Biometrika (2003), **90,** 2, *pp.* 355–366

Printed in Great Britain

A serially correlated gamma frailty model for longitudinal count data

BY ROBIN HENDERSON

Department of Mathematics and Statistics, Lancaster University, Lancaster LA1 4YF, U.K.

robin.henderson@lancaster.ac.uk

AND SILVIA SHIMAKURA

Departamento de Estatística, Universidade Federal do Paraná, Cx.Postal 19081, 81531-990, Curitiba-PR, Brazil

silvia.shimakura@est.ufpr.br

SUMMARY

A Poisson-gamma model is introduced to account for between-subjects heterogeneity and within-subjects serial correlation occurring in longitudinal count data. The model extends the usual time-constant shared frailty approach to allow time-varying serially correlated gamma frailty whilst retaining standard marginal assumptions. A composite likelihood approach to estimation and testing for serial correlation is proposed. The work is motivated by a clinical trial on patient-controlled analgesia where the number of analgesic doses taken by hospital patients in successive time intervals following abdominal surgery is recorded.

Some key words: Composite likelihood; Multivariate gamma distribution; Patient-controlled analgesia; Recurrent event data.

1. INTRODUCTION

We consider longitudinal count data, with N_{ij} events observed in interval j ($j = 1, \ldots, p$) for subject i ($i = 1, \ldots, s$). An example to be considered in § 5 concerns the number of analgesic doses taken by hospital patients in 12 successive four-hour intervals following abdominal surgery. Other examples include monthly cases of rare diseases (Zeger, 1988; Hay & Pettitt, 2001), epileptic seizures (Thall & Vail, 1990) and daily mortality numbers (Kelsall et al., 1999).

Parameter-driven models (Hay & Pettitt, 2001) for data of this type are usually based on a foundational assumption of Poisson regression with the possibility of a latent process to generate overdispersion and perhaps serial correlation. Within this general strategy three broad classes of models can be distinguished. The first class is based on the standard overdispersion approach for univariate count data, namely Poisson counts mixed by gamma random effects to give negative binomial marginals. In extending this to longitudinal data a common assumption is that event counts are conditionally independent Poisson variables given the value Z of a gamma-distributed subject-specific frailty term which remains constant in time; see Thall (1988), Diggle et al. (1994, Ch. 9) and references cited

by van Duijn & Böckenholt (1995). This approach allows for both extra-Poisson heterogeneity between individuals and within-subject association between intervals. A disadvantage is that with time-constant Z the association in event counts cannot depend on time-separation between the intervals. The second class induces both overdispersion and serial correlation through a multiplicative latent process, $Z_j = \exp(W_j)$ for period j, where $W_1, W_2, \ldots$ form a Gaussian time series (Chan & Ledolter, 1995; Hay & Pettitt, 2001). Efficient estimation is possible but requires computationally intensive methods and closed-form expressions do not exist for the marginal distributions or moments. The third approach is based on generalised estimating equations, specifying first- and second-order moments and estimating from the appropriate sample statistics (Zeger, 1988; Kelsall et al., 1999; Jowaheer & Sutradhar, 2002). This is robust but can be inefficient.

In this work we aim to combine the advantages of the first and second modelling approaches by specifying a time-varying frailty model which satisfies the following criteria: gamma frailty Z_{ij} for interval j and subject i, with mean one and variance ξ, written $\mathrm{Ga}(1/\xi, 1/\xi)$; conditionally independent Poisson N_{ij} given Z_{ij}; and within-subject correlation $\mathrm{corr}(Z_{ij}, Z_{ik}) = \rho^{|j-k|}$ $(0 \leqslant \rho \leqslant 1)$.

The first two of these criteria maintain the proven advantages of Poisson-gamma frailty models, principally flexibility and the availability of a closed, negative binomial, form for the marginal distribution of the event counts. The third subsumes the standard shared frailty approach into a potentially more realistic correlation structure. Additional advantages of our approach are improved efficiency in comparison with estimating equation methods, availability of closed-form expressions for covariances and other moments, and estimation using a standard Newton–Raphson technique rather than the EM algorithm, Markov chain Monte Carlo or other computationally intensive methods.

The proposed correlated gamma frailty model is described in § 2. A composite likelihood estimation procedure based on pairwise contributions is developed in §§ 3 and 4. The application to patient-controlled analgesia data is described in § 5, and some brief remarks in § 6 complete the paper.

2. CORRELATED GAMMA FRAILTY MODEL

We suppose that each subject is observed over a common observation period which is partitioned into p intervals of equal length. This assumption simplifies the presentation but is not necessary. For the moment we consider just a single subject and drop the subscript i. Thus $N_1, \ldots, N_p$ are the event counts and $Z_1, \ldots, Z_p$ are the corresponding frailties. Fixed covariates are denoted by x and are assumed to be time-constant, though again this is not a necessary assumption. We assume that the event counts are conditionally independent Poisson variables with

$$N_j | Z_j \sim \mathrm{Po}\{Z_j \exp(\alpha_j + \beta x)\},$$

where α_j determines the interval-specific baseline rate.

