ELSEVIER

Journal of Statistical Planning and Inference 47 (1995) 129–139

journal of statistical planning and inference

Dynamic generalized linear models and repeated measurements

J.K. Lindsey*, P. Lambert

Fac. d'Economie, de Gestion et de Sciences Sociales, Universite de Liege, Sart Tilman B31, 4000 Liege, Belgium

Abstract

The dynamic generalized linear model for non-normal data is extended for use in repeated measurements, when series of observations are available for more than one individual. Examples are given for count and duration data.

Key words: Autoregression; Dynamic generalized linear model; Kalman filter; Longitudinal data; Overdispersion; Repeated measurements; State space model

1. Introduction

1.1. Motivation

Generalized linear regression models are widely used for univariate data analysis. However, nothing equivalent is available for multivariate situations, such as repeated measurements. Such data are frequently encountered in the longitudinal context, especially in the form of count data. Below, we consider an example of monthly counts of deaths from a given disease. When the evolution of such counts is to be compared for different subpopulations, we are in a repeated measurements situation. But, such models are also useful to study profiles of counts on individual people, for example, the number of infections or of epilepsy attacks in given periods of time.

A second area of application is to longitudinal studies of positive-valued data. If the response values are fairly close to zero, one may expect their distribution to be skewed. The common solution is to take logarithms, i.e. to use a log normal distribution. However, in many situations, some other survival-type distribution, such as the

* Corresponding author.

SSDI 0378-3758X(94)00126-X

gamma or inverse Gaussian, may be more appropriate. Below, we look at an example of the evolution of plasma citrate concentrations.

Lindsey (1993, Chs. 6 and 8) provides a comprehensive discussion of various modelling approaches to such non-normal longitudinal data and includes a comprehensive bibliography which the reader may wish to consult.

One of the most useful approaches to such longitudinal data, similar in some ways both to autoregression and to random effects, and in fact able to encompass both, is to have the coefficients of the regression equations evolve over time according to a Markov process. This is called a dynamic generalized linear model and is usually estimated by a procedure called the Kalman filter. It was originally proposed as the dynamic linear model for single series of normal data, but has more recently been used for repeated measures (Jones and Ackerson, 1990; Jones and Boadi-Boateng, 1991; Schlain et al., 1992; Wilson, 1988). Although it has been extended to other distributions, as the dynamic *generalized* linear model, for single time series, notably by West et al. (1985), Kitagawa (1987), Fahrmeir (1989, 1992), Harvey (1989) and Harvey and Fernandes (1989), few applications of such models to repeated measurements have been proposed. That is the object of the present paper. For a somewhat different approach, see the recent paper by Singh and Roberts (1992).

1.2. The dynamic generalized linear model

Suppose some given quantity Y is measured repeatedly at equally spaced times ($t = 0, 1, 2, \ldots$) on several units ($i = 1, 2, \ldots, n$) together with a vector of covariates $\boldsymbol{x}$. We assume that the observations across units are independent. Then the linear regression model, now called the *observation* or *measurement* equation, is

$$g(u_{it}) = \lambda_{it} + \boldsymbol{\beta}_{it}^{\mathrm{T}} \boldsymbol{x}_{it}$$

where μ_{it} denotes the mean of Y_{it}, $g(\cdot)$ is some link function, $\boldsymbol{\beta}_{it}$ a possibly random vector of regression coefficients, λ_{it} the random 'base' mean. $(\lambda_{it}, \boldsymbol{\beta}_{it})^{\mathrm{T}}$ is called the *state vector*.

That approach would be equivalent to generalized linear models if it were not for the dependence between observations on the same unit. Instead of directly modelling that dependence (by imposing some covariance structure for example) we propose a first-order Markov chain model for the state vector

$$E\left[\begin{pmatrix} \lambda_{it} \\ \boldsymbol{\beta}_{it} \end{pmatrix}\right] = \boldsymbol{T}_{it} \begin{pmatrix} \lambda_{i,t-1} \\ \boldsymbol{\beta}_{i,t-1} \end{pmatrix}$$

called the *state transition* equation where $\boldsymbol{T}_{it}$ is the first-order Markovian state transition matrix, assumed to be known for all t.

