[H''_{(-1)}\right]^{\frac{1}{2}} X M^{-1} H'\right\}_i^T \tag{5.51}$$

$$Var[Y] = \left\{H'_{(-1)} X M^{-1} H'\right\} (N_0 - A''_0)^{-1} A_0 \left\{H'_{(-1)} X M^{-1} H'\right\}^T \tag{5.52}$$

where Y_i, φ_i are the *i*-th components of Y, φ and

$$\nu_0 = E\left[h^{-1}(M\beta)\right] = N_0^{-1} w_0 \qquad \text{[by (5.42)]} \tag{5.53}$$

$$H' = DIAG\, h'(\nu_0) \tag{5.54}$$

$$H'' = DIAG\, h''(\nu_0) \tag{5.55}$$

$$H'_{(-1)} = DIAG\left(h^{-1}\right)'(\eta_0) \tag{5.56}$$

$$H''_{(-1)} = DIAG\left(h^{-1}\right)''(\eta_0) \tag{5.57}$$

$$\eta_0 = X\beta_0 \tag{5.58}$$

and $\{...\}_i$ denotes the *i*-th row of the matrix argument and $\{...\}_i^T$ denotes the transpose of that row.

For the case of a canonical link, the Bayesian revisions of $E[Y]$ and $Var[Y]$ are made by substitution of (5.39) in (5.51) and (5.52), ie

$$w_0 \to w_1, \qquad n_0 \to n_1, \qquad M \to PM. \tag{5.59}$$

The case of other links involves a considerably more complex transformation than $M \to PM$, and is not pursued further here.

Canonical link

Considerable further simplification occurs in the case of canonical link $h = (b')^{-1}$. Specifically,

$$t(\beta) = M\beta \quad \text{[by (5.1)]} \tag{5.60}$$

$$g(t) = XM^{-1}t \quad \text{[by (5.4)]} \tag{5.61}$$

$$\left(Dg\right)\left(t\right) = XM^{-1},\ \left(D^2 g\right)\left(t\right) = 0 \tag{5.62}$$

$$u = PM\beta,\ u_0 = PM\beta_0 \quad \text{[by (5.34)]} \tag{5.63}$$

$$G = DIAG\, b''\left(X\beta_0\right) \quad \text{[by (5.29)]} \tag{5.64}$$

$$B_0 = DIAG\, b''\left(M\beta_0\right) \quad \text{[by (5.27)]} \tag{5.65}$$

The parameter β_0 is defined by (5.30) and (5.42):

$$\beta_0 = M^{-1} h\left(N_0^{-1} w_0\right).$$

Then, by (5.32) and (5.62),

$$P^T DP = N_0 B_0 + \left(Dg\right)^T G\Lambda\left(Dg\right). \tag{5.66}$$

Note that $G\Lambda\left(= \Lambda G\right)$ is diagonal, and so (5.66) becomes

$$P^T DP = N_0 B_0 + \left(XM^{-1}\right)^T \Lambda G\left(XM^{-1}\right). \tag{5.67}$$

Now the prior mean of $h^{-1}\left(M\beta\right)$ is given by (5.42), and the posterior mean by the same expression with N_0, w_0 replaced by N_1, w_1.

Thus

$$\begin{aligned} E\left[h^{-1}\left(PM\beta\right) \mid y\right] &= N_1^{-1} w_1 \\ &= N_1^{-1} P\left(XM^{-1}\right)^T \Lambda\left[y - h^{-1}\left(X\beta_0\right)\right] + b'\left(PM\beta_0\right) \end{aligned} \tag{5.68}$$

by (5.37) and (5.62).

By (5.35) and (5.67),

$$\begin{aligned} N_1^{-1} P &= \left[DIAG\, b''\left(PM\beta_0\right)\right] D^{-1} P \\ &= \left[DIAG\, b''\left(PM\beta_0\right)\right] P\left[N_0 B_0 + \left(XM^{-1}\right)^T \Lambda G\left(XM^{-1}\right)\right]^{-1} \end{aligned} \tag{5.69}$$

whence

$$E\left[h^{-1}\left(PM\beta\right) \mid y\right] = b'\left(PM\beta_0\right) + PJ\left[y - h^{-1}\left(X\beta_0\right)\right] \tag{5.70}$$

with

$$J = \left(P^T B_1 P\right)\left[N_0 B_0 + \left(XM^{-1}\right)^T \Lambda G\left(XM^{-1}\right)\right]^{-1}\left(XM^{-1}\right)^T \Lambda \tag{5.71}$$

and

$$B_1 = DIAG\, b''\left(PM\beta_0\right). \tag{5.72}$$

Also, $Var\left[h^{-1}\left(PM\beta\right) \mid y\right]$ is given by (5.43) – (5.47) with N_0 replaced by N_1 and *M* by *PM*:

$$Var\left[h^{-1}\left(PM\beta\right) \mid y\right] = \left(N_1 - A_1''\right)^{-1} A_1 \tag{5.73}$$

with

$$A_1 = DIAG\, \alpha\left(h^{-1}\left(PM\beta_0\right)\right) \tag{5.74}$$

$$A_1'' = DIAG\, \alpha''\left(h^{-1}\left(PM\beta_0\right)\right). \tag{5.75}$$

Example 5.2: normal error, identity link. This is the case

$$b\left(\theta\right) = \tfrac{1}{2}\theta^2 \tag{5.76}$$

$$h\left(\nu\right) = \nu. \tag{5.77}$$

The identity link is canonical, and so the immediately preceding results may be applied. With $b(.)$ given by (5.76), those results give:

$$G = 1, \quad B_0 = 1, \; B_1 = 1 \tag{5.78}$$

$$\alpha\left(\nu\right) = 1, \qquad \alpha''\left(\nu\right) = 0 \tag{5.79}$$

and then (5.70) reduces to

$$E\left[\left(PM\beta\right) \mid y\right] = \left(PM\beta_0\right) + PJ\left[y - X\beta_0\right] \tag{5.80}$$

where

$$J = \left[N_0 + \left(XM^{-1}\right)^T \Lambda\left(XM^{-1}\right)\right]^{-1}\left(XM^{-1}\right)^T \Lambda \tag{5.81}$$

and, by (5.30) and (5.42),

$$\beta_0 = E\left[\beta\right] = M^{-1} N_0^{-1} w_0. \tag{5.82}$$

Then

$$E[\beta \mid y] = M^{-1} P^T E[PM\beta \mid y] \\ = \beta_0 + K(y - X\beta_0) \tag{5.83}$$

where

$$K = M^{-1} J = M^{-1} N_0^{-1} \left[1 + \left(M^{-1}\right)^T X^T \Lambda X M^{-1} N_0^{-1}\right]^{-1} \left(M^{-1}\right)^T X^T \Lambda \\ = \Gamma X^T \Lambda \left(1 + X \Gamma X^T \Lambda\right)^{-1} \tag{5.84}$$

by (2.9) and where $\Gamma = M^{-1} N_0^{-1} \left(M^{-1}\right)^T$.

By (5.73) and (5.79),

$$Var[PM\beta \mid y] = N_1^{-1}. \tag{5.85}$$

Then, by (5.69),

$$Var[\beta \mid y] = M^{-1} \left[N_0 + \left(XM^{-1}\right)^T \Lambda \left(XM^{-1}\right)\right]^{-1} \left(M^{-1}\right)^T \\ = \left(\Gamma^{-1} + X^T \Lambda X\right)^{-1} \\ = \left(1 + \Gamma X^T \Lambda X\right)^{-1} \Gamma.$$

But note that

$$1 - KX = 1 - \Gamma X^T \Lambda \left(1 + X \Gamma X^T \Lambda\right)^{-1} X \\ = \left(1 + \Gamma X^T \Lambda X\right)^{-1},$$

by (2.10).

Thus

$$Var[\beta \mid y] = (1 - KX)\Gamma. \tag{5.86}$$

By (5.43) the prior variance of $M\beta$ was N_0^{-1}, and so the prior variance of β was Γ. Thus, the Bayesian revision has "contracted" this variance by a factor of 1-*KX*, where *K* is the weight given to the data in the revision (5.83) of the mean.

Example 5.3: gamma error, reciprocal link. This is the case

$$b(\theta) = -\log(-\theta) \tag{5.87}$$

$$h(\nu) = -1/\nu. \tag{5.88}$$

The link is canonical, and so the results preceding Example 5.2 may be applied. They give:

$$G = \left[DIAG\, X\beta_0\right]^{-2}, \qquad B_0 = \left[DIAG\, M\beta_0\right]^{-2}, \qquad B_1 = \left[DIAG\, PM\beta_0\right]^{-2} \tag{5.89}$$

$$\alpha(\nu) = \nu^2 \qquad \alpha''(\nu) = 2. \tag{5.90}$$

Then (5.70) gives

$$E\left[1/PM\beta \mid y\right] = 1/PM\beta_0 - PJ\left(y + 1/X\beta_0\right) \tag{5.91}$$

where P is the orthogonal matrix given by (5.67), and

$$\beta_0 = -M^{-1}N_0\left(1/w_0\right).$$

The results (5.50) to (5.59) are then applied to obtain the Bayesian revisions of $E[Y]$ and $Var[Y]$. These results give

$$\phi(\nu) = 1/XM^{-1}\left(1/\nu\right) \tag{5.92}$$

$$H' = \left[DIAG\,\nu_1\right]^{-2} \qquad H'' = -2\left[DIAG\,\nu_1\right]^{-3} \tag{5.93}$$

$$H'_{(-1)} = \left[DIAG\,\eta_1\right]^{-2} \qquad H''_{(-1)} = -2\left[DIAG\,\eta_1\right]^{-3} \tag{5.94}$$

$$\nu_1 = -E\left[1/PM\beta \mid y\right], \quad \eta_1 = XM^{-1}P^T\left(-1/\nu_1\right), \tag{5.95}$$

where the last relation has been obtained with the help of (5.46).

5.5.3 Companion GLM

For given GLM $\mathcal{G}$, the transformation to the companion GLM $\mathcal{G} \to \mathcal{G}^*$ is, according to (4.16), given by $h \to h^*$, with h^* defined by (4.8). The log-likelihood L^* of $\mathcal{G}^*$ is

$$L^*\left(y;\beta,\Lambda\right) = -1^T\Lambda\theta + y^T\Lambda b^{-1}(\theta). \tag{4.15}$$

Define t, g by (5.1) and (5.4) as for $\mathcal{G}$. Note that t, g are still defined in terms of h, not h^*.

