Les Cahiers du GERAD ISSN: 0711–2440

A Hierarchical Bayes Approach to Estimation and Prediction for Time Series of Counts

P.J. Farrell, B. MacGibbon, T.J. Tomberlin

G–2003–25

April 2003
Revised: February 2007

A Hierarchical Bayes Approach to Estimation and Prediction for Time Series of Counts

Patrick J. Farrell
Carleton University

Brenda MacGibbon
GERAD and
Université du Québec à Montréal

Thomas J. Tomberlin
Concordia University

April 2003
Revised: February 2007

Les Cahiers du GERAD
G–2003–25

Abstract

In this paper, we are concerned with the statistical methodology of epidemiological surveillance; that is, the ongoing procedure of analyzing and interpreting public health data of infectious disease incidence. In particular, we propose a hierarchical Bayes approach for the estimation of generalized linear mixed models for time series count data, and their use in the prediction of counts for future time periods. The estimators are obtained by Gibbs sampling and their performance is compared to those of other methods on the polio data originally analysed by Zeger (1988), which consist of the monthly number of U.S. polio cases between 1970 and 1983. Their properties are also investigated via simulation. Our aim is to illustrate how easily the hierarchical Bayes methodology lends itself to model checking and model comparisons. The proposed methodology, in particular, hierarchical Bayes prediction, is applied to a series of *Campylobacter* infection cases in the Montreal-Center region.

Key Words: Adaptive Rejection Metropolis Sampling, Epidemiological Surveillance, Gibbs Sampling, Infectious Diseases, Longitudinal Data, Poisson Counts.

Résumé

Dans cet article, nous nous intéressons à la méthodologie statistique de la surveillance épidémiologique, c'est-à-dire, la procédure de l'analyse et l'interprétation des données de santé publique concernant l'incidence de la maladie infectieuse. Plus précisément, nous proposons une approche bayésienne hiérarchique pour l'estimation de modèles linéaires mixtes généralisés pour une série chronologique à valeurs entières, ainsi que leur utilité pour la prévision des valeurs entières futures de la série. Nous obtenons les estimateurs par l'échantillonnage de Gibbs et nous comparons leur comportement par rapport aux autres méthodes en utilisant les données qui concernent la polio originalement analysées par Zeger (1988) et qui concernent le nombre de cas mensuels de polio aux États-Unis entre 1970 et 1983. Nous étudions également leurs propriétés par des simulations. Notre objectif est de démontrer avec quelle facilité la méthodologie bayésienne hiérarchique se prête à la vérification et à la comparaison de modèles. La méthodologie proposée, et plus particulièrement, la prévision bayésienne hiérarchique, est appliquée à une série de cas d'infection à la *Campylobacter* dans la région de Montréal-Centre.

Acknowledgments: The authors are grateful to the Bureau régional de maladies infectieuses, Régie régionale de la santé et des services-sociaux de Montréal-Centre for the data used here, and to the referees, whose comments greatly improved this manuscript. This work was supported through funds from the Natural Sciences and Engineering Research Council of Canada.

1 Introduction

Statistical research in modeling and inference for correlated and clustered longitudinal data of discrete outcomes such as count or binary data owes much of its development to Zeger and Liang (See Zeger, Liang, and Self 1985; Liang and Zeger 1986; Zeger 1988; Zeger and Qaqish 1988; Zeger and Karim 1991). If there are no explanatory variables, then integer-valued autoregressive moving average (INARMA) models previously studied by McKenzie (1988), Alzard and Al-Osh (1990), Du and Liu (1991), Latour (1997) and Belisle et al. (1998) could be used. However, in the study of disease incidence rates, nonstationarity and/or explanatory variables must often be taken into account. A useful model in such a context is the parameter-driven model defined in Cox (1981). Such a model is one in which there is an underlying unobservable latent process that influences the observed process. Estimation in parameter-driven models has previously been studied for discrete correlated data by many authors including Zeger (1988), Zeger and Qaqish (1988), Chan and Ledolter (1995), and Davis et al. (1997).

In fact, Zeger (1988) successfully modeled serially correlated count data with covariates by assuming that the observed counts were conditionally independent and Poisson distributed given the latent process. He assumed that this process was stationary and autoregressive; however he did not specify its distribution. Zeger (1988) used generalized estimating equation techniques and illustrated his method on a polio incidence series. Chan and Ledolter (1995) studied a similar parameter-driven model, assuming an underlying stationary Gaussian AR(1) latent process, used a Monte Carlo EM algorithm for parameter estimation and reanalyzed the polio incidence data of Zeger (1988). Kuk and Chen (1997) showed that the Monte Carlo Newton-Raphson algorithm is a viable alternative to the Monte Carlo EM algorithm of Chan and Ledolter (1995) for finding maximum likelihood estimates for this parameter-driven model of Poisson count data. They also reanalyzed the polio data.

Delampady et al. (1993) obtained Laplacian approximations to hierarchical Bayes estimators of a smooth but varying intensity function of a Poisson process by the analysis of a discrete series of Poisson counts. Jorgensen et al. (1995, 1996) used a state space non-stationary model for multivariate longitudinal count data driven by a latent gamma Markov process. They showed by reanalyzing the polio data of Zeger (1988) that an analysis such as theirs may differ substantially from one based on a stationary model.