Notice that this structure allows for heterogeneity across units, an important feature in repeated measurements modelling.

As an illustration consider the dynamic generalized linear model for an autoregression of order M. It has measurement and state transition equation

$$g(\mu_{it}) = [1,\ 0, \ldots, 0]\begin{pmatrix} \lambda_{i,t-1} \\ \boldsymbol{\beta}_{i,t-1} \end{pmatrix},$$

$$E\left[\begin{pmatrix} \lambda_{it} \\ \beta_{it} \\ \vdots \\ \beta_{i,t-M+2} \end{pmatrix}\right] = \begin{pmatrix} \rho_{i1} & \cdots & \rho_{i,M-1} & \rho_{iM} \\ 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 \end{pmatrix}\begin{pmatrix} \lambda_{i,t-1} \\ \beta_{i,t-1} \\ \vdots \\ \beta_{i,t-M+1} \end{pmatrix}.$$

Another simple DGLM is the random effects model. The equations are

$$g(\mu_{ikt}) = \mu_k + \lambda_{it},$$

$$E[\lambda_{it}] = 0,$$

where μ_{ikt} denotes the mean of the kth measure on the ith unit. Simple models, such as these, can be combined in any desired way.

West et al. (1985), West and Harrison (1989), Harvey (1989) and Harvey and Fernandes (1989) use the canonical link for all response distributions, so that the distribution of the random coefficients can be conjugate. Kitagawa (1987) approximates the probability by a piecewise linear function, a form of numerical integration.

Now arises the problem of inference. Denote by $\mathscr{F}_{it}$ the history of the responses for unit i up to and including time t. The problem that we are now facing is the estimation of $\boldsymbol{\beta}_{it}$ given $\mathscr{F}_{it}$. The Bayesian approach provides a satisfactory answer to this. If $p(\lambda_{it}, \boldsymbol{\beta}_{it} | \mathscr{F}_{it})$, $\Pr(y_{it} | \lambda_{it}, \boldsymbol{\beta}_{it}, \mathscr{F}_{i,t-1})$ and $\Pr(y_{it} | \mathscr{F}_{i,t-1})$ denote, respectively, the prior distribution of $(\lambda_{it}, \boldsymbol{\beta}_{it})^{\mathrm{T}}$, the conditional distribution of $(\lambda_{it}, \boldsymbol{\beta}_{it})^{\mathrm{T}}$ and the marginal distribution of $(\lambda_{it}, \boldsymbol{\beta}_{it})^{\mathrm{T}}$ before observing y_{it}, then the posterior distribution of the state vector is (according to Bayes theorem)

$$p(\lambda_{it}, \boldsymbol{\beta}_{it} | \mathscr{F}_{it}) = \frac{\Pr(y_{it} | \lambda_{it}, \mathscr{F}_{i,t-1}) p(\lambda_{it}, \boldsymbol{\beta}_{it} | \mathscr{F}_{i,t-1})}{\Pr(y_{it} | \mathscr{F}_{i,t-1})},$$

where

$$p(\lambda_{it}, \boldsymbol{\beta}_{it} | \mathscr{F}_{i,t-1}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(\lambda_{it}, \boldsymbol{\beta}_{it} | \lambda_{i,t-1}, \boldsymbol{\beta}_{i,t-1}) \times p(\lambda_{i,t-1}, \boldsymbol{\beta}_{i,t-1} | \mathscr{F}_{i,t-1})\, \mathrm{d}\lambda_{it}\, \mathrm{d}\boldsymbol{\beta}_{i,t-1}.$$

Note that the last expression can be used to compute predictions for Y_{it} given past observations $\mathscr{F}_{i,t-1}$.

To avoid intractability and time-consuming numerical calculations, one usually chooses the conjugate distribution for $(\lambda_{it}, \boldsymbol{\beta}_{it})^{\mathrm{T}}$. For instance one would take a gamma prior for λ_{it} if Y_{it} is a Poisson r.v., or a beta prior if Y_{it} is a binomial r.v. We shall develop the Poisson example in the coming sections.

1.3. Trend and seasonality

Very often data such as repeated counts show a trend pattern mixed with seasonal effects. An example of this is provided by the monthly numbers of deaths from bronchitis, emphysema and asthma in the UK (see next section).