Then (4.15) may be re-written in the form

$$L^*\left(y;\beta,\Lambda\right) = -1^T\Lambda\left(g\circ t\right)(\beta) + y^T\Lambda\left(b^{-1}\circ g\circ t\right)(\beta). \tag{5.96}$$

Choose the prior

$$\pi^*(\beta) = -w_0^{*T} t(\beta) + n_0^{*T}\left(b^{-1} \circ t\right)(\beta). \tag{5.97}$$

Comparison of (5.96) and (5.97) for $\mathcal{G}^*$ with their counterparts (5.5) and (5.6) for $\mathcal{G}$ shows that the transformation $\mathcal{G} \to \mathcal{G}^*$ and the transformation of the associated prior are given by:

$$b \to b^{-1}, \qquad \Lambda y \leftrightarrow -\Lambda 1. \tag{5.98}$$

The companion of the posterior log-likelihood (5.9) is then

$$\begin{aligned} L^*(\beta \mid y, \Lambda) = &-\left[w_0^{*T} t(\beta) + 1^T \Lambda (g \circ t)(\beta)\right] \\ &+\left[n_0^{*T}\left(b^{-1} \circ t\right)(\beta) + y^T \Lambda \left(b^{-1} \circ g \circ t\right)(\beta)\right]. \end{aligned} \tag{5.99}$$

Now (5.9) was approximated by (5.25) using the results of Appendix A, and (5.25) was approximated by (5.36). Appendix E provides the parallel development for $L^*(\beta \mid y, \Lambda)$, leading to the following result corresponding to (5.36):

$$L^*(\beta \mid y, \Lambda) = -w_1^{*T} u^* + n_1^{*T} b^{-1}\left(u^*\right) \tag{5.100}$$

where

$$w_1^* = -P^*(Dg)^T \Lambda\left[\left(H_{(-1)}^*\right)^{-1} y - 1\right] + N_1^*\left[1/\left(b' \circ b^{-1}\right)(u_0)\right] \tag{5.101}$$

$$n_1^* = VEC\, N_1^* \tag{5.102}$$

$$u^* = P^* t, \qquad u_0 = P^* t_0 \tag{5.103}$$

$$H_{(-1)}^* = DIAG\left(h^*\right)^{-1}\left(X\beta_0\right) \tag{5.104}$$

$$N_1^* = D^*\left[DIAG\left(b^{-1}\right)''\left(u_0\right)\right]^{-1} \tag{5.105}$$

P^*, D^* are orthogonal and diagonal matrices respectively such that

$$\begin{aligned} P^{*T} D^* P^* = &N_0^* B_0^* + (Dg)^T \left(DIAG\ G^* \Lambda y\right)(Dg) \\ &+\left[\Lambda\left[\left(H^*_{(-1)}\right)^{-1} y - 1\right] \otimes 1\right]^T \left(D^2 g\right) \end{aligned} \tag{5.106}$$

$$N_0^* = DIAG\, n_0^* \tag{5.107}$$

$$B_0^* = DIAG\left(b^{-1}\right)'' \circ \left(b'\right)^{-1} \circ h^{-1}\left(M\beta_0\right) \tag{5.108}$$

$$G^* = DIAG\left[\left(b^{-1}\right)'' \circ \left(b'\right)^{-1} \circ h^{-1}\left(X\beta_0\right)\right] \tag{5.109}$$

and β_0 is chosen to satisfy

$$1/\left(h^*\right)^{-1}\left(M\beta_0\right) = \left(N_0^*\right)^{-1} w_0. \tag{5.110}$$

Thus, Bayesian revision of the prior (5.97) is effected, to second order accuracy, by the substitutions

$$w_0^* \to w_1^*, \quad n_0^* \to n_1^*, \quad t(\beta) \to P^* t(\beta). \tag{5.111}$$

Companion canonical link

Recall that the companion canonical link is given by (4.8) with $h = \left(b'\right)^{-1}$, which is the canonical link. In this case

$$t(\beta) = M\beta \tag{5.112}$$

just as in (5.10), and so the last substitution in (5.111) reduces to $M \to P^* M$, in parallel with the case of the canonical link (see just after (5.39)). Relations (5.61) and (5.62) continue to hold.

This yields the following result corresponding to (5.70):

$$E\left[1/\left(h^*\right)^{-1}\left(P^* M\beta\right) \mid y\right] = \left(b^{-1}\right)'\left(P^* M\beta_0\right) - P^* J^*\left(y - \hat{y}\right)$$

where

$$J^* = P^{*T} B_1^* P^* \left\{N_0^* B_0^* + \left(XM^{-1}\right)^T \left[DIAG\, G^* \Lambda y\right]\left(XM^{-1}\right)\right\}\left(XM^{-1}\right)^T \Lambda \left(H_{(-1)}^*\right)^{-1}$$

$$\hat{y} = H_{(-1)}^* 1 = \left(h^*\right)^{-1}\left(X\beta_0\right)$$

$$B_1^* = DIAG\left(b^{-1}\right)''\left(P^* M\beta_0\right)$$

Revision of mean and variance

This sub-section is developed in parallel with Section 5.5.2 but in relation to the companion GLM $\mathcal{G}^*$. This is done by replicating Appendix C with the appropriate replacements, as set out in Table 5.1.

Table 5.1
Correspondences for mean and variance of GLM and its companion

$\mathcal{G}$	$\mathcal{G}^*$
b	b^{-1}
$h^{-1}\left(=b'\right)$	$\left(h^*\right)^{-1}\left(=b'\circ b^{-1}=1/\left(b^{-1}\right)'\right)$
Λy	$-\Lambda 1$
$\Lambda 1$	$-\Lambda y$
w_0	$-w_0^*$
n_0	$-n_0^*$
$h^{-1}\left(X\beta\right)$	$\left(h^*\right)^{-1}\left(X\beta\right)$
$h^{-1}\left(M\beta\right)$	$1/\left(h^*\right)^{-1}\left(M\beta\right)$

The first three rows of the table simply repeat (5.98). The next two are derived from a comparison of (5.97) with (5.6). The last two require some comment.

Equation (5.48) relates the mean of a GLM to its linear response through the inverse of its link. If the link changes from h to h^*, but no other aspect of the GLM changes, as is the case in the substitution $\mathcal{G}\rightarrow\mathcal{G}^*$, then h is simply replaced by h^* in (5.48). This is the case whenever h^{-1} operates on $X\beta$.

On the other hand, when h^{-1} operates on $M\beta$, it is being used to express moments of the prior π^* and, as shown in (E.14), it is replaced by $1/\left(h^*\right)^{-1}$.

As an example of the application of Table 5.1, the counterpart of (5.43) for the companion GLM is

$$Var_\beta\left[1/\left(h^*\right)^{-1}\left(M\beta\right)\right]=-\left[N_0^*+\left(A_0^*\right)''\right]^{-1}A^*_0 \tag{5.113}$$

where

$$A_0^*=DIAG\,\alpha^*\left(\nu_0^*\right) \tag{5.114}$$

$$\left(A_0^*\right)''=DIAG\left(\alpha^*\right)''\left(\nu_0^*\right) \tag{5.115}$$

$$\nu_0^*=1/\left(h^*\right)^{-1}\left(M\beta_0\right) \tag{5.116}$$

$$\alpha^*=\left(b^{-1}\right)''\circ\left(\left(b^{-1}\right)'\right)^{-1}. \tag{5.117}$$

Example 5.4: gamma error, log link. This is the case

$$b(\theta) = -\log(-\theta) \tag{5.118}$$

$$h(\nu) = \log \nu. \tag{5.119}$$

By (4.8), this link is companion to that in Example 5.3, and so henceforth will be written

$$h^*(\nu) = \log \nu \tag{5.120}$$

companion to

$$h(\nu) = -1/\nu. \tag{5.121}$$

The GLM of the present example is thus the companion of that in Example 5.3, and so the results of the present sub-section may be applied. They give:

$$G^* = -\left[DIAG \exp - X\beta_0\right], \qquad B_0^* = -\left[DIAG \exp - M\beta_0\right] \tag{5.122}$$

$$g(t) = XM^{-1}t,\ Dg = XM^{-1}, \quad D^2 g = 0 \tag{5.123}$$

$$N_1^* = -D^*\left[DIAG \exp P^* M\beta_0\right] \tag{5.124}$$

where D^* and P^* are obtained from the orthogonal decomposition

$$P^{*T} D^* P^* = N_0^* B_0^* + \left(XM^{-1}\right)^T \left[DIAG\ G^* \Lambda y\right]\left(XM^{-1}\right) \tag{5.125}$$

$$w_1^* = N_1^* \exp\left(-P^* M\beta_0\right) - P^*\left(XM^{-1}\right)^T \Lambda\left[\left[DIAG \exp - X\beta_0\right] y - 1\right] \tag{5.126}$$

The posterior version of (B.9) is

$$E\left[\exp - P^* M\beta \mid y, \Lambda\right] = \left(N_1^*\right)^{-1} w_1^* \tag{5.127}$$

$$= \left[DIAG \exp - P^* M\beta_0\right]\left[\underset{q\times 1}{1} - K\left(y - \hat{y}\right)\right] \tag{5.128}$$

with

$$K = \left(D^*\right)^{-1} P^* \left(XM^{-1}\right)^T \Lambda G^* \tag{5.129}$$

and

$$\hat{y} = \exp X\beta_0.$$

Expression (5.128) takes a neat form for the present special case of gamma error and log link. It may be desirable for some purposes, however, to express it in a form analogous to (5.70), as follows:

$$E\left[\exp - P^* M\beta \mid y, \Lambda\right] = \exp - P^* M\beta_0 - P^* J^* \left(y - \hat{y}\right)$$

where

$$\left[DIAG \exp - P^* M\beta_0\right] K = P^* J^*$$

ie

$$J^* = -P^{*T} B_1^* \left(D^*\right)^{-1} P^* \left(XM^{-1}\right)^T \Lambda G^*$$

with

$$B_1^* = -DIAG \exp - P^* M\beta_0.$$

To convert (5.128) to the posterior mean of $\exp X\beta$, define

$$\nu_1^* = E\left[\exp - P^* M\beta \mid y, \Lambda\right]. \qquad (5.130)$$

Define β_1 through the relation

$$1/\left(h^*\right)^{-1}\left(P^* M\beta_1\right) = E\left[1/\left(h^*\right)^{-1}\left(P^* M\beta\right) \mid y, \Lambda\right] \qquad \text{[c.f. (5.30)]}$$

ie

$$\beta_1 = -M^{-1} P^{*T} \log \nu_1^*. \qquad (5.131)$$

Also define

$$\eta_1^* = X\beta_1. \qquad (5.132)$$

By (5.130) and (5.132), ν_1^* and η_1^* serve the same purpose in the calculation of the posterior mean of $\exp X\beta$ as did ν_0 and η_0 in the calculation of its prior mean in Section 5.5.2; and similarly ν_0^* in the application of Table 5.1 immediately before the present example. Therefore, ν_0^* is replaced by ν_1^* there, giving

$$\alpha^*\left(\nu\right) = -\nu, \quad A_1^* = -DIAG \exp - M^* \beta_1, \quad \left(A_1^*\right)'' = 0, \qquad (5.133)$$

where A_1^* replaces A_0^*, and $M^* = P^* M$.

Table 5.1 then yields the following replacements in Example 5.3:

$$h^{-1}\left(\nu_1^*\right) \to 1/\left(h^*\right)^{-1}\left(\nu_1^*\right) = \exp-\nu_1^*, \quad h^{-1}\left(\eta_1^*\right) \to \left(h^*\right)^{-1}\left(\eta_1^*\right) = \exp\eta_1^* \qquad (5.134)$$

$$h\left(\nu_1^*\right) \to \log\nu_1^*, \quad h\left(\eta_1^*\right) \to \log\eta_1^* \qquad (5.135)$$

$$\begin{aligned} &H' \to DIAG\,1/\nu_1^* = DIAG\exp M^*\beta_1, \qquad H'' \to -DIAG\,1/\nu_1^{*2} = -DIAG\exp 2M^*\beta_1 \\ &H'_{(-1)} \to DIAG\exp\eta_1^* = DIAG\exp X\beta_1, \quad H''_{(-1)} \to DIAG\exp\eta_1^* = DIAG\exp X\beta_1 \end{aligned} \qquad (5.136)$$

$$H'_{(-1)}XM^{-1}H'' \to -\left[DIAG\exp X\beta_1\right]XM^{*-1}\left[DIAG\exp 2M^*\beta_1\right] \qquad (5.137)$$

$$\left[H''_{(-1)}\right]^{\frac{1}{2}} XM^{-1}H' \to \left[DIAG\exp\tfrac{1}{2}X\beta_1\right]XM^{*-1}\left[DIAG\exp M^*\beta_1\right]. \qquad (5.138)$$

Making the replacements (5.137) and (5.138) in (5.51), and recalling (5.133), gives

$$\begin{aligned} &E\left[\exp X_i^T\beta \mid y,\Lambda\right] = \phi_i\left(\nu_1^*\right) + \tfrac{1}{2}\left\{\left[DIAG\exp X\beta_1\right]XM^{*-1}\left[DIAG\exp M^*\beta_1\right]\right\}_i\left(n_1^*\right)^{-1} \\ &-\tfrac{1}{2}\left\{\left[DIAG\exp\tfrac{1}{2}X\beta_1\right]XM^{*-1}\left[DIAG\exp M^*\beta_1\right]\right\}_i\left(N_1^*\right)^{-1}\left[DIAG\exp-M^*\beta_1\right] \\ &\left\{\left[DIAG\exp\tfrac{1}{2}X\beta_1\right]XM^{*-1}\left[DIAG\exp M^*\beta_1\right]\right\}_i^T \end{aligned} \qquad (5.139)$$

where X_i^T is the *i*-th row of *X* and, by (5.50), (5.134) and (5.135),

$$\phi\left(\nu_1^*\right) = \exp-XM^{*-1}\log\nu_1^* = \exp X\beta_1 \qquad (5.140)$$

(by 5.131).