To derive a multivariate gamma distribution with the required properties we begin by recalling a standard result. Suppose $Y_1, \ldots, Y_q$ are independent p-variate Gaussian with standard marginals and common $p \times p$ correlation matrix C. Write $Y_j = (Y_{j1}, \ldots, Y_{jp})'$ and let $Z_k = \sum_{j=1}^{q} Y_{jk}^2/q$, for $k = 1, \ldots, p$. Then the standard result is that $Z = (Z_1, \ldots, Z_p)'$ is multivariate gamma with marginal $\mathrm{Ga}(q/2, q/2)$ distributions and Laplace transform

$$E\{\exp(-u'Z)\} = |I + 2C\,\mathrm{diag}(u)/q|^{-q/2}, \tag{1}$$

where $u = (u_1, \ldots, u_p)'$ (Krishnamoorthy & Parthasarathy, 1951). Moreover, by differentiation it is straightforward to show that the correlation matrix for Z is a matrix R of elementwise squares of C, that is $R_{jk} = C_{jk}^2$.

As given, the distribution requires positive integer q and hence is of limited use. The question as to whether or not (1) defines a proper distribution for non-integer q has had some attention (Krishnaiah & Rao, 1961; Moran & Vere-Jones, 1969; Griffiths, 1970) with Griffiths (1984) showing that under certain conditions (1) does indeed define a valid distribution for any $q > 0$. Equivalent conditions that are more simply evaluated are given by Bapat (1989): provided the off-diagonal elements of $(MCM)^{-1}$ are nonpositive for some diagonal matrix M with diagonal elements 1 or -1 such that all entries in MCM are nonnegative, the form (1) defines a Laplace transform of an infinitely divisible distribution.

Now suppose that C has elements $C_{jk} = \rho^{|j-k|/2}$ for $0 \leqslant \rho \leqslant 1$. For $M = I$, $(MCM)^{-1}$ has nonpositive off-diagonal elements, the Bapat condition is satisfied and (1) defines a proper multivariate gamma distribution. Letting $\xi = 2/q$ we thus have the following definition of a multivariate gamma distribution with the required marginal and correlation properties.

DEFINITION. *With* $C_{jk} = \rho^{|j-k|/2} = R_{jk}^{1/2}$ $(0 \leqslant \rho \leqslant 1)$, *the p-vector* Z *with Laplace transform*

$$\mathscr{L}(u) = E\{\exp(-u'Z)\} = |I + \xi C \operatorname{diag}(u)|^{-1/\xi} \tag{2}$$

has a proper multivariate gamma distribution, with $\mathrm{Ga}(1/\xi, 1/\xi)$ *marginal distributions and correlation matrix* R, *for all* $\xi > 0$.

This definition applies also for other correlation structures, including compound symmetry $C_{jk} = \rho^{1/2}$ and any Markovian-type correlation matrix such that $C_{ik} = C_{ij}C_{jk}$ $(i \leqslant j \leqslant k)$.

Although there is no known closed form for the density of Z, inference will be based on the marginal distribution of the count data after integrating out the frailty terms, and in effect on the properties of the Laplace transform of Z, which we have seen to have a very simple form. For instance, if we write $u_j = \exp(\alpha_j + \beta x)$ and if Z has the multivariate gamma distribution defined at (2) then the count data $N_1, \ldots, N_p$ have the following moments: $E(N_j) = u_j$, $\operatorname{var}(N_j) = \xi u_j^2 + u_j$ and $\operatorname{cov}(N_j, N_k) = \rho^{|j-k|}\xi u_j u_k$.

The marginal distributions are negative binomial, i.e.

$$\operatorname{pr}(N_j = n_j) = \frac{u_j^{n_j}}{n_j!(1 + \xi u_j)^{n_j + 1/\xi}} \prod_{k=0}^{n_j - 1} (1 + k\xi),$$

and the joint distribution can be obtained in principle by differentiation of the Laplace transform at (2):

$$\begin{aligned} \operatorname{pr}(N_1 = n_1, \ldots, N_p = n_p) &= \left(\prod_{j=1}^{p} \frac{u_j^{n_j}}{n_j!} \right) \times E\{Z_1^{n_1} \ldots Z_p^{n_p} \exp(-u'Z)\} \\ &= (-1)^{\sum n_j} \left(\prod_{j=1}^{p} \frac{u_j^{n_j}}{n_j!} \right) \frac{\partial^{(n_1 + \ldots + n_p)} \mathscr{L}(u)}{\partial^{n_1} u_1 \ldots \partial^{n_p} u_p}. \end{aligned} \tag{3}$$

In practice however the number of terms involved in the derivatives quickly becomes unmanageable for realistic problems.

3. BIVARIATE CASE

We consider the special case of two intervals as a prelude to the composite likelihood technique to be described in § 4. We consider intervals 1 and 2, with counts N_1 and N_2, frailties Z_1 and Z_2 and frailty correlation $\mathrm{corr}(Z_1, Z_2) = \rho$. Results apply equally to any intervals j and k provided ρ is replaced with $\rho^{|j-k|}$.