Davis, Dunsmuir, and Wang (1997) showed that for a parameter-driven model, the Poisson maximum likelihood estimator of the regression parameter based on a model without serial correlation is consistent and asymptotically normal with an easily computable covariance matrix that depends on the covariance structure of the latent process. They also used an approximation to the likelihood in order to estimate the regression parameters and those of the latent process, after having tested for the existence of the latent process. They, as well as Fahrmeir et al. (1994), also reanalyzed the polio data in Zeger (1988).

Although estimation for a time series of counts has been fairly well studied, the prediction problem has not received the same consideration. There are some notable exceptions; however, mainly using state space modeling. We cite the work of Jorgensen et al. (1995, 1996, 1999) inspired by earlier work of Smith (1979), Key and Godolphin (1981), Smith and Miller (1986), West (1986), and West and Harrison (1989), which generalized the Bayesian forecasting model. Another interesting frequency approach assuming independence can be found in Li and Heckman (2003). Our approach to the prediction problem described below is a hierarchical Bayes one, which differs from the state space modeling approach.

In this study, we propose a hierarchical Bayes procedure using the Gibbs sampler for estimating the parameters in the same model as Chan and Ledolter (1995). We also extend the methodology so that future values can be predicted. The advantage to the hierarchical Bayes approach is the ease with which model selection can be performed using Bayes factors. Our choice of different priors in order to increase robustness has been influenced by the viewpoint of Berger (1984), and by previous work of Farrell et al. (1994). With the simple choice of a different prior, different modes can be compared. The proposed methodology is applied to the polio incidence series of Zeger (1988). Our estimation results are compared to those of the other authors mentioned above. Our results are also used to form the basis of a simulation study to investigate the properties of the estimators of the model parameters under the proposed approach.

We also wish to investigate the ability of the proposed estimation methodology to forecast for future unobserved time periods. This is an extremely important consideration in the epidemiological context when interest lies, for example, in the detection of the outbreak of an epidemic, or of a trend. Again, the Bayesian methodology allows us to easily obtain prediction estimates and intervals. An example of our method consisting of an analysis of a time series of infectious disease counts previously studied by Cardinal (1995) and Cardinal et al. (1999) using INARMA models is also given.

The paper is organized as follows. In Section 2, we describe the model in detail as well as the calculation of the hierarchical Bayes estimators of the parameters. The properties of these estimators are studied in Section 3. A simulation study is also included here. The data example is presented in Section 4, while a conclusion is given in Section 5.

2 The model

Let $Y_t, t = 1, \ldots, T$, be a time series of count data and x_t an associated covariate vector augmented by the constant one. We wish to model Y_t on x_t, as well as on its dependence on past Y values. Assuming that the Y_t are generated from a Poisson distribution with mean λ_t, we propose the following model:

$$Y_t \sim Poisson(\lambda_t),\ \log(\lambda_t) = x_t'\beta + \delta_t,\ \delta_t = \rho\delta_{t-1} + \varepsilon_t,\ \text{ where } \varepsilon_t \sim \text{ i.i.d. } N(0, \sigma^2). \quad (2.1)$$

Here β is a vector of parameters associated with the covariates in x_t and δ_t is a random effect to account for the dependence of the count data on its past values. Specifically, we

assume that $\delta_t, t = 1, \ldots, T$, is a stationary Gaussian AR(1) latent process with correlation coefficient ρ.

Estimation of the parameters in (2.1) can be accomplished via a hierarchical Bayes approach. This requires the specification of prior distributions for β, σ^2, and ρ. We consider a fairly diffuse proper version of an inverse gamma distribution for β and the random effect variance and a uniform prior for ρ. As Hobert and Casella (1996) have shown that the use of improper priors in hierarchical linear mixed models may lead to improper posteriors resulting in ill-behaved behaviour in the Gibbs sampler, the priors here have been adjusted here so that the posteriors are proper. Hence, if Y is a vector containing Y_t, then essentially the joint distribution of Y and the parameters in the prior distributions is proportional to

$$\prod_t \lambda_t^{y_t} e^{-\lambda_t} \frac{1}{\sigma^2} \prod_t \exp\Big(-\frac{1}{2} \sum_t \frac{(\delta_{t+1} - \rho\delta_t)^2}{\sigma^2} \Big) f(\rho) f(\beta) f(\sigma^2),$$

where $\log(\lambda_t) = \exp(x_t'\beta + \delta_t)$. Thus, with our choice of priors, the posterior can be shown to be proper by invoking Theorem 2 of Hobert and Casella (1996). Thus, Markov chain Monte Carlo (MCMC) would be appropriate here.

2.1 Hierarchical Bayes Estimates for the Model Parameters

Although software such as WinBUGS is now readily available, and the model proposed here could be implemented in it, we include here a short description of our hierarchical Bayes estimates, and the MCMC procedure we use.

