A Regression Model for Time Series of Counts

Author(s): Scott L. Zeger

Source: *Biometrika*, Dec., 1988, Vol. 75, No. 4 (Dec., 1988), pp. 621-629

Published by: Oxford University Press on behalf of Biometrika Trust

Stable URL: https://www.jstor.org/stable/2336303

Biometrika (1988), **75**, 4, *pp.* 621–9
Printed in Great Britain

A regression model for time series of counts

BY SCOTT L. ZEGER
Department of Biostatistics, Johns Hopkins University, Baltimore, Maryland 21205, U.S.A.

SUMMARY

This paper discusses a model for regression analysis with a time series of counts. Correlation is assumed to arise from an unobservable process added to the linear predictor in a log linear model. An estimating equation approach used for parameter estimation leads to an iterative weighted and filtered least-squares algorithm. Asymptotic properties for the regression coefficients are presented. We illustrate the technique with an analysis of trends in U.S. polio incidence since 1970.

Some key words: Dependence; Estimating equation; Log linear; Parameter driven; Poisson; Quasilikelihood; Regression.

1. INTRODUCTION

This paper develops methods for regression when the outcomes, y_t, are a time series of counts. The objective of analysis is to describe $\mu_t = E(y_t)$ as a function of a $p\times 1$ vector of covariates, x_t. With independent observations, log linear models can be used for this purpose. Under the log linear model, $\mu_t = \exp(x_t'\beta)$ where β is a $p\times 1$ vector of unknown coefficients to be estimated. If y_t is Poisson, likelihood methods can be used to estimate β. Under the Poisson model, $\text{var}(y_t) = \mu_t$. It is common in practice, however, to find $\text{var}(y_t) > \mu_t$. In this case, quasilikelihood methods (Wedderburn, 1974; McCullagh, 1983; McCullagh & Nelder, 1983, Ch. 8) which allow a variety of variance-mean relations are appropriate. The two most common assumptions are: (i) $\text{var}(y_t) = \mu_t\phi$; (ii) $\text{var}(y_t) = \mu_t + \mu_t^2\sigma^2$, where ϕ and σ^2 are unknown scale parameters. See Cox (1983) or Breslow (1984) for a discussion of overdispersion in log linear models.

With time series, it is unlikely that neighbouring observations are independent. Extensions of log linear models which account for dependence are necessary to obtain valid inferences about the relationship of y_t and x_t. Cox (1981) characterized two classes of models of time-dependent data: observation-driven, and parameter-driven models. In an observation-driven model, the conditional distribution of y_t is specified as a function of past observations, $y_{t-1}, \ldots, y_1$. Autoregressive models for Gaussian series and Markov chains for discrete data are examples. See Kaufmann (1987) and Zeger & Qaqish (1988) for recent discussions of observation-driven generalized linear models.

In parameter-driven models, autocorrelation is introduced through a latent process. Let $\theta_t = \log\mu_t$ be the canonical parameter for the log linear model. Then θ_t is assumed to depend on an unobservable noise process, ε_t, that is $\theta_t = \theta(\varepsilon_t, y_{t-1}, \ldots, y_1)$. For example, y_t given ε_t might be Poisson with $E(y_t \mid \varepsilon_t) = \exp(x_t'\beta)\varepsilon_t$. The latent process, ε_t, introduces both overdispersion and autocorrelation in y_t. Parameter-driven models in regression analysis have previously been discussed by West, Harrison & Migon (1985) from a Bayesian perspective. See also Azzalini (1982), Stiratelli, Laird & Ware (1984) and Anderson & Aitkin (1985).

The present paper develops a parameter-driven extension of log linear models focusing on the regression problem. An estimating equation approach analogous to quasilikelihood is used to estimate regression coefficients. A computationally simple approximation is proposed that leads to an iterative weighted and filtered least-squares algorithm. To illustrate the methodology, we estimate the trend in U.S. polio incidence data and present results of a brief simulation study.

2. Parameter-driven model

We specify only the first two moments of y_t, because this is all that is needed for the estimating equation approach to regression used in § 3. Conditional on ε_t, a weakly stationary latent process, y_t is assumed to follow a log linear model. The marginal moments of y_t can then be expressed as a function of the log linear regression coefficients and the parameters of ε_t.

Conditional on a latent process, ε_t, suppose y_t is a sequence of independent counts with mean and variance given by

$$u_t = E(y_t \mid \varepsilon_t) = \exp(x_t'\beta)\varepsilon_t, \quad w_t = \operatorname{var}(y_t \mid \varepsilon_t) = u_t. \tag{1}$$

Also suppose ε_t is an unobserved stationary process with $E(\varepsilon_t)=1$ and $\operatorname{cov}(\varepsilon_t, \varepsilon_{t+\tau}) = \sigma^2\rho_\varepsilon(\tau)$. Then the marginal moments of y_t are

$$\mu_t = E(y_t) = \exp(x_t'\beta), \quad v_t = \operatorname{var}(y_t) = \mu_t + \sigma^2\mu_t^2, \tag{2}$$

$$\rho_y(t, \tau) = \operatorname{corr}(y_t, y_{t+\tau}) = \frac{\rho_\varepsilon(\tau)}{[\{1+(\sigma^2\mu_t)^{-1}\}\{1+(\sigma^2\mu_{t+\tau})^{-1}\}]^{\frac{1}{2}}}. \tag{3}$$

The moment restriction, $E(\varepsilon_t)=1$, forces the unconditional mean, μ_t, to depend only on $x_t'\beta$ and not on moments of the ε_t series, as is desirable in the regression content. However, this restriction is largely conventional. As long as ε_t is mean stationary, $\log \mu_t = c + x_t'\beta$. Hence all coefficients except the intercept are invariant under changing assumptions about $E(\varepsilon_t)$.