We write the bivariate Laplace transform (2) as

$$\mathscr{L}(u) = \begin{vmatrix} 1+\xi u_1 & \rho^{1/2}\xi u_2 \\ \rho^{1/2}\xi u_1 & 1+\xi u_2 \end{vmatrix}^{-1/\xi} = \Delta^{-1/\xi}$$

say, where $\Delta = 1 + \xi u_1 + \xi u_2 + \xi^2 u_1 u_2 (1-\rho)$. If we use (3) and induction, the joint distribution of the event counts N_1 and N_2 can be shown to be

$$\mathrm{pr}(N_1 = n_1, N_2 = n_2) = \frac{u_1^{n_1} u_2^{n_2}}{n_1! n_2!} \sum_{j=0}^{\min(n_1,n_2)} \left[(-1)^j \binom{n_1}{j} \binom{n_2}{j} j! \left\{ \prod_{k=0}^{n_1+n_2-j-1} (1/\xi + k) \right\} \right.$$
$$\left. \times \Delta^{-(1/\xi + n_1 + n_2 - j)} \left(\frac{\partial \Delta}{\partial u_1} \right)^{n_1 - j} \left(\frac{\partial \Delta}{\partial u_2} \right)^{n_2 - j} \left(\frac{\partial^2 \Delta}{\partial u_1 \, \partial u_2} \right)^{j} \right] \quad (4)$$

with the convention that $\prod_a^b = 1$ if $b < a$. Writing

$$\frac{\partial \Delta}{\partial u_1} = \xi\{1 + \xi u_2 (1-\rho)\} = \xi D_1, \quad \frac{\partial \Delta}{\partial u_2} = \xi\{1 + \xi u_1 (1-\rho)\} = \xi D_2, \quad \frac{\partial^2 \Delta}{\partial u_1 \, \partial u_2} = \xi^2 (1-\rho),$$

setting $m_1 = \min(n_1, n_2)$, $m_2 = \max(n_1, n_2)$ and defining $f = \Delta(1-\rho)/D_1 D_2$, we can rearrange (4) as

$$\mathrm{pr}(N_1 = n_1, N_2 = n_2) = \frac{u_1^{n_1} u_2^{n_2}}{n_1! n_2!} \left\{ \prod_{k=0}^{m_2 - 1} (1 + k\xi) \right\} \Delta^{-1/\xi} \left(\frac{D_1}{\Delta} \right)^{n_1} \left(\frac{D_2}{\Delta} \right)^{n_2}$$
$$\times \sum_{j=0}^{m_1} \left[(-1)^j \binom{m_1}{j} \binom{m_2}{j} j! \left\{ \prod_{k=m_2}^{m_1 + m_2 - j - 1} (1 + k\xi) \right\} \xi^j f^j \right]. \quad (5)$$

At $\rho = 1$ the model is equivalent to the shared frailty model and (5) reduces to

$$\mathrm{pr}(N_1 = n_1, N_2 = n_2) = \frac{u_1^{n_1} u_2^{n_2}}{n_1! n_2!} \left\{ \prod_{k=0}^{n_1 + n_2 - 1} (1 + k\xi) \right\} \{1 + \xi(u_1 + u_2)\}^{-(\xi^{-1} + n_1 + n_2)}.$$

At $\rho = 0$ the frailty terms are independent between intervals and the probability becomes

$$\mathrm{pr}(N_1 = n_1, N_2 = n_2) = \frac{u_1^{n_1} u_2^{n_2}}{n_1! n_2!} \left\{ \prod_{k=0}^{n_1 - 1} (1 + k\xi) \right\} \left\{ \prod_{k=0}^{n_2 - 1} (1 + k\xi) \right\}$$
$$\times (1 + \xi u_1)^{-(\xi^{-1} + n_1)} (1 + \xi u_2)^{-(\xi^{-1} + n_2)}. \quad (6)$$

4. COMPOSITE LIKELIHOOD APPROACH

4·1. *Parameter estimation*

Now reintroduce the subscript i ($i = 1, \ldots, s$) to index subjects and let $\theta = (\{\alpha_j\}, \beta, \xi, \rho)'$ be the combined vector of unknown parameters. Full likelihood analysis requires the joint probabilities $\mathrm{pr}(N_{i1} = n_{i1}, \ldots, N_{ip} = n_{ip})$, which we have seen to be intractable. As an alternative we propose a consistent composite likelihood procedure (Lindsay, 1988) based

on all pairwise comparisons of intervals. Let

$$l_i(\theta) = \sum_{j=1}^{p-1} \sum_{k=j+1}^{p} \frac{1}{p-1} \log \text{pr}(N_{ij} = n_{ij}, N_{ik} = n_{ik}) = \sum_{j=1}^{p-1} \sum_{k=j+1}^{p} \frac{l_{ijk}(\theta)}{p-1}$$

be the composite loglikelihood contribution for subject i, obtained using the results of § 3. Here the divisor $(p-1)$ is included so that in the independence case, $\rho = 0$, the composite likelihood is equivalent to the standard full likelihood. The combined composite loglikelihood is then

$$l(\theta) = \sum_{i=1}^{s} l_i(\theta).$$

As each term l_{ijk} is a proper loglikelihood based on counts in intervals j and k, the composite score $\partial l/\partial\theta$ provides an unbiased estimating equation and standard asymptotic results apply for the maximum composite likelihood estimator $\hat{\theta}$. In particular, as $s \to \infty$, $\sqrt{s}(\hat{\theta} - \theta_0)$ converges in distribution under regularity conditions (Godambe & Heyde, 1987) to $N(0, V)$, where θ_0 is the true parameter vector, $V = V_1^{-1} V_0 V_1^{-1}$ with

$$V_1 = \lim_{s\to\infty} \frac{1}{s} \sum_i \frac{\partial^2 l_i}{\partial\theta\, \partial\theta'}, \quad V_0 = \lim_{s\to\infty} \frac{1}{s} \sum_s E\left\{\left(\frac{\partial l_i}{\partial\theta}\right)\left(\frac{\partial l_i}{\partial\theta}\right)'\right\}$$

and the derivatives are evaluated at θ_0. The components V_0 and V_1, can be estimated from the composite score and information in the usual way to obtain a robust variance estimator (Zeger & Brookmeyer, 1986; Liang, 1987). Expressions for the required differentials are quite involved and are hence omitted, though further details are available from the authors.