Similarly, by (5.52) with the appropriate replacements,

$$\begin{aligned} Var\left[\exp X\beta \mid y,\Lambda\right] &= -\left[DIAG\exp X\beta_1\right]XM^{*-1}\left[DIAG\exp M^*\beta_1\right] \\ &\quad \left(N_1^*\right)^{-1}\left(XM^{*-1}\right)^T\left[DIAG\exp X\beta_1\right] \\ &= \left[DIAG\exp X\beta_1\right]\left(XM^{*-1}\right)D^{*-1}\left(XM^{*-1}\right)^T \\ &\quad \left[DIAG\exp X\beta_1\right] \end{aligned} \qquad (5.141)$$

by (5.124).

6. Dynamic generalised linear models

6.1 Preliminary commentary

Section 5 shows how to obtain a second order approximation to the Bayesian revision of a GLM.

The form of the revision is given by (5.39). The passage immediately following that result points out that (5.39) becomes a recursion in the case of a canonical link.

A similar result is obtained for a companion canonical link (see (5.111) and (5.112) and associated text).

These recursions enable one to carry out multiple Bayesian revisions, retaining distributional forms and just varying parameters at each step. This leads to a filtering algorithm akin to the Kalman filter for producing a sequence of Bayesian revisions of a location parameter and its associated dispersion.

6.2 Canonical link

6.2.1 Framework

Consider the Bayesian framework described by (5.11) and (5.12), but generalise to the situation in which it applies to a sequence of epochs s = 1, 2, etc. Specifically,

$$\pi_{s|s-1}\left(\beta(s)\right) = w^T(s|s-1)M(s|s-1)\beta(s) - n^T(s|s-1)b\left(M(s|s-1)\beta(s)\right) \qquad (6.1)$$

$$L_s\left(y(s);\beta(s),\Lambda(s)\right) = y^T(s)\Lambda(s)X(s)\beta(s) - 1^T\Lambda(s)b\left(X(s)\beta(s)\right) \qquad (6.2)$$

where generally the notation $s|s-j$ indicates an estimate of a quantity current at epoch s but where the estimate takes account of data up to and including s-j.

As to the form of the matrices $M(s|s-1)$, note from (6.1) that the prior distribution of the vector $M(s|s-1)\beta(s)$ has independent components. One choice of $M(s|s-1)$ is therefore the orthogonal matrix that diagonalises $Var\left[\beta(s)|Y(s-1)\right]$ as follows:

$$Var\left[\beta(s)|Y(s-1)\right] = M^T(s|s-1)Q(s|s-1)M(s|s-1). \qquad (6.3)$$

where $Q(s|s-1)$ is diagonal, and $Y(s)$ denotes the total information $y(1),...,y(s)$.

Now (6.1) denotes the distribution for $\beta(s)$ taking account of data up to (s-1), and it is the prior that is revised by observations made at s. Note that $b(.)$ does not depend on s.

Equation (6.2) gives the log-likelihood for the GLM $\mathcal{G}(b,h,\Lambda(s),X(s))$ with canonical link $h=(b')^{-1}$. Suppose in addition that the parameter vector $\beta(.)$ evolves over time as follows:

$$\underset{q_s \times 1}{\beta(s+1)} = \underset{q_{s+1} \times q_s}{\chi(s+1)} \underset{q_s \times 1}{\beta(s)} + \underset{q_{s+1} \times 1}{\varepsilon(s+1)} \qquad (6.4)$$

where the matrix $\chi(s+1)$ is non-stochastic and $\varepsilon(s+1)$ is a centred stochastic quantity that is independent of $\beta(0),...,\beta(s),y(1),...,y(s)$ and with $Var[\varepsilon(s+1)]=R(s+1)$.

Thus it is assumed that the dimension of the parameter vector may change over time.

The objective is to evaluate $E\left[h^{-1}(M(s|s)\beta(s))|Y(s)\right]$ and $E\left[h^{-1}(M(s+1|s)\beta(s+1))|Y(s)\right]$ and the corresponding variances in terms of (6.1) and (6.2), where $M(s|s)$ will be defined shortly.

6.2.2 Estimation

With a slight abuse of notation, let $h^{-1}(s|s-j)$ denote $E\left[h^{-1}(M(s|s-j)\beta(s))|Y(s-j)\right]$. Let $\Gamma(s|s-j)$ denote $Var\left[h^{-1}(M(s|s-j)\beta(s))|Y(s-j)\right]$. The objective is then the evaluation of $h^{-1}(s|s)$ and $\Gamma(s|s)$, and then $h^{-1}(s+1|s)$ and $\Gamma(s+1|s)$.

Estimation of $h^{-1}(s|s)$ and $\Gamma(s|s)$

Consider the first of these. It may be expanded as follows:

$$\begin{aligned} h^{-1}(s|s) &= E\left[h^{-1}(M(s|s)\beta(s))|Y(s)\right] \\ &= E\left[E\left[h^{-1}(M(s|s)\beta(s))|Y(s-1)\right]|y(s)\right]. \end{aligned} \qquad (6.5)$$

Define

$$M(s|s)=P(s)M(s|s-1) \qquad (6.6)$$

where orthogonal $P(s)$ will be defined below. Note that the orthogonality of both matrices on the right side of (6.6) implies orthogonality of $M(s\,|\,s)$. Then (6.5) becomes:

$$h^{-1}(s\,|\,s)=E\left[E\left[h^{-1}\left(P(s)M(s\,|\,s-1)\beta(s)\right)|\,Y(s-1)\right]|\,y(s)\right]. \tag{6.7}$$

This is the Bayesian revision of $E\left[h^{-1}\left(P(s)M(s\,|\,s-1)\beta(s)\right)|\,Y(s-1)\right]$ to take account of data $y(s)$, and is given by (5.70) with the replacements $M\to M(s\,|\,s-1), P\to P(s), \beta\to\beta(s)$. The relevant quantities are defined below, with their corresponding equation numbers from Section 5 attached in square brackets.

Define $\beta(s\,|\,s-1)$ by the relations

$$\begin{aligned} h^{-1}\left(M(s\,|\,s-1)\beta(s\,|\,s-1)\right) &= E\left[h^{-1}\left(M(s\,|\,s-1)\beta(s)\right)|\,Y(s-1)\right] \quad [(5.30)] \\ &= N^{-1}(s\,|\,s-1)\,w(s\,|\,s-1) \quad [(5.7)] \end{aligned} \tag{6.8}$$

$$G(s)=DIAG\,b''\left(X(s)\beta(s\,|\,s-1)\right) \quad [(5.64)] \tag{6.9}$$

$$B(s\,|\,s-j)=DIAG\,b''\left(M(s\,|\,s-j)\beta(s\,|\,s-1)\right),\ j=0,1 \ \text{[(5.65) and (5.72)]} \tag{6.10}$$

Define $P(s)$ and $D(s)$ as the orthogonal and diagonal matrices respectively satisfying:

$$\begin{aligned} P^T(s)D(s)P(s) &= N(s\,|\,s-1)B(s\,|\,s-1) \\ &\quad +M(s\,|\,s-1)X^T(s)\Lambda(s)G(s)X(s)M^T(s\,|\,s-1) \quad [(5.67)] \end{aligned} \tag{6.11}$$

$$\begin{aligned} J(s) = {} & \left[P^T(s)B(s\,|\,s)P(s)\right]\left\{N(s\,|\,s-1)B(s\,|\,s-1)+\right. \\ & \left.M(s\,|\,s-1)X^T(s)\Lambda(s)G(s)X(s)M^T(s\,|\,s-1)\right\}^{-1} \\ & M(s\,|\,s-1)X^T(s)\Lambda(s) \quad [(5.71)] \end{aligned} \tag{6.12}$$

$$\begin{aligned} h^{-1}(s\,|\,s) &= h^{-1}\left(M(s\,|\,s)\beta(s\,|\,s-1)\right) \\ &\quad +P(s)J(s)\left[y(s)-h^{-1}\left(X(s)\beta(s\,|\,s-1)\right)\right] \quad [(5.70)] \end{aligned} \tag{6.13}$$

With $\alpha=b''\circ(b')^{-1}$ $[(5.47)]$

$$\begin{aligned} A(s) &= DIAG\,\alpha\left(h^{-1}\left(M(s\,|\,s)\beta(s\,|\,s-1)\right)\right) \quad [(5.74)] \\ A''(s) &= DIAG\,\alpha''\left(h^{-1}\left(M(s\,|\,s)\beta(s\,|\,s-1)\right)\right) \quad [(5.75)] \end{aligned} \tag{6.14}$$

$$N(s|s) = D(s)B^{-1}(s|s) \qquad [(5.65)] \tag{6.15}$$

$$\Gamma(s|s) = \left[N(s|s) - A''(s)\right]^{-1} A(s). \quad [(5.73)] \tag{6.16}$$

Estimation of $E\left[\beta(s+1)|Y(s)\right]$ and $Var\left[\beta(s+1)|Y(s)\right]$

By (6.4),

$$E\left[\beta(s+1)|Y(s)\right] = \chi(s+1)E\left[\beta(s)|Y(s)\right] \tag{6.17}$$

$$Var\left[\beta(s+1)|Y(s)\right] = \chi(s+1)Var\left[\beta(s)|Y(s)\right]\chi^T(s+1) + R(s+1). \tag{6.18}$$

To evaluate these quantities, express $\beta(s)$ in the form $\beta(s) = M^T(s|s)h\left[h^{-1}\left(M(s|s)\beta(s)\right)\right]$.

Take the Taylor expansions (2.14) and (2.15) of $h[...]$ about $E\left[h^{-1}\left(M(s|s)\beta(s)\right)|Y(s)\right]$ to obtain the second order approximations

$$E\left[\beta(s)|Y(s)\right] = M^T(s|s)\left\{h\left(h^{-1}(s|s)\right) + \tfrac{1}{2}\Gamma(s|s)H''(s|s)1\right\} \tag{6.19}$$

$$Var\left[\beta(s)|Y(s)\right] = M^T(s|s)H'(s|s)\Gamma(s|s)\left[H'(s|s)\right]^T M(s|s) \tag{6.20}$$

where it has been recognised from (6.16) that $\Gamma(s|s)$ is diagonal, and

$$H'(s|s) = DIAG\,h'\left(h^{-1}(s|s)\right), \quad H''(s|s) = DIAG\,h''\left(h^{-1}(s|s)\right). \tag{6.21}$$

Substitution of (6.19) and (6.20) into (6.17) and (6.18) gives

$$E\left[\beta(s+1)|Y(s)\right] = \chi(s+1)M^T(s|s)\left\{h\left(h^{-1}(s|s)\right) + \tfrac{1}{2}\Gamma(s|s)H''(s|s)1\right\} \tag{6.22}$$

$$\begin{aligned} Var\left[\beta(s+1)|Y(s)\right] = \chi(s+1)M^T(s|s)H'(s|s)\Gamma(s|s)\left[H'(s|s)\right]^T \\ M(s|s)\chi^T(s+1) + R(s+1). \end{aligned} \tag{6.23}$$

Estimation of $M(s+1|s)$

$M(s+1|s)$ is the orthogonal matrix that diagonalises $Var\left[\beta(s+1)|Y(s)\right]$, just as in (6.3):

$$Var\left[\beta(s+1)|Y(s)\right] = M^T(s+1|s)Q(s+1|s)M(s+1|s). \tag{6.24}$$

Estimation of $h^{-1}(s+1|s)$ and $\Gamma(s+1|s)$

Expand $h^{-1}(s+1|s)$ as a second order Taylor series:

$$h^{-1}(s+1|s) = E\left[h^{-1}\left(M(s+1|s)\beta(s+1)|Y(s)\right]\right] = h^{-1}\left(E\left[M(s+1|s)\beta(s+1)|Y(s)\right]\right) + \tfrac{1}{2}Var\left[M(s+1|s)\beta(s+1)|Y(s)\right]H''_{(-1)}(s+1|s)1 \tag{6.25}$$

where

$$H''_{(-1)}(s+1|s) = DIAG\left(h^{-1}\right)'' E\left[M(s+1|s)\beta(s+1)|Y(s)\right] \tag{6.26}$$

and advantage has been taken of the fact that, by (6.24), the variance matrix in (6.25) is diagonal.