To develop hierarchical Bayes estimates for the parameters in the model given in (2.1) requires posterior distributions of the model parameters. However, it is only possible to know these distributions up to a constant of proportionality (see Gilks, Best, and Tan 1995); specifically the posterior distribution for any given parameter is proportional to the product of all terms in the model that contain it. Therefore, for the model in (2.1), if δ is a vectors containing δ_t, then

$$f(\beta_0 \mid Y, \beta_1, \ldots, \beta_m, \delta, \sigma^2, \rho) \ \alpha \prod_t \lambda_t^{y_t} e^{-\lambda_t},$$

$$f(\beta_u \mid Y, \beta_0, \beta_1, \ldots, \beta_{u-1}, \beta_{u+1}, \ldots, \beta_m, \delta, \sigma^2, \rho) \ \alpha \prod_t \lambda_t^{y_t} e^{-\lambda_t},$$

$$f(\delta_t \mid Y, \beta, \delta_1, \ldots, \delta_{t-1}, \delta_{t+1}, \ldots, \delta_T, \sigma^2, \rho) \ \alpha \prod_t \lambda_t^{y_t} e^{-\lambda_t} \prod_t \exp\Big(-\frac{1}{2} \sum_t \frac{(\delta_{t+1} - \rho\delta_t)^2}{\sigma^2} \Big),$$

$$f(\sigma^2 \mid Y, \beta, \delta, \rho) \ \alpha \frac{1}{\sigma^{T+2}} \prod_t \exp\Big(-\frac{1}{2} \sum_t \frac{(\delta_{t+1} - \rho\delta_t)^2}{\sigma^2} \Big),$$

$$f(\rho \mid Y, \beta, \delta, \sigma^2) \ \alpha \prod_t \exp\Big(-\frac{1}{2} \sum_t \frac{(\delta_{t+1} - \rho\delta_t)^2}{\sigma^2} \Big),$$

where u refers to the u-th covariate, and m is the number of covariates.

Under Gibbs sampling, an initial set of values would be assumed as the estimates for β, δ, σ^2, and ρ, say $\hat{\beta}_{\{0\}}, \hat{\delta}_{\{0\}}, \hat{\sigma}^2_{\{0\}}$, and $\hat{\rho}_{\{0\}}$. An updated estimate for β_0, say $\hat{\beta}_{0\{1\}}$, is obtained by sampling from the full conditional distribution $f(\beta_0 \big| Y, \hat{\beta}_{1\{0\}}, \ldots, \hat{\beta}_{m\{0\}}, \hat{\delta}_{\{0\}}, \hat{\sigma}^2_{\{0\}}, \hat{\rho}_{\{0\}})$. Sampling from the full conditional distribution $f(\beta_1 \big| Y, \hat{\beta}_{0\{1\}}, \hat{\beta}_{2\{0\}}, \ldots, \hat{\beta}_{m\{0\}}, \hat{\delta}_{\{0\}}, \hat{\sigma}^2_{\{0\}}, \hat{\rho}_{\{0\}})$ based on $\hat{\beta}_{0\{1\}}$ yields the revised estimate $\hat{\beta}_{1\{1\}}$ for β_1. The completion of a first iteration is realized once the revised estimates $\hat{\beta}_{\{1\}}, \hat{\delta}_{\{1\}}, \hat{\sigma}^2_{\{1\}}$, and $\hat{\rho}_{\{1\}}$ are obtained. This procedure of sampling from full conditional distributions using the most up-to-date revised estimates continues until the estimates of each parameter are deemed to have stabilized from one iteration to the next. See Geman and Geman (1984) and Gelfand and Smith (1990) for a general discussion on Gibbs sampling, and Gelman and Rubin (1992) for methods of convergence.

Note that a different full conditional distribution must be sampled every time a new estimate is obtained, regardless of which parameter is being estimated. Since many iterations are usually needed to ensure that estimates for each parameter have stabilized, efficient methods for constructing full conditional distributions and sampling from them are required. For log-concave distributions, this can be accomplished through adaptive rejection sampling (See Gilks and Wild, 1992). For applications where the full conditional distributions are not log-concave, Gilks, Best, and Tan (1995) propose appending a Hasting-Metropolis algorithm step to the adaptive rejection sampling scheme. They suggest using the resulting adaptive rejection Metropolis sampling scheme within the Gibbs sampling algorithm. We follow this approach here.

3 Properties of Model Parameter Estimators

In order to study the properties of the estimators of the parameters in the model given by (2.1), we make use of the data in Zeger (1988), consisting of the monthly number of cases of poliomyelitis from January 1970 to December 1983. This data has been analyzed by Zeger (1988), Fahrmeir and Tutz (1994), Chan and Ledolter (1995), Jorgensen et al. (1995), Davis et al. (1997), and Kuk and Chen (1997).

A central question studied by Zeger (1988) is whether the number of cases decrease over time. A time series plot of the data in Chan and Ledolter (1995) shows that there appears to be seasonality present. Similar to Zeger (1988) and Chan and Ledolter (1995), we model the seasonality with trigonometric components involving the first two harmonics. Thus, the covariate vector here is given by

$$x'_t = [1, t^*/1000, \cos(2\pi t^*/12), \sin(2\pi t^*/12), \cos(2\pi t^*/6), \sin(2\pi t^*/6)],$$

where $t^* = t - 73$. Our model is similar to that of Chan and Ledolter (1995). It differs from Zeger (1988) in that a Gaussian distribution is assumed for the ε_t. In addition to considering the central question studied by Zeger (1988), we also wish to investigate the

ability of the proposed model to forecast for future unobserved time periods. This is an extremely important consideration in the epidemiological context when interest lies, for example, in the detection of the outbreak of an epidemic. Therefore, to study the predictive ability of the proposed methodology, we hold out the final three observations and only make use of the first 165 observations in the polio data set to fit the model. Point and interval estimates for the last three observations that are based on the fitted model can then be determined and compared with their known true values. We explore this prediction problem later in some detail.