A final point about estimation is that care should be taken to avoid rounding error when counts are high and some of the factorial terms in the composite likelihood are large. Some robustness can be gained by writing (5) as the product of two terms, one of which reduces to the likelihood under independence when $\rho = 0$. Again detail is available on request. If all counts are high then a Normal approximation is likely to be adequate.

4·2. *Numerical results*

In this section we summarise simulation work designed to investigate the performance of the composite likelihood estimator $\hat{\theta}$ and its associated robust variance estimator. Additionally we compare with simpler methods using either an independence working assumption or a generalised estimating equation approach and we give some efficiency results.

A selection of results are shown in Table 1. For this study we simulated 500 datasets each of sample size 250, at each combination of $\xi = (0{\cdot}25, 0{\cdot}50)$ and $\rho = (0{\cdot}5, 0{\cdot}9)$. There were $p = 12$ counts per subject, with baseline $\exp(\alpha_j) = 4$ for each period, one binary covariate and $\beta = 0$. The table shows mean composite likelihood parameter estimates together with standard errors from mean robust variances and empirically observed over the 500 simulations. In addition there are efficiency estimates for independence working assumption and generalised estimating equation alternative estimation techniques. The independence working assumption method assumes the correct negative binomial marginal distributions but ignores correlation. The generalised estimating equation results are based on the moments of the N_{ij} as in Zeger (1988), and with the exception of ρ were obtained using the eponymous R module. This routine estimates extra-Poisson variation through a scale factor, $\text{var}(N_{ij}) = \phi E(N_{ij})$, whereas for the gamma frailty model $\text{var}(N_{ij}) =$

$\xi E(N_{ij})^2 + E(N_{ij})$, and hence normally there is no direct comparison with ξ. However, when all means are equal, $E(N_{ij}) = u$ say, we can set $\hat{\xi} = \hat{\phi} - 1/u$ and this explains our choice of $\alpha_1 = \ldots = \alpha_{12}$ and $\beta = 0$. We also obtained a direct estimate of ξ by equating the sample mean of N_{ij}^2 to its expectation, essentially following the procedure of Jowaheer & Sutradhar (2002), and obtained very similar results to those reported in Table 1. Whilst a working first-order autoregressive correlation structure was assumed under the generalised estimating equation method, that is $\mathrm{corr}(N_{ij}, N_{ik}) = v^{|j-k|}$, the parameter v is not comparable to the model correlation parameter ρ given that the true covariance is $\mathrm{cov}(N_{ij}, N_{ik}) = \rho^{|j-k|}\xi u_{ij} u_{ik}$. Instead, for illustration we estimated ρ using a moment estimator based on the sample lagged products. For this, first we estimated separately at each lag by equating $\sum_i N_{ij} N_{ik}$ to its appropriate expectation from the covariance above, and then adjusting by using the true values of u_{ij}, u_{ik} and ξ to obtain a sample estimate of $\rho^{|j-k|}$. Next we used weighted least squares on the combined log-estimators to derive an overall estimator of $\log \rho$ and hence ρ, weighting in inverse proportion to the number of available observations at each lag. We also estimated ρ more directly using the lag-one results only, as suggested by Zeger (1988). These results are not reported as efficiency was worse than for the first estimator, which made more use of the available data. The table also omits results for baseline parameters other than α_1 and α_{12} for space reasons.

Table 1. *Specimen simulation results, from* 500 *samples of size* $s = 250$. *Efficiencies under independence working assumption* (IWA) *and generalised estimating equation* (GEE) *approaches are estimated from empirical variances over the simulations. True* $\alpha_1 =$ *true* $\alpha_{12} = \log 4 = 1{\cdot}386$ *and true* $\beta = 0$

				Standard error		Efficiency (%)	
ξ	ρ		Mean	Robust	Empirical	IWA	GEE
0·5	0·5	$\hat{\alpha}_1$	1·380	0·059	0·061	100	100
		$\hat{\alpha}_{12}$	1·387	0·059	0·058	99	99
		$\hat{\beta}$	0·002	0·045	0·045	98	101
		$\hat{\xi}$	0·497	0·023	0·022	100	60
		$\hat{\rho}$	0·496	0·028	0·030	—	5
0·25	0·5	$\hat{\alpha}_1$	1·386	0·048	0·047	100	100
		$\hat{\alpha}_{12}$	1·386	0·048	0·047	100	99
		$\hat{\beta}$	0·000	0·035	0·034	100	100
		$\hat{\xi}$	0·247	0·014	0·014	100	80
		$\hat{\rho}$	0·491	0·032	0·038	—	6
0·5	0·9	$\hat{\alpha}_1$	1·381	0·066	0·064	99	99
		$\hat{\alpha}_{12}$	1·379	0·066	0·065	99	100
		$\hat{\beta}$	0·004	0·074	0·070	97	99
		$\hat{\xi}$	0·495	0·034	0·033	95	52
		$\hat{\rho}$	0·899	0·011	0·012	—	2
0·25	0·9	$\hat{\alpha}_1$	1·384	0·053	0·054	100	100
		$\hat{\alpha}_{12}$	1·380	0·053	0·055	100	101
		$\hat{\beta}$	0·000	0·055	0·055	99	101
		$\hat{\xi}$	0·248	0·019	0·020	96	73
		$\hat{\rho}$	0·900	0·014	0·014	—	1