Substitution of (6.24) into (6.25) yields

$$h^{-1}(s+1|s) = h^{-1}\left(E\left[M(s+1|s)\beta(s+1)|Y(s)\right]\right) + \tfrac{1}{2}Q(s+1|s)H''_{(-1)}(s+1|s)1 \tag{6.27}$$

where, by (6.22),

$$E\left[M(s+1|s)\beta(s+1)|Y(s)\right] = M(s+1|s)\chi(s+1)M^T(s|s) \left\{h\left(h^{-1}(s|s)\right) + \tfrac{1}{2}\Gamma(s|s)H''(s|s)1\right\}. \tag{6.28}$$

Similarly

$$\Gamma(s+1|s) = Var\left[h^{-1}\left(M(s+1|s)\beta(s+1)\right)|Y(s)\right] = H'_{(-1)}(s+1|s)Q(s+1|s)H'_{(-1)}(s+1|s) \tag{6.29}$$

where use has been made of (6.24) and

$$H'_{(-1)}(s+1|s) = DIAG\left(h^{-1}\right)' E\left[M(s+1|s)\beta(s+1)|Y(s)\right]. \tag{6.30}$$

Calculation of $N(s+1|s)$ and $w(s+1|s)$

By (6.8)

$$h^{-1}(s+1|s) = N^{-1}(s+1|s)w(s+1|s). \tag{6.31}$$

Recall (B.7), (5.60) and the fact that the canonical link $h = (b')^{-1}$:

$$\Gamma(s+1|s) = Var\left[h^{-1}\left(M(s+1|s)\beta(s+1)\right)|Y(s)\right]$$
$$= N^{-1}(s+1|s)\,DIAG\,E\left[\left(h^{-1}\right)'\left(M(s+1|s)\beta(s+1)\right)|Y(s)\right] \quad (6.32)$$

Take the second order Taylor series expansion of the expectation:

$$\Gamma(s+1|s) = N^{-1}(s+1|s)\left[H'_{(-1)}(s+1|s)+\tfrac{1}{2}Q(s+1|s)H'''_{(-1)}(s+1|s)\right] \quad (6.33)$$

by (6.24) and (6.30), and

$$H'''_{(-1)}(s+1|s) = DIAG\left(h^{-1}\right)'''\,E\left[\left(M(s+1|s)\beta(s+1)\right)|Y(s)\right]. \quad (6.34)$$

By (6.31) and (6.33),

$$N(s+1|s) = \Gamma^{-1}(s+1|s)\left[H'_{(-1)}(s+1|s)+\tfrac{1}{2}Q(s+1|s)H'''_{(-1)}(s+1|s)\right] \quad (6.35)$$

$$w(s+1|s) = N(s+1|s)h^{-1}(s+1|s). \quad (6.36)$$

Initiation of the estimates

The above estimation procedure provides a filter by which estimates of $\beta(\cdot)$ and their variances may be progressively revised in accordance with the data vectors $y(1), y(2)$, etc.

The filter is initiated with the prior estimates $E\left[\beta(1)\right]$ and $Var\left[\beta(1)\right]$. This yields $M(1|0)$ from (6.24), and then $h^{-1}(1|0)$ and $\Gamma(1|0)$ from (6.27) and (6.29).

The filter then proceeds through a sequence of iterations, each iteration revising $h^{-1}(s|s-1)$ and $\Gamma(s|s-1)$ to $h^{-1}(s+1|s)$ and $\Gamma(s+1|s)$, s = 1, 2, etc by means of the steps set out above, specifically:

(1) Calculation of $N(s|s-1)$ and $w(s|s-1)$ from (6.35) and (6.36);

(2) $h^{-1}(s|s-1)$, $\Gamma(s|s-1) \rightarrow h^{-1}(s|s), \Gamma(s|s)$ by (6.13) and (6.16);

(3) Estimation of $E\left[\beta(s+1)|Y(s)\right]$ and $Var\left[\beta(s+1)|Y(s)\right]$ from (6.22) and (6.23);

(4) Estimation of $M(s+1|s)$ from (6.24);

(5) Estimation of $h^{-1}(s+1|s)$ and $\Gamma(s+1|s)$ from (6.27) and (6.29).

Forecasts

Equation (6.27), together with substitution (6.28), provides one-step-ahead forecasts of $h^{-1}\left(M\left(s+1\,|\,s\right)\beta\left(s+1\right)\right)|\,Y\left(s\right)\right)$. Forecasts of $E\left[y\left(s+1\right)|\,Y\left(s\right)\right]=E\left[h^{-1}\left(X\left(s+1\right)\beta\left(s+1\right)\right)|\,Y\left(s\right)\right]$ may be obtained by the procedure set out in Section 5.5.2.

Equations (5.51) and (5.52) adapt simply to the present context to yield to the following:

$$\begin{aligned}E\left[y_i\left(s+1\right)|\,Y\left(s\right)\right]=\psi_i\left(s+1\,|\,s\right)&\left(h^{-1}\left(s+1\,|\,s\right)\right)\\&+\tfrac{1}{2}\left\{H'_{(-1)}\left(s+1\,|\,s\right)X\left(s+1\right)M^T\left(s+1\,|\,s\right)H''\left(s+1\,|\,s\right)\right\}_i\Gamma\left(s+1\,|\,s\right)1\\&+\tfrac{1}{2}\left\{\left[H''_{(-1)}\left(s+1\,|\,s\right)\right]^{\frac{1}{2}}X\left(s+1\right)M^T\left(s+1\,|\,s\right)H'\left(s+1\,|\,s\right)\right\}_i\Gamma\left(s+1\,|\,s\right)\left\{...\right\}_i^T\end{aligned} \qquad (6.37)$$

$$\begin{aligned}Var\left[y\left(s+1\right)|\,Y\left(s\right)\right]=&\left\{H'_{(-1)}\left(s+1\,|\,s\right)X\left(s+1\right)M^T\left(s+1\,|\,s\right)H'\left(s+1\,|\,s\right)\right\}\\&\Gamma\left(s+1\,|\,s\right)\left\{...\right\}^T\end{aligned} \qquad (6.38)$$

where

$$\phi\left(s+1\,|\,s\right)=h^{-1}\circ X\left(s+1\right)M^T\left(s+1\,|\,s\right)h \qquad (6.39)$$

and this time

$$H'_{(-1)}\left(s+1\,|\,s\right)=DIAG\left(h^{-1}\right)'\left(X\left(s+1\right)E\left[\beta\left(s+1\right)|\,Y\left(s\right)\right]\right) \qquad (6.40)$$

$$H''_{(-1)}\left(s+1\,|\,s\right)=DIAG\left(h^{-1}\right)''\left(X\left(s+1\right)E\left[\beta\left(s+1\right)|\,Y\left(s\right)\right]\right). \qquad (6.41)$$

Discussion

Consider (6.1), the prior associated with the GLM log-likelihood (6.2). It is applied for general *s*, which will be possible only if this family of priors is closed under Bayesian revision.

However, it was seen in Section 5.5.1 that this is not the case in general. That section was concerned with approximating the posterior likelihood with a member of the relevant family.

Thus, in the above sequence of iterations, a single iteration commences with $\pi_{s|s-1}$ approximated by a member of the family. Bayesian revision leads to $\pi_{s+1|s}$ which does not in general lie within that family. However, Step (1) of the procedure then replaces it by its second order approximation from the family.

6.3 Companion canonical link

Replace the Bayesian framework (6.1) and (6.2) by the companion form (see (5.96) and (5.97)):

$$\pi^*_{s|s-1}\left(\beta(s)\right) = -w^{*T}(s|s-1)M(s|s-1)\beta(s) + n^{*T}(s|s-1)b^{-1}\left(M(s|s-1)\beta(s)\right) \tag{6.42}$$

$$L^*_s\left(y(s);\beta(s),\Lambda(s)\right) = -1^T\Lambda(s)X(s)\beta(s) + y^T(s)\Lambda(s)b^{-1}\left(X(s)\beta(s)\right). \tag{6.43}$$

The correspondences between a GLM and its companion were identified in Table 5.1 for the purpose of calculating moments of transformed variates. Table 6.1 adapts these to the present context of iterative Bayesian revision, and then it may be used to translate all of the working of Section 6.2 to the companion case.

Table 6.1
Correspondences between GLM and its companion

$\mathcal{G}$	$\mathcal{G}^*$
b	b^{-1}
$h^{-1}(=b')$	$\left(h^*\right)^{-1}\left(=b'\circ b^{-1}=1/\left(b^{-1}\right)'\right)$
$\Lambda(s)y(s)$	$-\Lambda(s)1$
$\Lambda(s)1$	$-\Lambda(s)y(s)$
$w(s\|s-1)$	$-w^*(s\|s-1)$
$n(s\|s-1)$	$-n^*(s\|s-1)$
$h^{-1}\left(X(s)\beta(s)\right)$	$\left(h^*\right)^{-1}\left(X(s)\beta(s)\right)$
$h^{-1}\left(M(s\|s-j)\beta(s)\right)$	$1/\left(h^*\right)^{-1}\left(M^*(s\|s-j)\beta(s)\right)$

Most of these substitutions are straightforward, but a few call for comment. For example, $y(s)$ appears in (6.13) without a pre-multiplying $\Lambda(s)$. However, if one considers the product $P(s)J(s)y(s)$ in (6.13), one finds the term $\Lambda(s)y(s)$, and the appropriate substitution from Table 6.1 can be made.

The required filter for $\mathcal{G}^*$ is produced below by making the necessary substitutions throughout Section 6.2.2. Note that Table 6.1 is **not** required for the Bayesian revision $(s|s-1)\to(s|s)$, the algorithm for which is derived in Section 5.5.3. The table is used for transformations such as $E\left[\beta(s+1)|Y(s)\right]\to E\left[1/\left(h^*\right)^{-1}\left(M^*(s+1)\beta(s+1)|Y(s)\right)\right]$. In the following, the numbers of the corresponding equations from Section 6.2.2 are shown in square brackets. The equation numbers from Section 5.5.3 are also shown in square brackets when used.

Estimation of $1/\left(h^*\right)^{-1}(s\,|\,s)$ and $\Gamma^*(s\,|\,s)$

In the following $1/\left(h^*\right)^{-1}(s\,|\,s-j)$ will denote $E\left[1/\left(h^*\right)^{-1}\left(M^*(s\,|\,s-j)\beta(s-j)\right)|\,Y(s)\right]$ and $\Gamma^*(s\,|\,s)$ will denote the corresponding variance. Further $\left(h^*\right)^{-1}(s\,|\,s-j)$ will denote the reciprocal of $1/\left(h^*\right)^{-1}(s\,|\,s-j)$.

Define $M^*(s\,|\,s-1)$ and $Q^*(s\,|\,s-1)$ as the orthogonal and diagonal matrices respectively satisfying

$$Var\left[\beta(s)\,|\,Y(s-1)\right]=\left[M^*(s\,|\,s-1)\right]^T Q^*(s\,|\,s-1)M^*(s\,|\,s-1)\quad\left[(6.3)\right]\qquad (6.44)$$

$$M^*(s\,|\,s)=P^*(s)M^*(s\,|\,s-1)\qquad\left[(6.6)\right]\qquad (6.45)$$

where $P^*(s)$ is defined below.