Thus it is important to have a simple tool for modelling such fundamental behaviour. Basically this is performed by a judicious choice for the Markov transition matrix $\boldsymbol{T}_{it}$ in the state transition equation.

For clarity we shall consider two special cases:

- Suppose y_{it} can be viewed as a piecewise linear process. Then a possible model for Y_{it} could be

$$\mu_{it} = [1,\ 0]\begin{pmatrix}\lambda_{it}\\ \boldsymbol{\beta}_{it}\end{pmatrix},$$

$$E\left[\begin{pmatrix}\lambda_{it}\\ \beta_{it}\end{pmatrix}\right] = \begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix}\begin{pmatrix}\lambda_{i,t-1}\\ \beta_{i,t-1}\end{pmatrix}.$$

λ_{it} is to be interpreted as the current intercept for y_{it} and β_{it} as the local 'trend' for y_{it}. Low values for the discount (see below) will allow for quick changes in both λ_{it} and β_{it}.

- Now suppose that y_{it} only displays seasonal patterns. That hypothesis can be satisfied by first removing the trend (using the above procedure) or any systematic effect (modelled by including regressors in $\boldsymbol{T}_{it}$ for example). Suppose for simplicity that we have monthly data ($p = 12$, $\omega = 2\pi/p$). Then a possible DGLM is

$$\mu_{it} = \boldsymbol{\beta}_{it}^{\mathrm{T}}\boldsymbol{x}_{it},$$

$$E[\boldsymbol{\beta}_{it}] = \boldsymbol{T}_{it}\boldsymbol{\beta}_{i,t-1},$$

with

$$\boldsymbol{x}_{it} = (1,0;1,0;1,0;1,0;1,0;1)$$

$$\boldsymbol{T}_{it} = \text{blockdiag}(G_1,\ldots,G_5,\ -1), \qquad G_r = \begin{pmatrix}\cos(\omega r) & \sin(\omega r)\\ -\sin(\omega r) & \cos(\omega r)\end{pmatrix}.$$

Most often, such dynamic models are used in forecasting. However, here, they are particularly important as a unifying framework for many of the models of repeated measurements, as well as a means of calculating the likelihood function and estimating the parameters in difficult cases. Thus, two advantages of this approach, even in the case of a model based on the normal distribution, are that it can be extended to handle unequally spaced time intervals and missing observations and that it encompasses many useful models as special cases. However, little work has been done on unequally spaced time intervals and missing observations for such models for non-normal data.

The applications to discrete and positive-valued data in the subsequent sections will assume that the 'base' mean λ_{it} is random, but that the regression coefficients $\boldsymbol{\beta}_{it}$ are non-random and time-varying.

2. Discrete data

We now apply the above models to count data. Consider the Poisson distribution for modelling Y_{it}. Quite naturally we adopt a log-link together with gamma conjugate distribution for μ_{it}, the mean of Y_{it}. Hence

$$\Pr(y_{it} \mid \lambda_{it}, \boldsymbol{\beta}_{it}) = \frac{e^{-\lambda_{it}} e^{\boldsymbol{\beta}_{it}^{\mathrm{T}} \boldsymbol{x}_{it}} \lambda_{it}^{y_{it}} e^{\boldsymbol{\beta}_{it}^{\mathrm{T}} \boldsymbol{x}_{it} y_{it}}}{y_{it}!},$$

$$p(\lambda_{it}) = \frac{\upsilon_{it}^{-\kappa_{it}} \lambda_{it}^{\kappa_{it}-1} e^{-\lambda_{it}/\upsilon_{it}}}{\Gamma(\kappa_{it})}.$$

Note that the marginal distribution of Y_{it} is the negative binomial, commonly used when dealing with overdispersed count data.