From these and other simulation results we conclude that the composite likelihood and robust variance estimators perform well. There is no noticeable bias and the estimated

variances match the observed values closely. The same is true also for all parameters estimated under the independence working assumption method, and for the mean parameters when using generalised estimating equations, results not shown. The estimating equation method has poor efficiency however in estimation of the variance parameter ξ, with the performance deteriorating as ξ increases, and extremely poor efficiency in estimation of the correlation parameter, despite the method exploiting known u_{ij}, u_{ik} and ξ in this simulation study. To illustrate this further, Table 2 shows the asymptotic relative efficiency of generalised estimating equation estimators of ξ and ρ in comparison to model-based likelihood estimation in the bivariate case $p = 2$, with as above baseline count $\exp(\alpha_j) = 4$ and no covariate effect. These values were obtained using the expected information based on (5) and the moments of N_{ij}^2 and $N_{i1}N_{i2}$, which can be calculated under the assumed model. The moment-based estimator of ρ has very low efficiency and we note that at this smaller number of counts for each subject the generalised estimating equation estimator of ξ also performs poorly.

Table 2. *Simulation results. Asymptotic relative efficiency of estimating equation method relative to likelihood estimation, with $p = 2$ observations per subject and all $E(N_{ij}) = 4$*

	Efficiency (%)					
	$\xi = 0{\cdot}25$		$\xi = 0{\cdot}50$		$\xi = 1{\cdot}00$	
	$\rho = 0{\cdot}5$	$\rho = 0{\cdot}9$	$\rho = 0{\cdot}5$	$\rho = 0{\cdot}9$	$\rho = 0{\cdot}5$	$\rho = 0{\cdot}9$
$\hat{\xi}$	11	10	11	11	10	10
$\hat{\rho}$	11	6	10	3	8	1

5. Application

5·1. *Background*

Patient-controlled analgesia is a mechanism which allows patients to control their own pain relief following surgery. On request, a machine infuses a bolus of drug provided that a sufficient period, the lock-out time, has elapsed since the previous delivery. Table 3 shows the average number of requests in 12 successive four-hourly intervals following abdominal surgery for 65 patients in a clinical trial to compare two bolus/lock-out combinations. In group 1, containing 30 patients, the bolus was 2 mg of morphine and the machine was locked out for 8 minutes after each delivery. In group 2, of 35 patients, the bolus was 1 mg of morphine with four-minute lock-out. The difference in the lock-out times restricts the number of requests per interval to at most 30 for patients in group 1 and 60 for those in group 2. Such a restriction can in principle make comparisons of raw counts misleading but need not be of concern here as the observed counts are all below the corresponding upper limits. There is substantial extra-Poisson variation, see Table 3, and evidence of serial correlation decaying with time; see Fig. 1. Thus a time-varying random effect model may be appropriate.

5·2. *Results*

Parameter estimates under three different frailty models and a generalised estimating equation approach are shown in Table 4, where β denotes the effect of the bolus/lock-out combination. We fitted the proposed time-varying serially correlated gamma frailty model

Table 3. *Summary statistics for the* 12 *four-hourly patient-controlled analgesia request counts*

		Interval											
Group		1	2	3	4	5	6	7	8	9	10	11	12
1	$\bar{N}$	9·3	5·5	5·4	5·1	7·6	5·3	3·9	3·7	4·6	4·9	3·5	3·4
	s^2	33·5	22·0	16·0	17·0	25·1	11·3	17·6	9·2	10·0	13·2	4·3	8·0
2	$\bar{N}$	10·2	6·5	7·9	9·3	9·6	7·4	6·6	5·7	4·9	6·3	6·3	5·9
	s^2	59·2	31·1	60·0	50·7	43·1	30·7	24·9	22·8	16·1	30·3	29·7	33·8

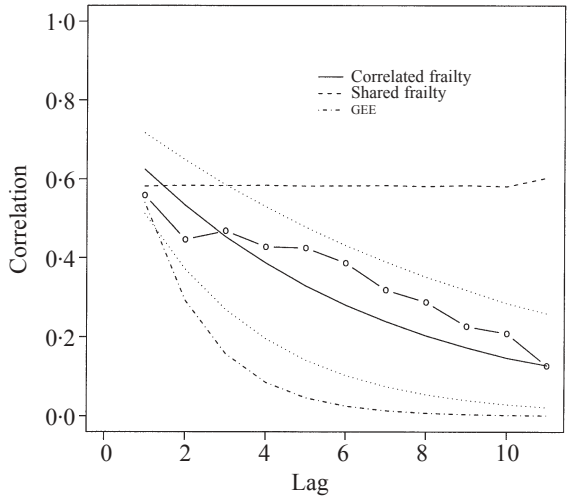

Fig. 1. Empirical (circles) and fitted correlation functions for the patient-controlled analgesia count data; GEE, generalised estimating equation approach; dotted lines, 95% simulation envelope.