The procedure employed by Gilks, Best, and Tan (1995) was used to fit the model in (2.1) to the data. Specifically, the Gibbs sampler was run for 15,000 iterations twice, each with a different set of starting values for the parameter estimates. Initially, the full conditional distributions for the estimator of each parameter were evaluated at six different values for the parameter that were based on the 5^{th}, 30^{th}, 45^{th} 55^{th}, 70^{th}, and 95^{th} percentiles of a piecewise linear function from the previous Gibbs iteration (See Gilks, Best, and Tan 1995). The method of Gelman and Rubin (1992) was used to assess convergence of the Gibbs sampler. To ensure proper convergence, only the last 3,000 iterations of each of the two runs were used to construct posterior distributions. In particular, the results over these two sets of 3,000 iterations were combined in order to approximate these distributions.

The results of the model fit using the proposed approach are presented in Table 1(a). It appears that the $\sin(2\pi t^*/6)$ term may not be needed in the model. To investigate this further, a Bayes factor (see Congdon 2005) was computed to compare models with and without this term, assuming a priori that each model was equally likely. A value of 1.06 was obtained for this factor, suggesting that the more parsimonious model is supported. Parameter estimates determined under this simpler model are provided in Table 1(b). In addition, note that an analogous Bayes factor of 0.23 was obtained when the $\cos(2\pi t^*/12)$ term was deleted from the model without the $\sin(2\pi t^*/6)$ term; thus, further removal of terms was deemed to be inappropriate.

As a result of our interest in robustness influenced by Berger (1984) and Farrell et al. (1994), we also obtained parameter estimates for the model under the assumption that the random effects follow a Laplace distribution. Starting with the covariate vector x_t' as given above, the model fit suggested that the $\sin(2\pi t^*/6)$ term may not be needed. A comparison of models with and without this term yielded a Bayes factor of 1.21. Parameter estimates for the model without the $\sin(2\pi t^*/6)$ term are presented in Table 1(c). Further removal of covariates was not supported by the appropriate Bayes factor.

A comparison of the model estimates in Tables 1(b) and 1(c) suggest that the results are extremely similar under the two different distributions specified for the random effects. However, the terms in the model are somewhat more statistically significant when the random effects are assumed to follow a Laplace distribution. For example, the estimate of the trend covariate $(t^*/1000)$ is -4.88 with a standard error of 1.05 under a Laplace

Table 1(a): Hierarchical Bayes model estimates (estimated standard errors in brackets) for the polio data time series based on a normal distribution for the random effects, where seasonality is modeled with trigonometric components involving the first two harmonics.

Intercept	$t^*/1000$	$\cos(2\pi t^*/12)$	$\sin(2\pi t^*/12)$	$\cos(2\pi t^*/6)$	$\sin(2\pi t^*/6)$	σ^2	ρ
0.37	-4.82	0.15	-0.52	0.45	-0.07	0.50	0.92
(0.12)	(1.19)	(0.08)	(0.10)	(0.09)	(0.09)	(0.19)	(0.13)

Table 1(b): Hierarchical Bayes model estimates (estimated standard errors in brackets) for the polio data time series based on a normal distribution for the random effects, where seasonality is modeled with trigonometric components involving the first two harmonics, excluding the $\sin(2\pi t^*/6)$ term.

Intercept	$t^*/1000$	$\cos(2\pi t^*/12)$	$\sin(2\pi t^*/12)$	$\cos(2\pi t^*/6)$	σ^2	ρ
0.39	-4.76	0.14	-0.58	0.48	0.56	0.89
(0.10)	(1.12)	(0.08)	(0.09)	(0.08)	(0.21)	(0.14)

Table 1(c): Hierarchical Bayes model estimates (estimated standard errors in brackets) for the polio data time series based on a Laplace distribution for the random effects, where seasonality is modeled with trigonometric components involving the first two harmonics, excluding the $\sin(2\pi t^*/6)$ term.

Intercept	$t^*/1000$	$\cos(2\pi t^*/12)$	$\sin(2\pi t^*/12)$	$\cos(2\pi t^*/6)$	σ^2	ρ
0.42	-4.88	0.15	-0.60	0.49	0.55	0.91
(0.08)	(1.05)	(0.08)	(0.08)	(0.07)	(0.18)	(0.13)

distribution, and -4.76 with a standard error of 1.12 when the random effects are assumed normal.

Table 2 presents a comparison of our estimates of the parameter for the trend covariate ($t^*/1000$) with analogous estimates published in Davis et al. (1997) that were previously obtained by Zeger (1988), Fahrmeir and Tutz (1994), Chan and Ledolter (1995), Jorgensen et al. (1995), Davis et al. (1997), and Kuk and Chen (1997). Our results for all parameter estimates are quite similar to those of Chan and Ledolter (1995); however the standard errors for the proposed hierarchical Bayes approach used here are uniformly smaller than those obtained by Chan and Ledolter (1995). We obtain the same negative trend estimate as Davis et al. (1997), but with a smaller estimated standard error. In addition, as mentioned above, the standard errors obtained under the Laplace distribution are slightly smaller than those for the normal.

For statistical inference, a good indicator of the performance of an estimator is its coverage probability. To study the properties of the estimators of the model parameters under normal and Laplace distributions for the random effects, we generated 500 simulated

Table 2: Comparison of the parameter estimates of trend ($t^*/1000$) for the polio data time series.