Starting with some initial values, we have the prediction equations

$$\kappa_{i,t|t-1} = \zeta_i \kappa_{i,t-1},$$

$$\frac{1}{\upsilon_{i,t|t-1}} = \frac{\zeta_i}{\upsilon_{i,t-1}},$$

when ζ_i is a *discount factor* between zero and one and the updating equations

$$\kappa_{it} = \kappa_{i,t|t-1} + y_{it},$$

$$\frac{1}{\upsilon_{it}} = \frac{1}{\upsilon_{i,t|t-1}} + e^{\boldsymbol{\beta}_{it}^{\mathrm{T}} \boldsymbol{X}_{it}}.$$

Note that the choice for ζ_i does not affect the predicted mean for μ_{it}. However the predicted variance is multiplied by ζ_i^{-1}. For example a small value for ζ_i enables the model to adapt itself quickly to changes of behaviour in the profiles, but makes it more sensible to outliers. Hence some kind of trade-off is required when choosing ζ_i. Of course one might simply choose the MLE. Estimates for the regression parameters $\boldsymbol{\beta}_{it}$ and possibly ζ_i are obtained by minimizing the deviance $-2 \sum_t \{\log \Pr(y_{it} \mid \mathcal{F}_{i,t-1})\}$ given by

$$-2 \sum_t \left\{ \log[\Gamma(y_{it} + \kappa_{i,t|t-1})] - \log[\Gamma(\kappa_{i,t|t-1})] - \kappa_{i,t|t-1} \log(\upsilon_{i,t|t-1}) \right.$$
$$\left. - (\kappa_{i,t|t-1} + y_{it}) \log\left(\frac{1}{\upsilon_{i,t|t-1}} + 1 \right) \right\}.$$

Note that the corresponding expression for the deviance in Lindsey (1993, Eq. (6.13), p. 207) is wrong and has to be corrected. Typically ζ_i and λ_{it} are, respectively, used for modelling the variance and the mean of Y_{it}. Higher-order moments are not directly used.

Consider, now, the reported numbers of UK deaths from bronchitis, emphysema, and asthma (Diggle, 1990, p. 238) each month from 1974 to 1979, distinguished by sex, presented in Table 1. We shall fit models with a linear time trend to see if the number of deaths is changing over the years. As well, we require a seasonal component, since

Table 1
Monthly numbers of deaths from bronchitis, emphysema and asthma in the UK, 1974–1979 (Diggle, 1990, p. 238)

Males									
2134	1863	1877	1877	1492	1249	1280	1131	1209	1492
1621	1846	2103	2137	2153	1833	1403	1288	1186	1133
1053	1347	1545	2066	2020	2750	2283	1479	1189	1160
1113	970	999	1208	1467	2059	2240	1634	1722	1801
1246	1162	1087	1013	959	1179	1229	1655	2019	2284
1942	1423	1340	1187	1098	1004	970	1140	1110	1812
2263	1820	1846	1531	1215	1075	1056	975	940	1081
1294	1341								
Females									
901	689	827	677	522	406	441	393	387	582
578	666	830	752	785	664	467	438	421	412
343	440	531	771	767	1141	896	532	447	420
376	330	357	445	546	764	862	660	663	643
502	392	411	348	387	385	411	638	796	853
737	546	530	446	431	362	387	430	425	679
821	785	727	612	478	429	405	379	393	411
487	574								

Table 2
Deviances for various models for monthly deaths of Table 1

Effect	Separate sexes		Different level		Sexes together	
	Deviance	Par.	Deviance	Par.	Deviance	Par.
Gamma–Poisson						
Trend	7510.16	6	7510.18	5	29237.84	3
Seasonal	645.91	26	672.97	14	22400.61	13
Both	639.72	28	671.86	15	22399.55	14
Beta–negative Binomial						
Trend	280.30	8	281.36	6	479.81	4
Seasonal	0.27	28	4.62	16	447.72	14
Both	0.00	30	4.62	17	447.72	15

the number of deaths varies regularly over the year. We use seasonal harmonics for the twelve-month period. Thus, three models will be fitted: (1) separately to the data for each sex, (2) with the same trend and seasonal, but a different level for each sex, and (3) with all components the same for each sex. The resulting deviances for the gamma–Poisson (negative binomial) and beta–negative binomial (hypergeometric) are displayed in Table 2. We arbitrarily take a fixed discount of 0.7, although this could also have been estimated.