using the composite loglikelihood of § 4, and for reference used standard likelihood methods to fit a shared frailty model and a model that assumed between-interval independence. The independent-frailty parameter estimates are identical to those obtained under a negative binomial independence working assumption and so we also give robust standard errors for this model, appropriate should independence be considered a working assumption only. Estimates of mean parameters are similar under all models, though consistently slightly higher when generalised estimating equations are used. Robust standard errors are also similar where available, with standard information-based estimates under independent and shared frailty estimates being slightly smaller. Note that the estimate of ξ is substantially lower under the shared frailty model than the others. This is because with shared frailty ξ influences both the between-patient heterogeneity and the within-patient association, whereas under the other models there is more separation of these components. The suggestion is that ξ is reduced under shared frailty as a consequence of less overall within-patient association than would be anticipated given the heterogeneity. There is no estimate of ξ when generalised estimating equations are used, as this method produces instead an estimate of scale ϕ as discussed earlier. A first-order autoregressive correlation structure, $\mathrm{corr}(N_{ij}, N_{ik}) = v^{|j-k|}$, was assumed for the estimating equation method, and v was estimated to be substantially smaller than the time-varying frailty parameter ρ.

Although not shown in Table 4, there are also differences between models in the esti-

Table 4. *Parameter estimates and standard errors, in parentheses, for patient-controlled analgesia data under three Poisson-gamma frailty models and a generalised estimating equation approach with autoregressive working correlation matrix, elements* $\mathrm{corr}(N_{ij}, N_{ik}) = v^{|j-k|}$. *For the independent frailty model two standard errors are given, information-based and robust*

	Frailty							
Parameter	Independent		Shared		Time-varying		GEE	
α_1	2·10	(0·10, 0·10)	2·09	(0·10)	2·10	(0·10)	2·11	(0·10)
α_2	1·61	(0·11, 0·13)	1·61	(0·10)	1·62	(0·13)	1·63	(0·13)
α_3	1·71	(0·10, 0·12)	1·72	(0·10)	1·71	(0·12)	1·74	(0·12)
α_4	1·78	(0·10, 0·11)	1·80	(0·10)	1·79	(0·11)	1·83	(0·11)
α_5	1·96	(0·10, 0·11)	1·97	(0·10)	1·97	(0·10)	1·99	(0·10)
α_6	1·66	(0·11, 0·10)	1·67	(0·10)	1·66	(0·10)	1·69	(0·10)
α_7	1·47	(0·11, 0·14)	1·49	(0·11)	1·47	(0·13)	1·51	(0·13)
α_8	1·37	(0·11, 0·12)	1·38	(0·11)	1·37	(0·12)	1·40	(0·12)
α_9	1·38	(0·11, 0·11)	1·37	(0·11)	1·37	(0·11)	1·39	(0·11)
α_{10}	1·55	(0·11, 0·11)	1·55	(0·11)	1·54	(0·11)	1·57	(0·11)
α_{11}	1·40	(0·11, 0·10)	1·42	(0·11)	1·39	(0·10)	1·44	(0·10)
α_{12}	1·35	(0·11, 0·12)	1·37	(0·11)	1·35	(0·12)	1·39	(0·12)
β	0·34	(0·06, 0·13)	0·33	(0·12)	0·34	(0·13)	0·30	(0·12)
ξ	0·49	(0·03, 0·05)	0·24	(0·04)	0·49	(0·06)	$\hat{\phi} = 3{\cdot}98$	—
ρ	0	—	1	—	0·85	(0·03)	$\hat{v} = 0{\cdot}54$	—
loglike	−2191·8		−2249·2		−2159·5			

GEE, generalised estimating equation approach.
loglike, loglikelihood for independent and shared frailty models, composite loglikelihood for time-varying frailty model.

mated correlations between parameter estimates. For example, under the independent frailty model the estimated baseline intensities are close to being orthogonal, with correlation between α_1 and α_2 for instance just 0·09. The robust estimate of the same quantity is 0·71, under shared frailty it is 0·80, under time-varying frailty it is 0·70 and the generalised estimating ·equation estimate is 0·70. Intuitively, under independence between periods there is no reason for separate baseline parameters to be strongly correlated as the only dependence arises through the shared covariate. When there is strong positive serial correlation on the other hand, there is potential for consistent over- or underestimation of baselines across periods which are close in time, and this is reflected in the correlation structure. For α_1 and α_{12} the corresponding estimates are 0·09, under the independent frailty model, 0·33, robust, 0·78, shared, 0·30, time-varying, and 0·26 for the generalised estimating equation approach. Note that with wider separation there is less correlation between baselines under the models which allow the correlation in counts to decay with time. Clearly these differences are important when considering the cumulative baseline for total dose purposes or in comparing baselines in investigating the changing pattern over time.