Study	Trend Estimate	Standard Error	t-ratio
Zeger (1988)	-4.35	2.68	-1.62
Fahrmeir and Tutz (1994)	-3.33	2.00	-1.67
Chan and Ledolter (1995)	-4.62	1.38	-3.35
Jorgensen et al. (1995)	-1.64	0.02	-91.1
Davis et al. (1997)	-4.80	1.40	-3.43
Kuk and Chen (1997)	-3.79	2.95	-1.28
Farrell et al. (2005) Nor	-4.76	1.12	-4.25
Farrell et al. (2005) Lap	-4.88	1.05	-4.65

data sets analogous to the original polio data for each distribution in the following manner. For a given distribution, observations were generated for a particular data set by treating the estimates obtained for σ^2, ρ, and the vector β in Table 1 as true parameter values. Random effects were generated according to an AR(1) process based on the estimate for ρ, and the estimated prior distribution for the random effects. More specifically, the following procedure was used:

1) Set $\delta_0 = 0$.
2) Generate a value for ε_1 from a normal *or Laplace* distribution with mean zero and variance given by the estimate of σ^2.
3) Compute δ_1 using the relation $\delta_1 = \rho\delta_0 + \varepsilon_1$, where the value used for ρ is the estimate of the AR(1) correlation coefficient.
4) Compute λ_1 using $\log(\lambda_1) = x_1'\beta + \delta_1$, where the values used for the components of β are the estimates obtained for β under the hierarchical Bayes approach.
5) Generate a value for Y_1 from a Poisson distribution with mean λ_1.
6) Repeat steps (2) through (5) until observations $Y_2, \ldots, Y_T$ have been obtained.

A summary of the results over the 500 simulated data sets for each random effects distribution is presented in Tables 3(a) and 3(b). Regardless of the distribution assumed for the random effects, there is little bias in the estimators for β. Nevertheless, the means of the estimators computed over the 500 simulated data sets are slightly closer to the true parameter values when a Laplace distribution is used for the random effects. The average model-based standard errors for the estimators of the fixed effect parameters over the 500 simulated data sets are also presented in Tables 3(a) and 3(b), along with the associated empirical root mean square errors. A comparison of these standard errors with the empirical root mean square errors indicate that the former serve to appropriately describe the variability in the estimators of β from sample to sample. Also of note is the fact that, generally speaking, the Laplace prior provides more efficient estimators, in particular for the trend covariate. To explore these notions further, for both normal and Laplace priors, 500 normal symmetric confidence intervals were computed for each fixed effect

Table 3(a): Mean parameter estimates, mean standard errors, empirical root mean square errors and coverage rates for 90%, 95%, and 99% nominal rates over the 500 simulated data sets. Estimates are based on a model without $\sin(2\pi t^*/6)$ and a normal distribution for the random effects.

					Coverage		
Term	True Value	Mean	Mean SE	RMSE	90%	95%	99%
Intercept	0.39	0.402	0.114	0.108	88.4	93.0	97.8
$t^*/1000$	-4.76	-4.64	1.223	1.318	90.6	95.8	99.4
$\cos(2\pi t^*/12)$	0.14	0.136	0.079	0.092	90.8	95.2	98.8
$\sin(2\pi t^*/12)$	-0.58	-0.601	0.121	0.109	89.4	93.6	98.0
$\cos(2\pi t^*/6)$	0.48	0.466	0.082	0.094	92.6	97.4	99.6
σ^2	0.56	0.627	0.241	0.255	91.4	96.6	99.2
ρ	0.89	0.834	0.166	0.152	89.2	94.4	98.0
Y_{166}	1	1.216	0.385	0.502	92.8	97.4	99.8
Y_{167}	3	2.734	0.617	0.783	93.2	98.0	100.0
Y_{168}	6	5.376	1.079	1.304	93.8	98.2	100.0

Table 3(b): Mean parameter estimates, mean standard errors, empirical root mean square errors and coverage rates for 90%, 95%, and 99% nominal rates over the 500 simulated data sets. Estimates are based on a model without $\sin(2\pi t^*/6)$ and a Laplace distribution for the random effects.

					Coverage		
Term	True Value	Mean	Mean SE	RMSE	90%	95%	99%
Intercept	0.42	0.427	0.094	0.089	89.0	93.8	97.8
$t^*/1000$	-4.88	-4.81	1.110	1.097	90.2	95.4	99.4
$\cos(2\pi t^*/12)$	0.15	0.142	0.080	0.087	90.6	95.2	98.8
$\sin(2\pi t^*/12)$	-0.60	-0.619	0.122	0.114	89.6	94.0	98.0
$\cos(2\pi t^*/6)$	0.49	0.469	0.079	0.085	93.0	96.6	99.6
σ^2	0.55	0.574	0.232	0.237	91.4	96.2	99.2
ρ	0.91	0.877	0.157	0.149	89.2	94.6	98.0
Y_{166}	1	1.144	0.378	0.448	92.0	96.4	99.6
Y_{167}	3	2.857	0.598	0.702	92.4	96.8	99.6
Y_{168}	6	5.602	1.033	1.184	93.0	97.2	99.8

parameter at three different levels of confidence: 90%, 95%, and 99%. The coverage rates for these intervals were ascertained; they are also presented in Tables 3(a) and 3(b). The results for each prior indicate that, regardless of the parameter being estimated or the level of confidence being considered, the normal symmetric confidence intervals approximately attain the desired level of coverage. However, it is also worth mentioning that the coverage rates based on the Laplace prior are noticeably closer to the nominal rates.