Define $\beta(s\,|\,s-1)$ by the relation

$$\begin{aligned}1/\left(h^*\right)^{-1}\left(M(s\,|\,s-1)\beta(s\,|\,s-1)\right)&=E\left[1/\left(h^*\right)^{-1}\left(M^*(s\,|\,s-1)\beta(s)\right)|\,Y(s-1)\right]\\&=\left[N^*(s\,|\,s-1)\right]^{-1}w^*(s\,|\,s-1)\quad\left[(5.110)\text{ and }(6.8)\right]\end{aligned}\qquad (6.46)$$

Then

$$G^*(s)=DIAG\left(b^{-1}\right)''\left(X(s)\beta(s\,|\,s-1)\right)\qquad\left[(5.109)\text{ and }(6.9)\right]\qquad (6.47)$$

$$B^*(s\,|\,s-j)=DIAG\left(b^{-1}\right)''\left(M^*(s\,|\,s-j)\beta(s\,|\,s-1)\right),\ j=0,1\ \left[(5.108)\text{ and }(6.10)\right]\qquad (6.48)$$

Define $P^*(s)$ and $D^*(s)$ as the orthogonal and diagonal matrices respectively satisfying:

$$\begin{aligned}\left[P^*(s)\right]^T D^*(s)P^*(s)=&\,N^*(s\,|\,s-1)B^*(s\,|\,s-1)+M^*(s\,|\,s-1)X^T(s)\\&\left[DIAG\,G^*(s)\Lambda(s)y(s)\right]X(s)M^{*T}(s\,|\,s-1)\left[(5.106)\text{ and }(6.11)\right]\end{aligned}\qquad (6.49)$$

Then

$$J^*(s)=\left\{\left[P^*(s)\right]^T B^*(s\,|\,s)P^*(s)\right\}$$
$$\left\{N^*(s\,|\,s-1)B^*(s\,|\,s-1)+M^*(s\,|\,s-1)X^T(s)\left[DIAG\;G^*(s)\Lambda(s)y(s)\right]X(s)\right.$$
$$\left.\left[M^*(s\,|\,s-1)\right]^T\right\}^{-1} M^*(s\,|\,s-1)X^T(s)\Lambda(s)\quad\left[(6.12)\right] \tag{6.50}$$

$$1/\left(h^*\right)^{-1}(s\,|\,s)=1/\left(h^*\right)^{-1}\left(M^*(s\,|\,s)\beta(s\,|\,s-1)\right)$$
$$-P^*(s)J^*(s)\left\{\left[H^*_{(-1)}(s\,|\,s-1)\right]^{-1}y(s)-1\right\}\quad\left[(6.13)\right] \tag{6.51}$$

with

$$H^*_{(-1)}(s\,|\,s-1)=DIAG\left(h^*\right)^{-1}\left(X(s)\beta(s\,|\,s-1)\right).\quad[(5.104)] \tag{6.52}$$

Define

$$\alpha^*=\left(b^{-1}\right)''\circ\left(\left(b^{-1}\right)'\right)^{-1}\qquad[(5.117)] \tag{6.53}$$

$$A^*(s)=DIAG\,\alpha^*\left(1/\left(h^*\right)^{-1}\left(M(s\,|\,s)\beta(s\,|\,s-1)\right)\right)\quad\left[(5.114)\right]$$
$$A^{*\prime\prime}(s)=DIAG\,\alpha^{*\prime\prime}\left(\left(1/\left(h^*\right)^{-1}\right)\left(M^*(s\,|\,s)\beta(s\,|\,s-1)\right)\right)\;\left[(5.115)\text{ and }(6.14)\right] \tag{6.54}$$

$$N^*(s\,|\,s)=D^*(s)\left[B^*(s\,|\,s)\right]^{-1}\qquad\left[(5.105)\text{ and }(6.15)\right] \tag{6.55}$$

$$\Gamma^*(s\,|\,s)=-\left[N^*(s\,|\,s)+A^{*\prime\prime}(s)\right]^{-1}A^*(s).\quad\left[(5.113)\text{ and }(6.16)\right] \tag{6.56}$$

Estimation of $E\left[\beta(s+1)\,|\,Y(s)\right]$ and $Var\left[\beta(s+1)\,|\,Y(s)\right]$

Taylor series expansions corresponding to (6.19) and (6.20) are required. To obtain these, write

$$\beta(s)=\left[M^*(s\,|\,s)\right]^T\;h^*\left(\frac{1}{1/\left(h^*\right)^{-1}\left(M^*(s\,|\,s)\beta(s)\right)}\right)$$

and expand h^*, making use of the relations:

$$(d/dx)h^*(1/x)=-x^{-2}\left(h^*\right)'(1/x) \tag{6.57}$$

$$\left(d^2/dx^2\right)h^*\left(1/x\right)=2x^{-3}\left(h^*\right)'\left(1/x\right)+x^{-4}\left(h^*\right)''\left(1/x\right). \tag{6.58}$$

Thus

$$E\left[\beta\left(s+1\right)|Y\left(s\right)\right]=\chi\left(s+1\right)\left[M^*\left(s|s\right)\right]^T\left\{h^*\left(\left(h^*\right)^{-1}\left(s|s\right)\right)+\tfrac{1}{2}\Gamma^*\left(s|s\right)\right.$$
$$\left.\left[2\left[H^*_{(-1)}\left(s|s\right)\right]^3H^{*\prime}\left(s|s\right)+\left[H^*_{(-1)}\left(s|s\right)\right]^4H^{*\prime\prime}\left(s|s\right)1\right]\right\} \qquad \left[(6.22)\right] \tag{6.59}$$

$$Var\left[\beta\left(s+1\right)|Y\left(s\right)\right]=\chi\left(s+1\right)\left[M^*\left(s|s\right)\right]^T\left[H^*_{(-1)}\left(s|s\right)\right]^2H^{*\prime}\left(s|s\right)\Gamma^*\left(s|s\right)H^{*\prime}\left(s|s\right)$$
$$\left[H^*_{(-1)}\left(s|s\right)\right]^2M^*\left(s|s\right)\chi^T\left(s+1\right)+R\left(s+1\right) \qquad \left[(6.23)\right] \tag{6.60}$$

where

$$H^*_{(-1)}\left(s|s\right)=DIAG\left(h^*\right)^{-1}\left(s|s\right) \tag{6.61}$$

$$H^{*\prime}\left(s|s\right)=DIAG\left(h^*\right)'\left(\left(h^*\right)^{-1}\left(s|s\right)\right),$$
$$H^{*\prime\prime}\left(s|s\right)=\left(h^*\right)''\left(\left(h^*\right)^{-1}\left(s|s\right)\right). \qquad \left[(6.21)\right] \tag{6.62}$$

Estimation of $M^*\left(s+1|s\right)$

$M^*\left(s+1|s\right)$ is the orthogonal matrix that diagonalises $V\left[\beta\left(s+1\right)|Y\left(s\right)\right]$:

$$Var\left[\beta\left(s+1\right)|Y\left(s\right)\right]=\left[M^*\left(s+1|s\right)\right]^TQ^*\left(s+1|s\right)M^*\left(s+1|s\right)\ \left[(6.24)\right] \tag{6.63}$$

where $Q^*\left(s+1|s\right)$ is diagonal.

Estimation of $1/\left(h^*\right)^{-1}\left(s+1|s\right)$ and $\Gamma^*\left(s+1|s\right)$

$$1/\left(h^*\right)^{-1}\left(s+1|s\right)=1/\left(h^*\right)^{-1}\left(E\left[M^*\left(s+1|s\right)\beta\left(s+1\right)|Y\left(s\right)\right]\right)$$
$$+\tfrac{1}{2}Q^*\left(s+1|s\right)H^{*\prime\prime}_{(-1)}\left(s+1|s\right)1 \qquad \left[(6.27)\right] \tag{6.64}$$

$$\Gamma^*\left(s+1|s\right)=H^{*\prime}_{(-1)}\left(s+1|s\right)Q^*\left(s+1|s\right)H^{*\prime}_{(-1)}\left(s+1|s\right) \qquad \left[(6.29)\right] \tag{6.65}$$

where the expectation in (6.64) may be calculated from (6.59), and

$$H^{*}_{(-1)}{}'(s+1|s) = DIAG\left[1/\left(h^{*}\right)^{-1}\right]'\left(E\left[M^{*}(s+1|s)\beta(s+1)|Y(s)\right]\right). \ \left[(6.30)\right] \qquad (6.66)$$

Calculation of $N^{*}(s+1|s)$ and $w^{*}(s+1|s)$

$$N^{*}(s+1|s) = \left[\Gamma^{*}(s+1|s)\right]^{-1} \left[H^{*}_{(-1)}{}'(s+1|s) + \tfrac{1}{2}Q^{*}(s+1|s)H^{*}_{(-1)}{}'''(s+1|s)\right] \qquad \left[(6.35)\right] \qquad (6.67)$$

$$w^{*}(s+1|s) = N^{*}(s+1|s)\left[1/\left(h^{*}\right)^{-1}(s+1|s)\right] \ \left[(6.36)\right] \qquad (6.68)$$

where

$$H^{*}_{(-1)}{}'''(s+1|s) = DIAG\left[1/\left(h^{*}\right)^{-1}\right]''' \left(E\left[\left(M^{*}(s+1|s)\beta(s+1)\right)|Y(s)\right]\right) \quad \left[(6.34)\right] \qquad (6.69)$$

Forecasts

Define

$$\phi^{*}(s+1|s)(\nu) = \left(h^{*}\right)^{-1} X(s+1)\left[M^{*}(s+1|s)\right]^{T} h^{*}(1/\nu) \quad \left[(6.39)\right] \qquad (6.70)$$

so that

$$E\left[y(s+1)|Y(s)\right] = E\left[\psi^{*}(s+1|s)\left(1/\left(h^{*}\right)^{-1}\left(M^{*}(s+1|s)\beta(s+1)\right)\right)\right] \qquad (6.71)$$

It is possible to take the Taylor series expansion of this expression about $1/\left(h^{*}\right)^{-1}(s+1|s)$ and obtain results corresponding to (6.37) and (6.38). However, this is not done here as the general results are complicated and it will sometimes be simpler in practice to work directly from (6.71) for the specific h^{*} under consideration. This is illustrated in Example 6.2.

6.4 Special cases

Example 6.1: normal error, identity link

In this case, with h = identity, $h^{-1}(s|s-j)$ denotes $E\left[M(s|s-j)\beta(s)|Y(s-j)\right]$.

It will be convenient to begin the cycle of steps in the filter at a different point from that commencing (6.8) – (6.41).

Estimation of $N(s|s-1)$ and $w(s|s-1)$

Note that all H' matrices are identity matrices and all H'' and H''' are null matrices. Therefore

$$\begin{aligned} N(s|s-1) &= \Gamma^{-1}(s|s-1) \qquad [(6.35)] \\ &= M(s|s-1)Var^{-1}\left[\beta(s)|Y(s-1)\right]M^T(s|s-1) \end{aligned} \tag{6.72}$$

$$\begin{aligned} w(s|s-1) &= N(s|s-1)M(s|s-1)E\left[\beta(s)|Y(s-1)\right] \qquad [(6.36)] \\ &= M(s|s-1)Var^{-1}\left[\beta(s)|Y(s-1)\right]E\left[\beta(s)|Y(s+1)\right]. \end{aligned} \tag{6.73}$$

Estimation of $h^{-1}(s|s)$ and $\Gamma(s|s)$

The required estimation was carried for a single iteration of the filter in Example 5.2. The results obtained there, adapted to the present context, are as follows.

$$\begin{aligned} M^T(s|s)h^{-1}(s|s) &= E\left[\beta(s)|Y(s)\right] = E\left[\beta(s)|Y(s-1)\right] \\ &\quad + K(s)\left\{y(s) - X(s)E\left[\beta(s)|Y(s-1)\right]\right\} \qquad [(5.83)] \end{aligned} \tag{6.74}$$

where

$$K(s) = F(s)X^T(s)\left[\Lambda^{-1}(s) + X(s)F(s)X^T(s)\right]^{-1} \quad [(5.84)] \tag{6.75}$$

$$\begin{aligned} F(s) &= M^T(s|s-1)N^{-1}(s|s-1)M(s|s-1) \\ &= Var\left[\beta(s)|Y(s-1)\right], \end{aligned} \tag{6.76}$$

by (6.72).

Note that $\Lambda^{-1}(s) = Var\left[Y(s)|\beta(s)\right]$ by (3.4) and (3.5) with $p = 0$.

Now

$$\begin{aligned} M^T(s|s)\Gamma(s|s)M(s|s) &= Var\left[\beta(s)|Y(s)\right] \\ &= \left[1 - K(s)X(s)\right]F(s) \quad [(5.86)] \\ &= \left[1 - K(s)X(s)\right]Var\left[\beta(s)|Y(s-1)\right], \end{aligned} \tag{6.77}$$

by (6.76).