All models suggest the presence of unobserved heterogeneity, indicated by the frailty variance estimates and their standard errors; note that no standard error for ϕ or v is provided by the generalised estimating equation software. Moreover, the increase in the composite loglikelihood from the independent frailty to the time-varying frailty model, and the decrease from this to the shared frailty model, suggest that the unobserved

stochastic process which induces the extra-Poisson variation is unlikely to be either uncorrelated, $\rho = 0$, or perfectly correlated, $\rho = 1$, in time. This conclusion is supported by a Monte Carlo test of $H_0 : \rho = 1$ performed using 1000 simulations which resulted in a p-value of 0·001, and also by Fig. 1, which shows the empirical and the fitted autocorrelation curves, averaged over groups. The fitted time-varying frailty model certainly provides the best description for the correlation structure in the data. The empirical curve is not only close to that fitted, it is almost entirely contained within 95% bootstrap reference bands. The working correlation structure for the generalised estimating equation method estimates the average lag-1 correlation well but seriously underestimates at higher separation.

5·3. *Comments*

One of the original aims of the trial was to investigate whether patients in the two regimes tended to take either similar total doses of morphine or make similar numbers of requests for morphine. These hypotheses are equivalent to $\beta = \log 2$ and $\beta = 0$ respectively, both of which would be rejected by formal significance tests under all approaches considered. Patients in the 1 mg group made more requests on average than those in the 2 mg group, but not by the factor two needed to obtain the same total dose. A secondary aim of the trial was to describe both within- and between-patient variability with a long-term goal of setting patient-specific time-dependent maximum allowed doses, balancing the requirement to reduce the risk of overdose with the need for patient analgesia. In this context it is interesting to contrast predictions under each of the models given incomplete sequences of observations. To illustrate we consider the mean and standard deviation of the number of requests in intervals 3 and 12 given the observed numbers of requests in intervals 1 and 2. We omit details of the calculations, which for the time-varying frailty model are based on the pairwise distribution derived in § 3. Instead, in Table 5 we give numerical values for three different combinations of N_1 and N_2. Two examples are based on data from patients in the 2 mg group, using the 10% and 90% points of $N_1 + N_2$ as selection criteria, namely patients 4 and 13, with $(N_1, N_2) = (4, 1)$ and $(23, 6)$ respectively. We also consider prediction if the values for patient 13 had been reversed to $(6, 23)$. Under the marginal model there is no use of information in (N_1, N_2) and the three illustrations all have the same values. The difference between the interval 3 and interval 12 sections is simply due to an overall lower mean towards the end of the follow-up period. Both frailty models have smaller prediction standard deviations than the marginal model, this precision being bought of course by the additional assumptions on the correlation structure. The mean predictions under the frailty models are affected by (N_1, N_2) in the obvious way, and we note attenuation towards the marginal under the time-varying frailty model, as expected. The two-interval total $N_1 + N_2$ is sufficient for prediction under the shared frailty model and so we see the same results for the $(23, 6)$ and $(6, 23)$ illustrations, whereas under the time-varying model the most recent observation has more influence.

6. DISCUSSION

A Laplace transform similar to (1) though allowing nonstandard marginals for Y was used by Aalen (1987) to construct a nonnegative multivariate distribution for use in mixing intensities of Markov chains. Extension to non-integer q was not considered so that the marginals remained as χ^2, possibly noncentral, and inference was not considered. For our model, estimation methods based on the computation of the full likelihood seem

Table 5. *Patient-controlled analgesia data. Illustration of conditional mean and standard deviation,* SD, *of number of requests in intervals* 3 *and* 12 *given numbers in intervals* 1 *and* 2 *from* 2 *mg bolus patients, for three different models*

		$N_1 = 4$, $N_2 = 1$		$N_1 = 23$, $N_2 = 6$		$N_1 = 6$, $N_2 = 23$	
Interval	Model	Mean	SD	Mean	SD	Mean	SD
3	Marginal	5·52	4·52	5·52	4·52	5·52	4·52
	Shared	2·97	1·72	10·73	3·28	10·73	3·28
	Time-varying	2·86	1·68	8·77	2·91	12·93	3·55
12	Marginal	3·87	3·35	3·87	3·35	3·87	3·35
	Shared	2·10	1·45	7·60	2·76	7·60	2·76
	Time-varying	3·45	1·86	4·41	2·10	5·09	2·25

unmanageable but the alternative approach suggested in this work seems to work well. For simplicity, we have chosen to base our estimation on all pairwise combinations of time intervals for each individual. Thus, the probability elements used to construct our estimating function, the composite likelihood, involve only bivariate contributions.

Our model extends the available random effects methodology for recurrent count data. Other extensions to the standard shared-frailty approach for event time or count data include a replacement of the gamma assumption with power variance function frailties (Hougaard et al., 1997), a two-state switching model for intensities (Albert, 1991) and additive frailty models (Petersen, 1998). The last-mentioned can lead to a compound symmetry structure for the frailty correlation matrix, which can also be obtained directly by replacing $\rho^{|j-k|}$ in (2) with ρ, as the resulting distribution remains proper. Additive and direct distributions have the same marginal and association properties but not exactly the same distributions, though differences are small except in the tails, as shown in J. C. Chapman's unpublished 2000 Ph.D. Thesis at Lancaster University.