The models based on normal and Laplace distributions were also used to forecast polio counts for the last three periods, which were held out of the data used for estimation. The mean predictions computed over the 500 simulated data sets are presented in Tables 3(a) and 3(b), respectively, along with the true counts. It can be seen from these results that the Laplace prior fares better with regards to point estimation. Also of note is that the estimators for the true count values are more efficient under a Laplace prior. In addition, confidence intervals based on these estimators are seem to better achieve the nominal rate when compared to those based on a normal prior.

4 Data Example

Epidemiological surveillance is an ongoing procedure that consists of collecting, analyzing, and interpreting public health data for disease and prevention control programs. Infectious disease incidence, rendered available through epidemiologic surveillance, may have irregular values in different geographical areas, time periods, or demographic characteristics. In the dimension of time, irregular disease incidence values can provide an early indication of an upcoming epidemic. Hence, the detection of irregular disease incidence values is of importance for public health programs since government officials would be in a position to react and intervene promptly.

As shown by Cardinal et al. (1999), integer-valued time series models can play an important role in the analysis and interpretation of infectious disease incidence data. From a public health perspective, it is important to be able to detect trends as well as other more irregular patterns in the data. Specifically, it is extremely important to be able to identify high incidences that may foretell an upcoming epidemic.

Although the success of the integer-valued time series was shown in Cardinal et al. (1999), it would also be very useful for epidemiologists in the public health sector to be able to incorporate covariates into statistical models of disease incidence. The hierarchical Bayes approach described here allows for the incorporation of such covariates.

We apply the proposed hierarchical Bayes methodology to a time series consisting of the number of cases of *Campylobacter* infection in the Montreal-Center region from January 1986 to December 1993. Of interest is the detection of periods of irregular disease incidence values, in particular unusually high values. According to Cardinal (1995), *Campylobacter* infection is an "acute enteric disease characterized by diarrhea, abdominal pain, general malaise, fever, nausea, and vomiting". For each year that data are available, the number of cases were recorded over thirteen periods of approximately twenty eight days. The first twelve periods for each year consisted of twenty eight days, the last period of the remaining number of days making up the year.

A time series plot of the *Campylobacter* infection data is given in Figure 1. There appears to be a positive linear trend and a seasonal component to the series, where the number of cases of the disease is higher during the summer months. We therefore apply the

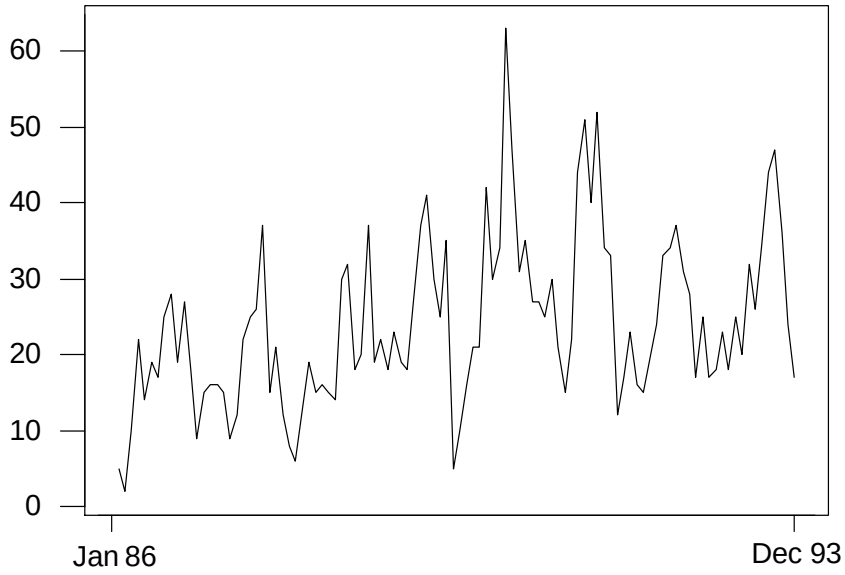

Figure 1: Time Series Plot of *Campylobacter* Data

proposed hierarchical Bayes approach to fit the model given by (2.1), with the following covariate vector

$$x_t' = [1, t/1000, \cos(2\pi t/13), \sin(2\pi t/13), \cos(2\pi t/6.5), \sin(2\pi t/6.5)]$$

where $t = 1, \ldots, 104$. Similar to the fit for the polio data, we model the seasonality with trigonometric components involving the first two harmonics.

The procedure for obtaining estimates of the model parameters was identical to that described above for the polio data fit. The final three observations were removed so that predictions could be made for these time periods. Separate fits based on normal and Laplace priors were considered. In both cases, the use of Bayes factors suggested that the $\sin(2\pi t/6.5)$ term was not significant. However, once this term was removed all other covariates were deemed necessary. Tables 4(a) and 4(b) present the estimates for the fixed effects parameter estimates. Regardless of which prior is considered, the results suggest that there is a significant linear positive trend to the *Campylobacter* infection data, as well as a significant seasonal component. This is illustrated by the fact that the coefficients for the first two terms in the trigonometric series are highly significant. Also of note is the fact that standard errors obtained under a Laplace prior are, generally speaking, smaller than those arrived at when a normal prior is assumed for the random effects.