Estimation of $E\left[\beta(s+1)|Y(s)\right]$ and $Var\left[\beta(s+1)|Y(s)\right]$

$$\begin{aligned} E\left[\beta(s+1)|Y(s)\right] &= \chi(s+1)M^T(s|s)h^{-1}(s|s) \qquad [\text{by } (6.22)] \\ &= \chi(s+1)E\left[\beta(s)|Y(s)\right] \end{aligned} \tag{6.78}$$

$$Var\left[\beta(s+1)\,|\,Y(s)\right]=\chi(s+1)M^T(s\,|\,s)\Gamma(s\,|\,s)M(s\,|\,s)\chi^T(s+1)+R(s+1)\ \left[\text{by }(6.23)\right]$$
$$=\chi(s+1)Var\left[\beta(s)\,|\,Y(s)\right]\chi^T(s+1)+R(s+1). \tag{6.79}$$

The filter that maps $E\left[\beta(s)\,|\,Y(s-1)\right]$ and $Var\left[\beta(s)\,|\,Y(s-1)\right]$ to $E\left[\beta(s+1)\,|\,Y(s)\right]$ and $Var\left[\beta(s+1)\,|\,Y(s)\right]$ finally comprises (6.74) – (6.79), which may be recognised as the **Kalman filter** (Kalman, 1960; Jazwinski, 1970).

Example 6.2: gamma error, log link. This case was dealt with in Example 5.4, where it was identified as companion to the GLM with gamma error and canonical reciprocal link (5.121).

Estimation of $1/\left(h^*\right)^{-1}(s\,|\,s)$ and $\Gamma^*(s\,|\,s)$

Here $1/\left(h^*\right)^{-1}(s\,|\,s)$ denotes $E\left[\exp-M^*(s\,|\,s)\beta(s)\,|\,Y(s)\right]$ and $\Gamma^*(s\,|\,s)$ the corresponding variance.

Define $\beta(s\,|\,s-1)$ by the relation

$$\exp-M(s\,|\,s-1)\beta(s\,|\,s-1)=\left[N^*(s\,|\,s-1)\right]^{-1}w^*(s\,|\,s-1).\quad\left[(6.46)\right] \tag{6.80}$$

Also define

$$G^*(s)=-\left[DIAG\exp-X(s)\beta(s\,|\,s-1)\right],\ \left[(5.122)\right]$$
$$B^*(s\,|\,s-j)=-\left[DIAG\exp-M^*(s\,|\,s-j)\beta(s\,|\,s-1)\right],\ j=0,1\ \left[(6.48)\right]. \tag{6.81}$$

Matrices $P^*(s),D^*(s)$, $J^*(s)$ and $M^*(s,s)$ are defined by (6.49), (6.50) and (6.45).

Then, using the form (5.128) that is specifically adapted to the present special case,

$$1/\left(h^*\right)^{-1}(s\,|\,s)=\left[DIAG\exp-M^*(s\,|\,s)\beta(s\,|\,s-1)\right]$$
$$\left\{\underset{q\times1}{1}-K(s)\left[y(s)-\hat{y}(s\,|\,s-1)\right]\right\}\qquad\left[(5.128)\text{ and }(6.51)\right] \tag{6.82}$$

with

$$\hat{y}(s\,|\,s-1)=\exp X(s)\beta(s\,|\,s-1) \tag{6.83}$$

$$K(s)=\left[B^*(s\,|\,s)\right]^{-1}P^*(s)J^*(s)G^*.$$

Also

$$N^*(s|s) = D^*(s)\left[B^*(s|s)\right]^{-1} \quad \left[(6.55)\right] \tag{6.84}$$

$$\begin{aligned}\Gamma^*(s|s) &= \left[N^*(s|s)\right]^{-1} DIAG \exp - M^*(s|s)\beta(s|s) \; \left[\text{by } (6.56) \text{ and } (5.133)\right] \\ &= -\left[D^*(s)\right]^{-1} B^2(s|s)\end{aligned} \tag{6.85}$$

by (6.81) and (6.84).

Estimation of $E\left[\beta(s+1)|Y(s)\right]$ and $Var\left[\beta(s+1)|Y(s)\right]$

Recall from (5.120) that $h^*(\nu) = \log \nu$, so that (6.61) and (6.62) yield

$$H^*_{(-1)}(s|s) = DIAG\left(h^*\right)^{-1}(s|s) \tag{6.86}$$

$$H^{*\prime}(s|s) = \left[H^*_{(-1)}(s|s)\right]^{-1}, \quad H^{*\prime\prime} = -\left[H^*_{(-1)}(s|s)\right]^{-2}. \tag{6.87}$$

Then

$$\begin{aligned}E\left[\beta(s+1)|Y(s)\right] = \chi(s+1)\left[M^*(s|s)\right]^T &\Big\{-\log\left[1/\left(h^*\right)^{-1}(s|s)\right] \\ &+ \tfrac{1}{2}\Gamma^*(s|s)\left[H^*_{(-1)}(s|s)\right]^2 1\Big\} \qquad \left[(6.59)\right]\end{aligned} \tag{6.88}$$

$$\begin{aligned}Var\left[\beta(s+1)|Y(s)\right] = \chi(s+1)\left[M^*(s|s)\right]^T &H^*_{(-1)}(s|s)\Gamma^*(s|s) \\ &H^*_{(-1)}(s|s)M^*(s|s)\chi^T(s+1) + R(s+1). \quad \left[(6.60)\right]\end{aligned} \tag{6.89}$$

Estimation of $M^*(s+1|s)$

Define $M^*(s+1|s)$ and $Q^*(s+1|s)$ as the orthogonal and diagonal matrices respectively that satisfy:

$$Var\left[\beta(s+1)|Y(s)\right] = \left[M^*(s+1|s)\right]^T Q^*(s+1|s)M^*(s+1|s). \qquad \left[(6.63)\right]$$

Estimation of $1/\left(h^*\right)^{-1}(s+1|s)$ and $\Gamma^*(s+1|s)$

$$1/\left(h^*\right)^{-1}(s+1|s) = \left[1 - \tfrac{1}{2}Q^*(s+1|s)\right]\exp - E\left[M^*(s+1|s)\beta(s+1)|Y(s)\right] \; \left[(6.64)\right] \tag{6.90}$$

$$\Gamma^*(s+1|s) = Q^*(s+1|s)DIAG\exp - 2E\left[M^*(s+1|s)\beta(s+1)|Y(s)\right] \quad \left[(6.65)\right] \tag{6.91}$$

Calculation of $N^*(s+1|s)$ and $w^*(s+1|s)$

$$N^*(s+1|s) = \left[\Gamma^*(s+1|s)\right]^{-1}\left[1+\tfrac{1}{2}Q^*(s+1|s)\right] \\ DIAG \exp - E\left[M^*(s+1|s)\beta(s+1)|Y(s)\right]. \quad \left[(6.67)\right] \tag{6.92}$$

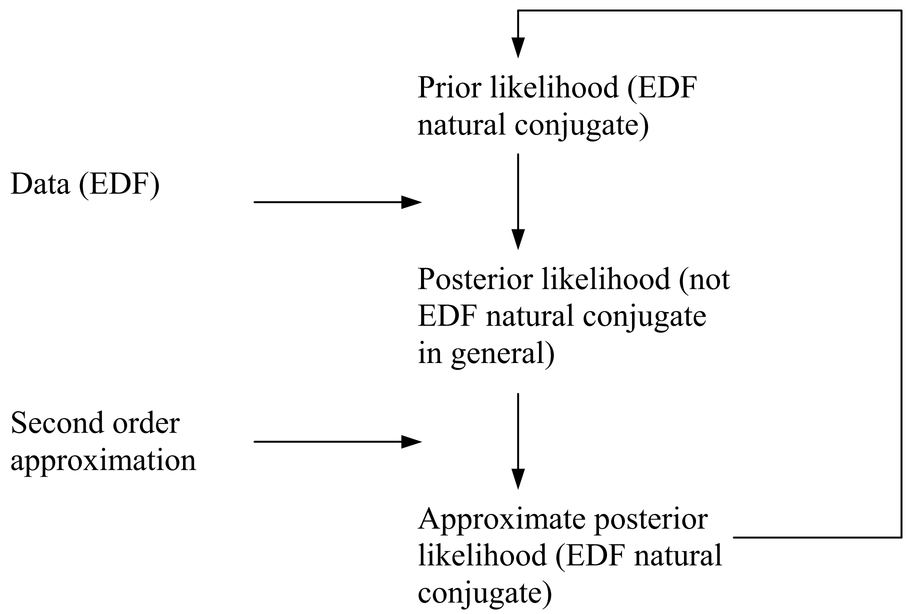
$$w^*(s+1|s) = N^*(s+1|s)\left[1/\left(h^*\right)^{-1}(s+1|s)\right]. \quad \left[(6.68)\right] \tag{6.93}$$

6.5 Commentary

Sections 6.2 and 6.3 extend the Kalman filter from dynamic general linear models to dynamic generalised linear models. However, the static equation (6.4) remains linear. This work is therefore distinct from non-linear extensions such as the **extended Kalman filter** (Jazwinski, 1970, pp.272-281; Harvey 1989, pp.160-162) and others (eg Naik-Nimbalkar and Rajarshi, 1995).

It is worthwhile summarising the logic of a single iteration of the DGLM filter produced here. It is as set out in Figure 6.1.

Figure 6.1
Single iteration of DGLM filter



The accuracy of the filter as an estimator will depend on the accuracy of the approximation step. As noted at the end of Section 5.5.1, there are indications that such estimators are highly accurate in certain circumstances. However, the author intends to carry out in the near future a program of numerical testing of the filter presented here.

Appendix A
Taylor expansion of (5.9)

A.1 Expansion of $\boldsymbol{n}_0^T(\boldsymbol{b} \circ \boldsymbol{t})$

By (2.1),

$$b(t) = \left[b(t_1),\ldots,b(t_q)\right]^T \tag{A.1}$$

where *t* denotes $t(\beta)$ and t_j denotes $t_j(\beta)$.

Expand $b(t_j)$ to second order:

$$b(t_j) = b(t_{j0}) + (t_j - t_{j0})b'(t_{j0}) + \tfrac{1}{2}(t_j - t_{j0})^2 b''(t_{j0}) \tag{A.2}$$

where

$$t_{j0} = t_j(\beta_0) \tag{A.3}$$

for an arbitrarily selected q-vector β_0.

Then

$$\begin{aligned} n_0^T b(t) &= \sum_j n_{0j} b(t_j) \\ &= n_0^T b(t_0) + (t - t_0)^T N_0 b'(t_0) + \tfrac{1}{2}(t - t_0)^T N_0 \left[DIAG\, b''(t_0)\right](t - t_0) \end{aligned} \tag{A.4}$$

where

$$t_0 = \underset{q \times q}{t}(\underset{q \times 1}{\beta_0}) \tag{A.5}$$

$$\underset{q \times q}{N_0} = DIAG\, n_0. \tag{A.6}$$

By (5.1) and (A.5),

$$b'(t_0) = h^{-1}(M\beta_0). \tag{A.7}$$

Substitution of (A.7) in (A.4) yields

$$n_0^T b(t) = n_0^T b(t_0) + (t - t_0)^T N_0 h^{-1}(M\beta_0) + \tfrac{1}{2}(t - t_0)^T N_0 B_0 (t - t_0) \tag{A.8}$$

with

$$B_0 = diag\, b''(t_0) = DIAG\, b'' \circ (b')^{-1} \circ h^{-1}(M\beta_0). \tag{A.9}$$

A.2 Expansion of $y^T \Lambda(g \circ t)$

Equation (5.4) defines g as an m x q transformation. The Taylor expansion (2.12) may be used to express $g(t)$ in the form

$$g(t) = g(t_0) + Dg(t_0)(t - t_0) + \tfrac{1}{2}\Big[(t - t_0) \otimes \underset{m \times m}{1}\Big]^T D^2 g(t_0)(t - t_0) \tag{A.10}$$

where Dg and $D^2 g$ are defined in Section 2, and $\otimes$ denotes the Kronecker product defined there.