Although there is no clear saturated model, we take our most complex model as a baseline, arbitrarily giving it zero deviance, so that it can easily be compared with the others. We immediately see that all of the gamma–Poisson models are

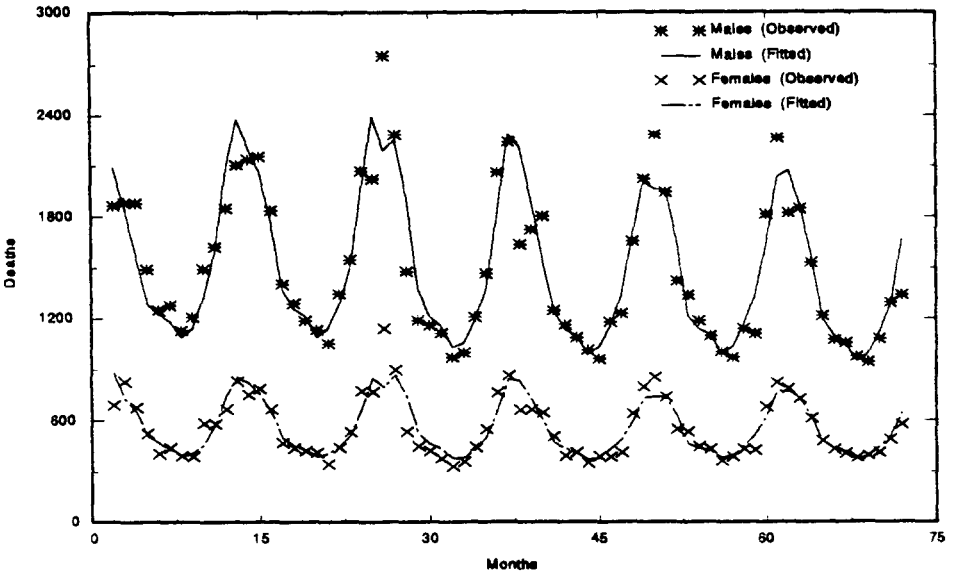

Fig. 1. Monthly deaths from bronchitis, emphysema and asthma, from Table 1, with fitted negative binomial DGLM.

unacceptable as compared to the beta–negative binomial ones. A trend is not necessary and the seasonal components can be the same for deaths of both sexes. However, the level must be different for the two sexes. This model has a deviance of only 4.62 greater than the most complex one, but with 14 fewer parameters. The only further simplification is to reduce the number of harmonics. It is only possible to eliminate the two highest ones, with a further increase in deviance of 3.76. The fitted values of the final model are plotted in Fig. 1, along with the observed numbers of deaths. The fitted line follows the observed deaths fairly closely, with the exception of three high numbers of the male deaths and one of females. Male deaths are consistently higher than female, but with the same seasonal variation. There is no indication of a change in the number of deaths over the years.

3. Positive-valued data

Dynamic generalized linear models can also be applied to duration data, or at least to longitudinal data having positive response values which might follow a gamma, inverse Gaussian, or log normal distribution, the most common appropriate members of the exponential family. Here, we consider the gamma distribution, whose conjugate is also a form of gamma, allowing for frailty or heterogeneity across units. The procedure for estimating the parameters is essentially the same as that described above, except for the change in distributions, and need not be repeated here.

We shall apply this model to measurements of plasma citrate concentration (μmol/l) for ten subjects over fourteen successive hourly observation points between eight in

Table 3
Plasma citrate concentrations (μmol/l) for 10 subjects at 14 successive times during the day (Andersen et al., 1981)

93	109	114	121	101	109	112	107	97	117
89	132	121	124						
116	116	111	135	107	115	114	106	92	98
116	105	135	83						
125	166	180	137	142	114	119	121	95	105
152	154	102	110						
144	157	161	173	158	138	148	147	133	124
122	133	122	130						
105	134	128	119	136	126	125	125	103	91
98	112	133	124						
109	121	100	83	87	110	109	100	93	80
98	100	104	97						
89	109	107	95	101	96	88	83	85	91
95	109	116	86						
116	138	138	128	102	116	122	100	123	107
117	120	119	99						
151	165	156	149	136	142	121	128	130	126
154	148	138	127						
137	155	145	139	150	141	125	109	118	109
112	102	107	107						

Table 4
Deviances for several dynamic generalized linear models for the plasma citrate data of Table 3

	Together		Different intercept	
	Deviance	Par.	Intercept	Par.
Null	1152.92	1	1079.68	10
Trend	1149.57	2	1075.17	11
Harmonics	1138.05	12	1053.87	21
Both	1134.96	13	1049.96	22

the morning and nine in the evening. The data, from Andersen et al. (1981), are reproduced in Table 3. Since interest centres on daily rhythms, this dynamic generalized linear model with harmonics may be appropriate.