In order to detect a period with an unusually high disease incidence value, the ratio of the estimate of the random effect for the period to its standard error can be considered. We will consider here an unusually high value for a time period as being reflected by a

Table 4(a): Hierarchical Bayes model estimates (estimated standard errors is brackets) for the *Campylobacter* data time series based on a normal distribution for the random effects, where seasonality is modeled with trigonometric components involving the first two harmonics, excluding the $\sin(2\pi t^*/6.5)$ term. Point estimates and prediction intervals are presented for the last three observations in the time series, which were not used in the fit.

Intercept	$t/1000$	$\cos(2\pi t/13)$	$\sin(2\pi t/13)$	$\cos(2\pi t/6.5)$	σ^2	ρ
2.89	5.97	-0.24	-0.33	-0.07	0.59	0.88
(0.042)	(0.669)	(0.029)	(0.027)	(0.029)	(0.15)	(0.12)

Term	True Value	Estimate	95% Prediction Interval
Y_{102}	36	34.30	(27.07, 41.88)
Y_{103}	24	25.63	(20.23, 31.71)
Y_{104}	17	19.78	(11.20, 28.45)

Table 4(b): Hierarchical Bayes model estimates (estimated standard errors is brackets) for the *Campylobacter* data time series based on a Laplace distribution for the random effects, where seasonality is modeled with trigonometric components involving the first two harmonics, excluding the $\sin(2\pi t^*/6.5)$ term. Point estimates and prediction intervals are presented for the last three observations in the time series, which were not used in the fit.

Intercept	$t/1000$	$\cos(2\pi t/13)$	$\sin(2\pi t/13)$	$\cos(2\pi t/6.5)$	σ^2	ρ
2.95	6.04	-0.24	-0.34	-0.08	0.51	0.89
(0.039)	(0.666)	(0.028)	(0.027)	(0.027)	(0.14)	(0.10)

Term	True Value	Estimate	95% Prediction Interval
Y_{102}	36	34.77	(28.03, 40.69)
Y_{103}	24	25.12	(20.78, 31.17)
Y_{104}	17	19.32	(12.91, 26.88)

ratio of random effect estimate to associated standard error of +3 or more. According to this criterion, regardless of whether a normal or Laplace prior is assumed for the random effects, there were six time periods with unusually high values: 39, 51, 57, 60, 71, and 72. Of note is the fact that these periods, with one exception, seem to be either in the month of December or between mid-May and mid-July. In addition, there does not appear to be any unusually high disease incidence values over the last two-and-a-half years of the series. Nevertheless, the significant positive trend in the model indicates that, generally speaking, cases of *Campylobacter* infection are on the rise.

Finally, Tables 4(a) and 4(b) also present point estimates and 95% symmetric normal prediction intervals for the last three observations in the *Campylobacter* data set. The Laplace prior appears to fare better in this regard. Point estimates obtained under this prior are slightly closer to the true count. In addition, even though all prediction intervals cover the true value for both normal and Laplace priors, those obtained under the latter are somewhat narrower.

5 Conclusion

We have proposed and illustrated a hierarchical Bayes approach to parameter estimation in a model to describe time series count data. We have focused on the AR(1) parameter-driven model as in Chan and Ledolter (1995), although other models of dependence could be easily implemented. Our method gives similar estimates to those of others for the trend component in models for the polio data in Zeger (1988). The performance of our estimators, as measured by the coverage probability of confidence intervals, is very good. Prediction for future time periods is also possible. Such a property is important in epidemiological applications where interest may lie in the detection of disease outbreaks. The proposed methodology was also successfully applied to a time series of Campylobacter disease counts. Alternative models of the correlation structure, as well as other priors could be pursued as a topic for future research. In addition, since estimated standard errors for the estimator of ρ are available, these might ultimately be used to derive a test for the hypothesis of independence of the random effects. The development of such an inferential approach could be another interesting possibility for future research.

In conclusion, the Bayesian methodology has been particularly helpful here. We have shown how easily Bayes factors can be applied to check the significance of model covariates, and how much improvement in performance of estimators and prediction can be obtained by using a robust Bayesian approach. In addition, the proposed hierarchical Bayes methodology should have a broad application to other statistical longitudinal data problems, including multivariate count data.

References

Alzard, A.A., and Al-Osh, M.A. (1990), "An Integer-Valued pth Order Autoregressive Structure (INAR(p)) Process", *Journal of Applied Probability*, 27, 314–324.

Belisle, P., Joseph, L., MacGibbon, B., Wolfson, D.B., and du Berger, R. (1998), "Change Point Analysis of Neural Spike Train Data", *Biometrics*, 54, 113–123.

Berger, J.O. (1984), "The Robust Bayesian Viewpoint (with Discussion)", in *Robustness in Bayesian Statistics*, edited by J. Kadane, Amsterdam: North Holland.

Cardinal, M. (1995), "Modélisation temporelle d'incidence de maladies", Mémoire de maîtrise, Université de Montréal.

Cardinal, M., Lambert, J., and Roy, R. (1999), "On the Application of Integer-Valued Time Series Models for the Analysis of Disease Incidence", *Statistics in Medicine*, 18, 2025–2039.