Then

$$\begin{aligned} y^T \Lambda g(t) &= y^T \Lambda g(t_0) + (t - t_0)^T (Dg)^T \Lambda y \\ &\quad + \tfrac{1}{2}(t - t_0)^T \Big[(\Lambda y) \otimes \underset{q \times q}{1}\Big]^T \big(D^2 g\big)(t - t_0). \end{aligned} \tag{A.11}$$

A.3 Expansion of $1^T \Lambda(b \circ g \circ t)$

The algebra proceeds largely as in Appendix A.2, but with $g(t)$ replaced by $b(g(t))$. In parallel with (A.11)

$$\begin{aligned} 1^T \Lambda b(g(t)) &= 1^T \Lambda b(g(t_0)) + (t - t_0)^T \big[D(b \circ g)\big]^T \Lambda 1 \\ &\quad + \tfrac{1}{2}(t - t_0)^T \Big[\underset{m \times 1}{\Lambda 1} \otimes \underset{q \times q}{1}\Big]^T \big[D^2(b \circ g)\big](t - t_0). \end{aligned} \tag{A.12}$$

Now, evaluated at t_0,

$$\begin{aligned} D(b \circ g) &= \Big[\underset{m \times m}{D}\, b(g(t_0))\Big] \underset{m \times q}{Dg}(t_0) \\ &= \big[DIAG(b' \circ g)\big](Dg) \end{aligned} \tag{A.13}$$

where both members are evaluated at t_0.

By (5.1), (5.4) and (A.5),

$$(b' \circ g)(t_0) = h^{-1}(X\beta_0) \qquad \text{[compare (A.7)]} \tag{A.14}$$

and so (A.13) may be written

$$D(b \circ g) = H_{(-1)}(Dg) \tag{A.15}$$

with

$$H_{(-1)} = DIAG\, h^{-1}\left(X\beta_0\right). \tag{A.16}$$

Moreover,

$$\underset{mq \times q}{D^2\left(b \circ g\right)} = \begin{bmatrix} D^2\left(b \circ g_1\right) \\ \vdots \\ D^2\left(b \circ g_m\right) \end{bmatrix} \tag{A.17}$$

and $D^2\left(b \circ g_i\right)$ has (*j*,*k*) element

$$\begin{aligned} \left[D^2\left(b \circ g_i\right)\right]_{jk} &= \partial^2\left(b \circ g_i\right)/\partial t_j \partial t_k \\ &= \left(b' \circ g_i\right)\partial^2 g_i / \partial t_j \partial t_k + \left(b'' \circ g_i\right)\left(\partial g_i / \partial t_j\right)\left(\partial g_i / \partial t_k\right) \end{aligned} \tag{A.18}$$

with all evaluations at $t = t_0$.

Then the (*j*,*k*) element of $\left[\Lambda 1 \otimes 1\right]^T D^2\left(b \circ g\right)$ is

$$\begin{aligned} \left\{\left[\Lambda 1 \otimes 1\right]^T D^2\left(b \circ g\right)\right\}_{jk} &= \sum_i \Lambda_{ii}\left(b' \circ g_i\right)\partial^2 g_i / \partial t_j \partial t_k \\ &\quad + \sum_i \Lambda_{ii}\left(b'' \circ g_i\right)\left(\partial g_i / \partial t_j\right)\left(\partial g_i / \partial t_k\right). \end{aligned} \tag{A.19}$$

Hence

$$\left[\Lambda 1 \otimes 1\right]^T \left[D^2\left(b \circ g\right)\right] = \left(H_{(-1)}\Lambda 1 \otimes \underset{q \times q}{1}\right)^T \left(D^2 g\right) + \left(Dg\right)^T G\Lambda\left(Dg\right) \tag{A.20}$$

where use has been made of (A.14),

$$G = DIAG\left(b'' \circ g\right) = DIAG\left[b'' \circ \left(b'\right)^{-1} \circ h^{-1}\left(X\beta_0\right)\right] \tag{A.21}$$

and all quantities are evaluated at $t = t_0$.

Substitution of (A.15) and (A.20) into (A.12) yields

$$\begin{aligned} 1^T \Lambda\left(b \circ g \circ t\right) &= 1^T \Lambda b\left(g\left(t_0\right)\right) + \left(t - t_0\right)^T \left(Dg\right)^T H_{(-1)}\Lambda 1 \\ &\quad + \tfrac{1}{2}\left(t - t_0\right)^T \left\{\left(H_{(-1)}\Lambda 1 \otimes \underset{q \times q}{1}\right)^T \left(D^2 g\right) + \left(Dg\right)^T G\Lambda\left(Dg\right)\right\}\left(t - t_0\right) \end{aligned} \tag{A.22}$$

Appendix B
Evaluation of $E_\beta\left[b'\left(t_j\right)\right]$ and $Var_\beta\left[b'\left(t_j\right)\right]$

Recall the log-likelihood of β (5.6), which is seen to be the sum of *q* independent log-likelihoods. The *j*-th of these is the log-likelihood of $t_j(\beta)$, and is

$$w_{0j}t_j(\beta) - n_{0j}b\,\mathrm{o}t_j(\beta) \tag{B.1}$$

Denote this by $\pi\left(t_j(\beta)\right)$. Then, with $t_j(\beta)$ written as just t_j,

$$\int \exp\pi\left(t_j\right)dt_j = K, \quad \text{normalising constant.} \tag{B.2}$$

It will be convenient to suppress the subscript *j* temporarily. Then consider the quantity

$$\begin{aligned}\int (d/dt)\exp\pi(t)d(t) &= \int\left[w_0 - n_0 b'(t)\right]\exp\pi(t)\,dt \\ &= K\left\{w_0 - n_0 E_\beta\left[b'(t)\right]\right\}.\end{aligned} \tag{B.3}$$

Now integration of the left side by parts gives

$$\int (d/dt)\exp\pi(t)\,dt = \exp\pi(t)\,|_{t\in T} \tag{B.4}$$

where *T* is the boundary of the range of integration. If the likelihood of $t_j(\beta)$ is zero on the boundary of its support, then (B.4) reduces to zero, and (B.3) yields (with subscript *j* reinstated)

$$E_\beta\left[b'\left(t_j\right)\right] = w_{0j}/n_{0j}. \tag{B.5}$$

Now repeat this reasoning with (B.3) replaced by the following:

$$\begin{aligned}n_0^{-2}\int\left(d^2/dt^2\right)\exp\pi(t)\,dt &= \int\left\{\left[w_0/n_0 - b'(t)\right]^2 - b''(t)/n_0\right\}\exp\pi(t) \\ &= V_\beta\left[b'(t)\right] - E_\beta\left[b''(t)\right]/n_0.\end{aligned} \tag{B.6}$$

If the first derivative of the likelihood of $t_j(\beta)$ is zero on the boundary of its support, then (B.6) reduces to zero, and so

$$Var_\beta\left[b'\left(t_j\right)\right] = E_\beta\left[b''\left(t_j\right)\right]/n_{0j}. \tag{B.7}$$

Companion GLM

The prior (5.97) associated with the companion GLM $\mathcal{G}^*$ discussed in Section 5.5.3 contains the following member in place of (B.1) from the prior of $\mathcal{G}$:

$$-w^*_{0j} t_j(\beta) + n^*_{0j} b^{-1} \circ t_j(\beta). \tag{B.8}$$

The reasoning from (B.2) to (B.7) may be repeated with (B.1) replaced by (B.8), yielding the following results:

$$E_\beta\left[\left(b^{-1}\right)'\left(t_j\right)\right] = w^*_{0j} / n^*_{0j} \tag{B.9}$$

$$Var_\beta\left[\left(b^{-1}\right)'\left(t_j\right)\right] = -E_\beta\left[\left(b^{-1}\right)''\left(t_j\right)\right] / n^*_{0j}. \tag{B.10}$$

Appendix C
Bayesian revision of mean and covariance

This appendix computes quantities required in Section 5.5.2.

C.1 Variance of $\boldsymbol{h}^{-1}(\boldsymbol{M}\beta)$

By (5.1),

$$\begin{aligned} Var_\beta\left[h^{-1}(M\beta)\right] &= Var_\beta\left[b'(t)\right] \\ &= N_0^{-1} DIAG\, E_\beta\left[b''(t)\right], \end{aligned} \tag{C.1}$$

by (B.7).

Express this in the form:

$$Var_\beta\left[h^{-1}(M\beta)\right] = N_0^{-1} DIAG\, E_\beta\left[\alpha(\nu)\right] \tag{C.2}$$

with

$$\alpha = b'' \text{o} (b')^{-1} \tag{C.3}$$

$$\nu = b'(t) = h^{-1}(M\beta). \tag{C.4}$$

Now take the Taylor series expansion (2.14) of $E\left[\alpha(\nu)\right]$ about $\nu = \nu_0 = E\left[h^{-1}(M\beta)\right]$, as defined by (5.46):

$$\begin{aligned} E\left[\alpha(\nu)\right] &= \alpha(\nu_0) + \tfrac{1}{2}\left[PRIN\, Var\left[h^{-1}(M\beta)\right]\right]\alpha''(\nu_0) \\ &= \alpha(\nu_0) + \tfrac{1}{2} N_0^{-1}\left[DIAG\, E\left[\alpha(\nu)\right]\right]\alpha''(\nu_0) \end{aligned} \tag{C.5}$$

by (C.2).

Thus

$$E\left[\alpha(\nu)\right] = \left(1 - N_0^{-1} A''\right)^{-1} \alpha(\nu_0) \tag{C.6}$$

where

$$A'' = DIAG\, \alpha''(\nu_0). \tag{C.7}$$

Substitution of (C.6) in (C.2) gives

$$Var_\beta\left[h^{-1}(M\beta)\right] = (N_0 - A'')^{-1} A \tag{C.8}$$

where

$$A = DIAG\,\alpha(\nu_0). \tag{C.9}$$

C.2 Mean of $h^{-1}(X\beta)$

By (3.2), (4.1) and the fact that $\mu(\theta) = b'(\theta)$,

$$\begin{aligned} E[Y \mid \beta] &= h^{-1}(X\beta) \\ &= \varphi(\nu) \end{aligned} \tag{C.10}$$

with

$$\varphi = h^{-1} \circ XM^{-1} \circ h. \tag{C.11}$$

The unconditional mean of Y is then given by

$$E[Y] = E_\beta[Y \mid \beta] = E_\beta\varphi(\nu). \tag{C.12}$$

Taking the Taylor series expansion (2.13) of $E[Y_i]$ gives

$$\begin{aligned} E[Y_i] &= \varphi_i(\nu_0) + \tfrac{1}{2} Tr\left[Var[\nu] D^2\varphi_i(\nu_0)\right] \\ &= \varphi_i(\nu_0) + \tfrac{1}{2}\sum_j Var\left[\nu_j\right]\partial^2\varphi_i(\nu_0)/\partial\nu_j^2. \end{aligned} \tag{C.13}$$

The second derivative is evaluated in Appendix D, and the variance in Appendix C.1. Substitution of (C.8) and (D.3) into (C.13) gives

$$\begin{aligned} E[Y_i] = \varphi_i(\nu_0) &+ \tfrac{1}{2}\left\{H'_{(-1)} XM^{-1} H''\right\}_i (N_0 - A'')^{-1} A1 \\ &+ \tfrac{1}{2}\left\{\left[H''_{(-1)}\right]^{\frac{1}{2}} XM^{-1} H'\right\}_i (N_0 - A'')^{-1} A\{\ldots\}_i^T \end{aligned} \tag{C.14}$$

where

$$H' = DIAG\,h'(\nu_0) \tag{C.15}$$

$$H'' = DIAG\,h''(\nu_0) \tag{C.16}$$

$$H'_{(-1)} = DIAG\left(h^{-1}\right)'(\eta_0) \tag{C.17}$$

$$H''_{(-1)} = DIAG\left(h^{-1}\right)''(\eta_0) \tag{C.18}$$

$\eta = X\beta$, as defined in Appendix D, and $\eta_0 = X\beta_0$, $\{...\}_i$ denotes the *i*-th row of the matrix argument and $\{...\}_i^T$ denotes the transpose of that row.