With short series, as in this example, fitting a different DGLM to each series is not reasonable; there would be too many parameters. Instead we fit 'parallel' and identical series, with 12 harmonics for a half-day and a trend which might pick up a longer period. The resulting deviances are given in Table 4. The ten series are not identical, but have different levels, as already could be seen from Table 3. For example, subjects

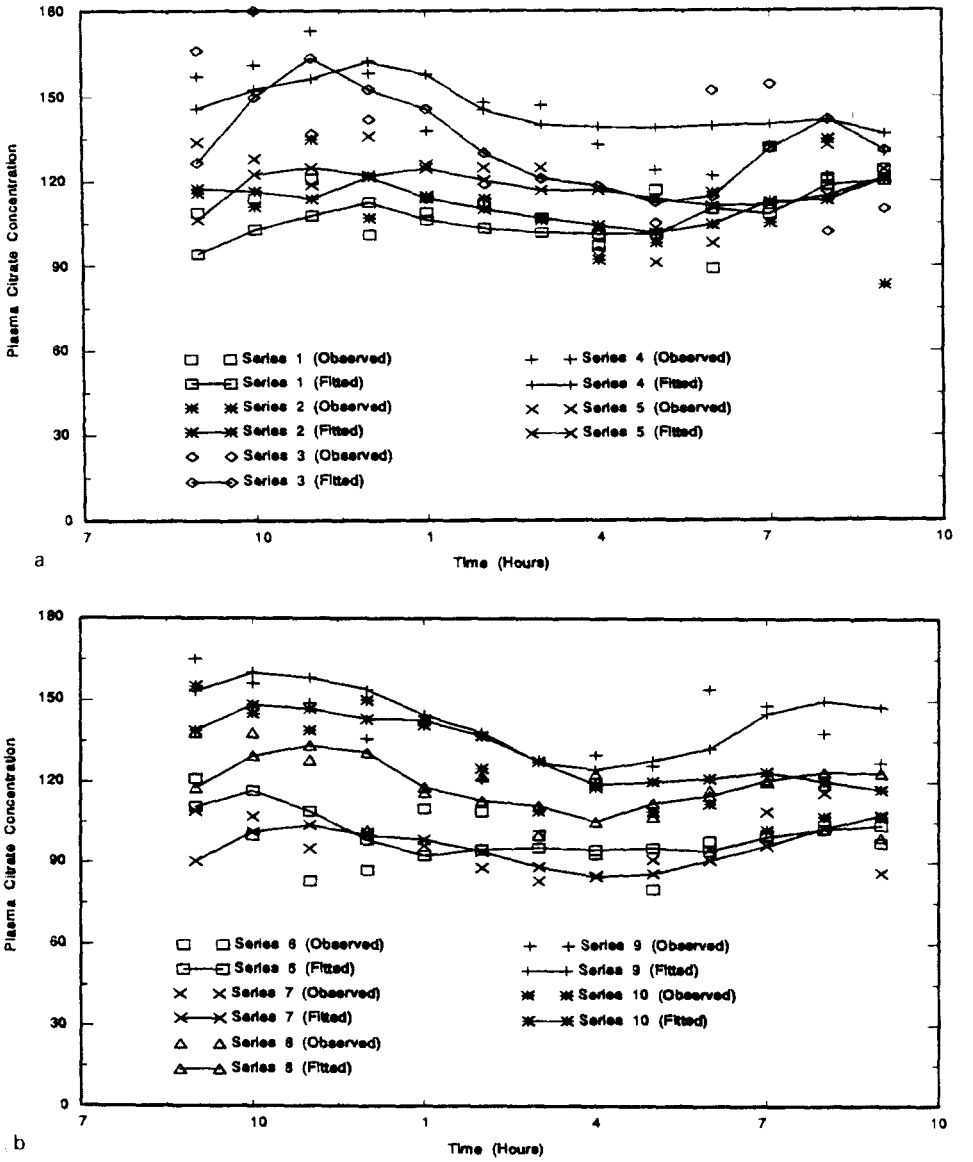

Fig. 2. Hourly plasma citrate concentrations, from Table 3, with fitted gamma DGLM.