Of course should interest lie only in the regression or other marginal parameters then there may be no need to use our estimation procedure as a marginal generalised estimating equation or independence working assumption analysis would provide consistent estimators. This would lead to a simpler estimation procedure, though at the price of losing assessment of dependence, goodness of fit or the possibility of prediction (Hougaard, 2000, p. 420). Additionally, as pointed out by a referee, our method can accommodate missing data and unbalanced time intervals. Arguments for and against marginal methods in comparison with conditional or random effects methods are well known; see Diggle et al. (1994) and Hougaard (2000) for précis in longitudinal and event time data respectively.

For this work we have concentrated on count data. In principle the time-varying gamma frailty model may also be applicable for repeated event time data. Lawless et al. (2001) discuss analysis methods for data of this type and criticise the standard assumption of time-fixed random effects independent of covariates as often being untenable. Our model allowing time-varying frailty overcomes part of this concern. There are further difficulties in estimation however, especially with a semiparametric intensity model. This is under investigation.

Fortran and R software for fitting the correlated gamma frailty model described in this paper is available at `www.maths.lancs.ac.uk/~henderr1/sercor.dir`.

ACKNOWLEDGEMENT

We thank Nouraddin Moussavinasab and Ian Warnell for providing the patient-controlled analgesia data. We are grateful to the editor and three referees for helpful comments. The research of S. Shimakura was supported in part by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brazil.

REFERENCES

AALEN, O. O. (1987). Mixing distributions on a Markov chain. *Scand. J. Statist.* **14**, 281–9.
ALBERT, P. S. (1991). A two-state Markov mixture model for a time series of epileptic seizure counts. *Biometrics* **47**, 1371–81.
BAPAT, R. B. (1989). Infinite divisibility of multivariate gamma distributions and *M*-matrices. *Sankyhā* A **51**, 73–8.
CHAN, K. S. & LEDOLTER, J. (1995). Monte Carlo EM estimation for time series models involving counts. *J. Am. Statist. Assoc.* **90**, 242–52.
DIGGLE, P. J., LIANG, K.-Y. & ZEGER, S. (1994). *Analysis of Longitudinal Data.* Oxford University Press.
GODAMBE, V. P. & HEYDE, C. C. (1987). Quasi-likelihood and optimal estimation. *Int. Statist. Rev.* **55**, 231–44.
GRIFFITHS, R. C. (1970). Infinitely divisible multivariate gamma distributions. *Sankhyā* A **32**, 393–404.
GRIFFITHS, R. C. (1984). Characterization of infinitely divisible multivariate gamma distributions. *J. Mult. Anal.* **15**, 13–20.
HAY, J. L. & PETTITT, A. N. (2001). Bayesian analysis of a time series of counts with covariates: an application to the control of an infectious disease. *Biostatistics* **2**, 433–44.
HOUGAARD, P. (2000). *Analysis of Multivariate Survival Data.* New York: Springer.
HOUGGARD, P., LEE, M.-L. & WHITMORE, G. A. (1997). Analysis of overdispersed count data by mixtures of Poisson variables and Poisson processes. *Biometrics* **53**, 1225–38.
JOWAHEER, V. & SUTRADHAR, B. C. (2002). Analysing longitudinal count data with overdispersion. *Biometrika* **89**, 389–99.
KELSALL, J. E., ZEGER, S. L. & SAMET, J. M. (1999). Frequency domain log-linear models; air pollution and mortality. *Appl. Statist.* **48**, 331–44.
KRISHNAIAH, P. R. & RAO, M. M. (1961). Remarks on a multivariate gamma distribution. *Am. Math. Mon.* **68**, 342–6.
KRISHNAMOORTHY, A. S. & PARTHASARATHY, M. (1951). A multivariate gamma-type distribution. *Ann. Math. Statist.* **22**, 549–57. Correction (1960), **31**, 229.
LAWLESS, J. F., WIGG, M. B., TULI, S., DRAKE, J. & LAMBERTI-PASCULLI, M. (2001). Analysis of repeated failures or durations, with application to shunt failures for patients with hydocephalus. *Appl. Statist.* **50**, 449–66.
LIANG, K.-Y. (1987). Extended Mantel-Haenzel estimating procedure for multivariate logistic regression models. *Biometrics* **43**, 289–99.
LINDSAY, B. G. (1988). Composite likelihood methods. *Contemp. Math.* **80**, 221–39.
MORAN, P. A. P. & VERE-JONES, D. (1969). The infinite divisibility of multivariate gamma distributions. *Sankhyā* A **31**, 191–4.
PETERSEN, J. H. (1998). An additive frailty model for correlated life times. *Biometrics* **54**, 646–61.
THALL, P. F. (1988). Mixed Poisson likelihood regression models for longitudinal interval count data. *Biometrics* **44**, 197–209.
THALL, P. F. & VAIL, S. C. (1990). Some covariance models for longitudinal count data with overdispersion. *Biometrics* **46**, 657–71.
VAN DUIJN, M. A. J. & BÖCKENHOLT, U. (1995). Mixture models for the analysis of repeated count data. *Appl. Statist.* **44**, 473–85.
ZEGER, S. L. (1988). A regression model for a time series of counts. *Biometrika* **75**, 621–9.
ZEGER, S. L. & BROOKMEYER R. (1986). Regression analysis with censored autocorrelated data. *J. Am. Statist. Assoc.* **81**, 722–9.

[*Received April* 2001. *Revised June* 2002]