C.3 Covariance of $h^{-1}(X\beta)$

By (C.10),

$$\begin{aligned} Var\left[h^{-1}(X\beta)\right] &= Var\left[\varphi(\nu)\right] \\ &= D\varphi(\nu_0)Var[\nu]\left[D\varphi(\nu_0)\right]^T \end{aligned} \tag{C.19}$$

to first approximation, where $D\varphi$ denotes the Jacobian matrix of φ. By (D.2),

$$D\varphi(\nu_0) = H'_{(-1)}XM^{-1}H'. \tag{C.20}$$

Substitution of (C.8) and (C.20) into (C.19) gives

$$Var\left[h^{-1}(X\beta)\right] = H'_{(-1)}XM^{-1}H'(N_0 - A'')^{-1} AH'\left(XM^{-1}\right)^T H'_{(-1)} \tag{C.21}$$

Appendix D
Derivatives of φ

This appendix calculates the first and second derivatives of the function φ introduced in (C.11). The *i*-th component of $\varphi(\nu)$ is

$$\varphi_i(\nu) = h^{-1}\left(\sum_k r_{ik} h(\nu_k)\right) \tag{D.1}$$

where r_{ik} is the (i,k) element of $R = XM^{-1}$.

Then

$$\begin{aligned}\partial\varphi_i(\nu)/\partial\nu_j &= \left(h^{-1}\right)'\left(\sum_k r_{ik} h(\nu_k)\right) r_{ij} h'(\nu_j) \\ &= \left(h^{-1}\right)'(\eta_i)\, r_{ij} h'(\nu_j) \quad \left[\text{by (C.10)}\right]\end{aligned} \tag{D.2}$$

with η denoting the linear predictor $X\beta$.

A second differentiation gives

$$\partial^2\varphi_i(\nu)/\partial\nu_j^2 = \left(h^{-1}\right)'(\eta_i)\, r_{ij} h''(\nu_j) + \left(h^{-1}\right)''(\eta_i)\left[r_{ij} h'(\nu_j)\right]^2. \tag{D.3}$$

Appendix E
Bayesian revision of companion GLM

In the revision of $\mathcal{G}$, Appendix A was used to convert (5.9) into the form (5.25). The present appendix will parallel Appendix A, but subject to the replacements (5.98), in order to obtain a Taylor expansion of (5.99).

E.1 Expansion of $\boldsymbol{n}_o^{*T}\left(\boldsymbol{b}^{-1}\circ\boldsymbol{t}\right)$

Note that

$$\begin{aligned}\left(b^{-1}\right)'(t_0) &= 1/\left(b'\circ b^{-1}\circ t\right)(\beta_0)\\ &= 1/\left(b'\circ b^{-1}\circ(b')^{-1}\circ h^{-1}\right)(M\beta_0) \qquad [\text{by } (5.1)]\\ &= 1/\left(h^*\right)^{-1}(M\beta_0) \end{aligned} \tag{E.1}$$

by (4.8).

Also

$$\left(b^{-1}\right)''(t_0) = \left(\left(b^{-1}\right)''\circ(b')^{-1}\circ h^{-1}\right)(M\beta_0). \tag{E.2}$$

Therefore, (A.8) is replaced by:

$$\begin{aligned} n_0^{*T}b^{-1}(t) = n_0^{*T}b^{-1}(t_0) + (t-t_0)^T\, N_0^*\left[1/\left(h^*\right)^{-1}(M\beta_0)\right]\\ + \tfrac{1}{2}(t-t_0)^T\, N_0^* B_0^*\,(t-t_0) \end{aligned} \tag{E.3}$$

with

$$\begin{aligned} B_0^* &= DIAG\left(b^{-1}\right)''\circ(b')^{-1}\circ h^{-1}(M\beta_0)\\ N_0^* &= DIAG\, n_0^*. \end{aligned} \tag{E.4}$$

E.2 Expansion of $\mathbf{1}^T\Lambda\left(\boldsymbol{g}\circ\boldsymbol{t}\right)$

This is just as in Appendix A.2 but with y replaced by 1.

E.3 Expansion of $\boldsymbol{y}^T\Lambda\left(\boldsymbol{b}^{-1}\circ\boldsymbol{g}\circ\boldsymbol{t}\right)$

Relation (A.13) is replaced by:

$$D\left(b^{-1}\circ g\right) = \left[DIAG\left(b^{-1}\right)'\circ g\right](Dg) \tag{E.5}$$

Now

$$\left(\left(b^{-1}\right)' \circ g\right)(t) = \left(b^{-1}\right)'(\theta) = 1/\left(b' \circ b^{-1}\right)(\theta)$$
$$= 1/\left(b' \circ b^{-1} \circ \left(b^{-1}\right)' \circ h^{-1}\right)(X\beta) = 1/\left(h^*\right)^{-1}(X\beta) \tag{E.6}$$

where the successive equations make use of (5.3), (4.3) and (4.8). Therefore

$$\left(\left(b^{-1}\right)' \circ g\right)(t_0) = 1/\left(h^*\right)^{-1}(X\beta_0) \tag{E.7}$$

Substitution of this into (E.5) gives the following result corresponding to (A.15):

$$D\left(\left(b^{-1}\right)' \circ g\right) = \left[H^*_{(-1)}\right]^{-1}(Dg) \tag{E.8}$$

where

$$H^*_{(-1)} = DIAG\left(h^*\right)^{-1}(X\beta_0) \qquad \text{[compare (A.16)]} \tag{E.9}$$

The reasoning from (A.17) to (A.22) holds in the companion case except that G is replaced by

$$G^* = DIAG\left[\left(\left(b^{-1}\right)'' \circ g\right)(t_0)\right] = DIAG\left[\left(b^{-1}\right)'' \circ \left(b'\right)^{-1} \circ h^{-1}(X\beta_0)\right] \tag{E.10}$$

This then yields the following in place of (A.22):

$$y^T\Lambda\left(b^{-1} \circ g \circ t\right) = y^T\Lambda b^{-1}\left(g(t_0)\right) + \left(t - t_0\right)^T (Dg)^T \left(H^*_{(-1)}\right)^{-1} \Lambda y$$
$$+ \tfrac{1}{2}\left(t - t_0\right)^T \left\{\left(\left(H_{(-1)}{}^*\right)^{-1} \Lambda y \otimes \underset{q \times q}{1}\right)^T D^2 g\right.$$
$$\left. + (Dg)^T \left[DIAG\ G^*\Lambda y\right](Dg)\right\}\left(t - t_0\right). \tag{E.11}$$

E.4 Counterpart to Section 5.5.1

The current appendix develops for the companion GLM $\mathcal{G}^*$ the reasoning given in Section 5.5.1 for $\mathcal{G}$.

Corresponding to (5.25) and (5.26):

$$L^*[\beta \mid y,\Lambda] = (t-t_0)^T a^* + \tfrac{1}{2}(t-t_0)^T \Big\{ N_0^* B_0^* + (Dg)^T \big[DIAG\ G^* \Lambda y \big](Dg) + \Big[\Lambda \Big[\big(H_{(-1)}{}^*\big)^{-1} y - 1 \Big] \otimes 1 \Big]^T \big(D^2 g\big) \Big\}(t-t_0) \tag{E.12}$$

where

$$a^* = \Big\{ N_0^* \Big[1/\big(h^*\big)^{-1}(M\beta_0) \Big] - w_0^* \Big\} + (Dg)^T \Lambda \Big[\big(H_{(-1)}{}^*\big)^{-1} y - 1 \Big]. \tag{E.13}$$

Now fix the arbitrary parameter β_0 by setting

$$\begin{aligned} 1/\big(h^*\big)^{-1}(M\beta_0) &= E_\beta \Big[1/\big(h^*\big)^{-1}(M\beta) \Big] \\ &= E_\beta \Big[\big(b^{-1}\big)'(t) \Big] \qquad [\text{by}(\text{E.1})] \\ &= \big(N_0^*\big)^{-1} w_0^* \quad [\text{by}(\text{B.9})]. \end{aligned} \tag{E.14}$$

This reduces (E.13) to:

$$a^* = (Dg)^T \Lambda \Big[\big(H_{(-1)}^*\big)^{-1} y - 1 \Big]. \tag{E.15}$$

Now diagonalise the matrix within the braces in (E.12):

$$P^{*T} D^* P^* = N_0^* B_0^* + (Dg)^T \big[DIAG\ G^* \Lambda y \big](Dg) + \Big[\Lambda \Big[\big(H_{(-1)}^*\big)^{-1} y - 1 \Big] \otimes 1 \Big]^T \big(D^2 g\big) \tag{E.16}$$

with D^* diagonal and P^* orthogonal.

Then applying to (E.12) the same reasoning as led from (5.25) to (5.36), one obtains

$$L^*(\beta \mid y,\Lambda) = -w_1^{*T} u^* + n_1^{*T} b^{-1}\big(u^*\big) \tag{E.17}$$

where

$$w_1^* = -P^* (Dg)^T \Lambda \Big[\big(H_{(-1)}^*\big)^{-1} y - 1 \Big] + N_1^* \Big[1/\big(b' \mathrm{o} b^{-1}\big)(u_0) \Big] \tag{E.18}$$

$$N_1^* = D^* \Big[DIAG \big(b^{-1}\big)''(u_0) \Big]^{-1} \tag{E.19}$$

$$n_1^* = VEC\ N_1^* \tag{E.20}$$

$$u^* = P^* t. \tag{E.21}$$

References

de Alba, E. (2002). Bayesian estimation of outstanding claim reserves. **North American Actuarial Journal**, 6(1), 1-20.

De Jong, P. (1977). The scan sampler for time series models. **Biometrika**, 84(4), 929-937.

Fahrmeir, L. (1992). Posterior mode estimation by extended Kalman filtering for multivariate dynamic generalised linear models. **Journal of the American Statistical Association**, 87, 501-509.

Fahrmeir, L. and Kaufmann, H. (1991). On Kalman filtering, posterior mode estimation and Fisher scoring in dynamic exponential family regression. **Metrika**, 38, 37-60.

Harvey, A.C. (1989). **Forecasting, structural time series and the Kalman filter**. Cambridge University Press, Cambridge, UK.

Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their application. **Biometrika**, 57, 97-109.

Jazwinski, A.H. (1970). **Stochastic processes and filtering theory**. Academic Press, New York.

Landsman, Z. (2002). Credibility theory: a new view from the theory of second order optimal statistics. **Insurance: mathematics and economics**, 30(3), 351-362.

Landsman, Z. (2004). Second order Bayes prediction of functionals of exponential dispersion distributions and an application to the prediction of the tails. **Astin Bulletin**, 34 (2), 285-298.

Landsman, Z. and Makov, U. (1998). Exponential dispersion models and credibility. **Scandinavian Actuarial Journal**, 89-96.

Levit, B. (1980). Second order minimaxity. **Theory of Probability and its Applications**, 25(3), 561-576.

McCullagh, P. and Nelder, J.A. (1989). **Generalised linear models** (3rd edition). Chapman and Hall, New York.

Naik-Nimbalkar, U.V. and Rajarshi, M.B. (1995). Filtering and smoothing via estimating functions. **Journal of the American Statistical Association**, 90, 301-306.

Nelder, J.A. and Verrall, R.J. (1997). Credibility theory and generalized linear models. **Astin Bulletin**, 27(3), 71-82.

Nelder, J.A. and Wedderburn, R.W.M. (1972). Generalized linear models. **Journal of the Royal Statistical Society, Series A**, 135, 370-384.

Ntzoufras, I. and Dellaportas, P. (2002). Bayesian modelling of outstanding liabilities incorporating claim count uncertainty, **North American Actuarial Journal**, 6(1), 113-136.

Scollnik, D.P.M. (2001). Actuarial modelling with MCMC and BUGS, **North American Actuarial Journal** 5(2), 96-125.

Scollnik, D.P.M. (2002). Regression models for bivariate loss data. **North American Actuarial Journal**, 6(4), 67-80.

Smith, J.Q. (1979). A generalization of the Bayesian steady forecasting model. **Journal of the Royal Statistical Society, Series B**, 41, 375-387.

Smith, A.F.M. and Roberts, G.O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo method. **Journal of the Royal Statistical Society, Series B**, 55, 3-24.

Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion). **The Annals of Statistics**, 22, 1701-1762.

West, M.P., Harrison, P.J. and Migon, H.S. (1995). Dynamic generalised models and Bayesian forecasting (with discussion). **Journal of the American Statistical Association**, 80, 73-97.