one and seven have consistently lower plasma citrate concentrations than the others. Since a trend does not appear, the model with only the harmonics is sufficient. In fact, this can be reduced from 12 to four harmonics with only a gain in deviance of 8.45. The observations are plotted, along with the fitted model, in Fig. 2, arbitrarily separated into two plots to make the presentation clearer. We see that the plasma citrate concentration is generally highest at about ten in the morning and lowest

about four or five in the afternoon. There seems to be no relationship to the meal times of eight in the morning, noon, and five in the afternoon.

4. Discussion

Procedures for dynamic generalized linear models are not yet well developed, other than in the normal case. There they are very useful for fitting random effects and autoregression when the observation times are unequally spaced (Jones and Ackerson, 1990; Jones and Boadi-Boateng, 1991). For this reason, the method is coming to be extensively used for repeated measurements data. In the more general, non-normal, case, the intractability of the integrals means that time-consuming numerical methods must be used (Kitagawa, 1987) or approximation made. Here, only the first two moments of the conjugate distribution were employed. Little is known about how to apply such methods to unequally spaced data. However, the power of the procedure makes it one of the most promising avenues of research in repeated measures.

Acknowledgements

The authors would like to thank Christiano Fernandes for this FORTRAN program for a single DGLM series, which they have modified to perform the analysis of sets of series as in the examples.

References

[1] Andersen, A.H., E.B. Jensen and G. Schou (1981). Two-way analysis of variance with correlated errors. *Internat. Statist. Rev.* **49**, 153–167.
[2] Diggle, P.J. (1990). *Time Series. A Biostatistical Introduction.* Oxford University Press, Oxford.
[3] Fahrmeir, L. (1989). Extended Kalman filtering for nonnormal longitudinal data. In: A. Decarli, B.J. Francis, R. Gilchrist and G.U.H. Seeber Eds, *Statistical Modelling.* Springer, Berlin, 151–156.
[4] Fahrmeir, L. (1992). Posterior mode estimation by extended Kalman filtering for multivariate dynamic generalized linear models. *J. Amer. Statist. Assoc.* **87**, 501–509.
[5] Harvey, A.C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge University Press, Cambridge.
[6] Harvey, A.C. and C. Fernandes (1989). Time series models for count or qualitative observations. *J. Bus. Econom. Statist.* **7**, 407–423.
[7] Jones, R.H. and L.M. Ackerson (1990). Serial correlation in unequally spaced longitudinal data. *Biometrika* **77**, 721–731.
[8] Jones, R.H. and F. Boadi-Boateng (1991). Unequally spaced longitudinal data with AR(1) serial correlation. *Biometrics* **47**, 161–175.
[9] Kitagawa, G. (1987). Non-Gaussian state-space modeling of nonstationary time series. *J. Amer. Statist. Assoc.* **82**, 1032–1063.
[10] Lindsey, J.K. (1993). *Models for Repeated Measurements.* Oxford Statistical Science Series, Oxford.

[11] Schlain, B.R., P.T. Lavin and C.L. Hayden (1992). Using an autoregressive model to detect departures from steady states in unequally spaced tumour biomarker data. *Statistics in Medicine* **11**, 515–532.

[12] Singh, A.C. and G.R. Roberts (1992). State space modelling of cross-classified time series of counts. *Internat. Statist. Rev.* **60**, 321–335.

[13] West, M. and P.J. Harrison (1989). *Bayesian Forecasting and Dynamic Models*. Springer, Berlin.

[14] West, M., P.J. Harrison and H.S. Migon (1985). Dynamic generalized linear models and Bayesian forecasting. *J. Amer. Statist. Assoc.* **80**, 73–97.

[15] Wilson, P.D. (1988). Autoregressive growth curves and Kalman filtering. *Statistics in Medicine* **7**, 73–86.