Chan, K.S., and Ledolter, J. (1995), "Monte-Carlo EM Estimation for Time Series Models Involving Counts", *Journal of the American Statistical Association*, 90, 242–252.

Congdon, P. (2005), *Bayesian Models for Categorical Data*, Toronto: Wiley.

Cox, D.R. (1981), "Statistical Analysis of Time Series: Some Recent Developments", *Scandinavian Journal of Statistics*, 8, 93–115.

Davis, R.A., Dunsmuir, W.T.M., and Wang, Y. (1997), "Modeling Time Series of Count Data", Technical Report.

Delampady, M., Yee, I.M.L., and Zidek, J.V. (1993), "Hierarchical Bayesian Analysis of a Discrete Time Series of Counts", *Statistics and Computing*, 3, 7–15.

Du, J.G., and Liu, Y. (1991), "The Integer-Valued Autoregressive (INAR(p)) Model", *Journal of Time Series Analysis*, 12, 129–142.

Fahrmeir, L., and Tutz, G. (1994), *Multivariate Statistical Modeling Based on Generalized Linear Models*, New York: Springer-Verlag.

Farrell, P.J., MacGibbon, B., and Tomberlin, T.J. (1994), "Protection Against Outliers in Empirical Bayes Estimation", *Canadian Journal of Statistics*, 22, 365–376.

Gelfand, A.E., and Smith, A.F.M. (1990), "Sampling-Based Approaches to Calculating Marginal Densities", *Journal of the American Statistical Association*, 85, 398-409.

Gelman, A., and Rubin, D.B. (1992), "Inference from Iterative Simulation Using Multiple Sequences", *Statistical Science*, 7, 457–472.

Geman, S., and Geman, D. (1984), "Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images", *IEEE Trans. Pattn Anal. Mach. Intell.*, 6, 721–741.

Gilks, W.R., Best, N.G., and Tan, K.K.C. (1995), "Adaptive Rejection Metropolis Sampling Within Gibbs Sampling", *Applied Statistics*, 44, 455–472.

Gilks, W.R., and Wild, P. (1992), "Adaptive Rejection Sampling for Gibbs Sampling", *Applied Statistics*, 41, 337–348.

Hobert, J.P., and Casella, G. (1996), "The Effect of Improper Priors on Gibbs Sampling in Hierarchical Linear Mixed Models", *Journal of the American Statistical Association*, 91, 1461–1473.

Jorgensen, B., Lundbye-Christensen, S., Song, X.K., and Sun, L. (1995), "A State Space Model for Multivariate Longitudinal Count Data of Mixed Types", Technical Report #148, Department of Statistics, University of British Columbia.

Jorgensen, B., Lundbye-Christensen, S., Song, X.K., and Sun, L. (1996), "State Space Models for Multivariate Longitudinal Count Data of Mixed Types", *Canadian Journal of Statistics*, 24, 385–402.

Jorgensen, B., Lundbye-Christensen, S., Song, X.K., and Sun, L. (1999), "A State Space Model for Multivariate Longitudinal Count Data", *Biometrika*, 86, 169–181.

Key, P.B., and Godolphin, E.J. (1981), "On the Bayesian Steady Forecasting Model", *Journal of the Royal Statistical Society, Series B*, 43, 92–96.

Kuk, A.Y.C., and Cheng, Y.W. (1997), "The Monte Carlo Newton-Raphson Algorithm", *Journal of Statistical Computation and Simulation*, 59, 233–250.

Latour, A. (1997), "The Mulitvariate GINAR(p) Process", *Advances in Applied Probability*, 29, 228–248.

Li, X., and Heckman, N.E. (2003), "Local Linear Extrapolation", *Nonparametric Statistics*, 15, 565–578.

Liang, K.Y., and Zeger, S.L. (1986), "Longitudinal Data Analysis Using Generalized Linear Models, *Biometrika*, 73, 13–22.

McKenzie, E. (1988), "Some ARMA Models for Dependent Sequences of Poisson Counts", *Advances in Applied Probability*, 20, 822–835.

Smith, J.Q. (1979), "A Generalization of the Bayesian Steady Forecasting Model", *Journal of the Royal Statistical Society, Series B*, 41, 375–387.

Smith, R.L., and Miller, J.E. (1986), "A Non-Gaussian State Space Model and Application to Prediction of Records", *Journal of the Royal Statistical Society, Series B*, 48, 79–88.

West, M. (1986), "Bayesian Model Monitoring", *Journal of the Royal Statistical Society, Series B*, 48, 70–78.

West, M., Harrison, P.J., and Migon, H.S. (1985), "Dynamic Generalized Linear Models and Bayesian Forecasting", *Journal of the American Statistical Association*, 80, 73–97.

Zeger, S.L. (1988), "A Regression Model for Time Series of Counts", *Biometrika*, 75, 621–629.

Zeger, S.L., Liang, K.Y., and Self, S.G. (1985), "The Analysis of Binary Longitudinal Data with Time Independent Covariates", *Biometrika*, 72, 31–38.

Zeger, S.L., and Karim, M.R. (1991), "Generalized Linear Models with Random Effects: A Gibbs Sampling Approach", *Journal of the American Statistical Association*, 86, 79–86.

Zeger, S.L., and Qaqish, B. (1988), "Markov Regression Models for Time Series: A Quasi-Likelihood Approach", *Biometrics*, 44, 1019–1033.