The latent process, ε_t, introduces both overdispersion and autocorrelation into y_t. By (2), $\operatorname{var}(y_t) = \mu_t(1+\sigma^2\mu_t)$ so that the degree of overdispersion relative to a Poisson variable depends on μ_t. This is in contrast to the assumption of constant overdispersion (Cox, 1983). The autocorrelation in y_t must be less than or equal that in ε_t. The degree of autocorrelation in y_t relative to ε_t decreases as μ_t and σ^2 decrease.

3. Estimation

3·1. *Estimation of β*

This section considers estimation of the regression parameters, β, given consistent estimates of the covariance parameters, $\theta = (\sigma^2, \theta_\rho)$, where θ_ρ completely specify the autocorrelation function $\rho_\varepsilon(\tau; \theta_\rho)$. An estimating equation like the one used in quasilikelihood and by Liang & Zeger (1986) for longitudinal data is proposed. An approximation which leads to simpler computation in the time series case is then introduced. Section 3·2 describes moment estimators for θ. Relative efficiencies of the two estimators from § 3·1 are discussed in § 3·3.

The estimating equation approach used here is a time series analogue of quasilikelihood as discussed by McCullagh (1983) for independent data. In that case, $\hat\beta$ is the root of the equation

$$U(\beta)=\sum_{t=1}^{n} \frac{\partial \mu_t'}{\partial \beta} v_t^{-1}(y_t-\mu_t)=0. \tag{4}$$

The quasilikelihood estimator is consistent, asymptotically Gaussian and is optimal in an extended Gauss–Markov sense (McCullagh, 1983). This approach is also robust in that consistent inferences about β can be made given only that $E(y_t)=\mu_t$ whether or not $v_t=\operatorname{var}(y_t)$. Our strategy is to generalize the quasilikelihood estimating equation to the time series case. Letting

$$y=(y_1,\ldots,y_n)', \quad X=(x_1,\ldots,x_n)', \quad \mu=(\mu_1,\ldots,\mu_n)', \quad V=\operatorname{var}(y),$$

equation (4) can be rewritten

$$\frac{\partial \mu'}{\partial \beta} V^{-1}(y-\mu)=0. \tag{5}$$

With independent data, V is diagonal. With time series data, V will include off-diagonal terms which depend on nuisance parameters. Nevertheless, solving (5) for $\hat\beta$ remains an effective approach with dependent observations. For example, Liang & Zeger (1986) use (5) when V is a block diagonal matrix, each block representing the covariance among repeated observations for one subject in a longitudinal study.

To adapt (5) for time series data, let R_ε be an $n\times n$ matrix with j, k element $\rho_\varepsilon(|j-k|)$. Then, for the parameter-driven model, $V=\operatorname{var}(y)=A+\sigma^2 A R_\varepsilon A$, where $A=\operatorname{diag}(\mu_1,\ldots,\mu_n)$. We propose to estimate β by solving the $p\times 1$ system of equations

$$\frac{\partial \mu'}{\partial \beta} V^{-1}(\beta,\theta)(y-\mu)=0. \tag{6}$$

Equation (6) depends on β but also on the nuisance parameters, θ. Let $\hat\theta$ be a $\surd n$-consistent estimate of θ depending on the observations and on β. Let $\hat\beta$ be the solution of

$$U(\beta)=\frac{\partial \mu'}{\partial \beta} V^{-1}\{\beta,\hat\theta(\beta)\}(y-\mu)=0. \tag{7}$$

We have the following.

PROPOSITION 1. *Suppose ε_t is a stationary process. Under mild regularity conditions and given $\surd n(\hat\theta-\theta)=o_p(1)$ for some fixed θ, $\surd n(\hat\beta-\beta)$ is asymptotically multivariate Gaussian with zero mean and covariance matrix*

$$V_{\hat\beta}=\lim_{n\to\infty}\left(\frac{\partial \mu'}{\partial \beta} V^{-1}\frac{\partial \mu}{\partial \beta}\Big/ n\right)^{-1}.$$

The proof relies on the fact that y_t is a mixing process when ε_t is since each y_t is conditionally independent given ε_t. Otherwise it is analogous to that of Liang & Zeger (1986) and is omitted.

To compute $\hat\beta$ for a given value of $\hat\theta(\beta)$, an iterative weighted least-squares can be used as is the case in quasilikelihood with independent data. The parameter estimates at the $j+1$st iteration, $\hat\beta^{(j+1)}$, are given by

$$\hat\beta^{(j+1)}=\left(\frac{\partial \mu'}{\partial \beta} V^{-1}\frac{\partial \mu}{\partial \beta}\right)^{-1}\left(\frac{\partial \mu'}{\partial \beta} V^{-1} Z\right), \tag{8}$$

where $Z=(\partial\mu'/\partial\beta)\beta+(y-\mu)$ and the right-hand side is evaluated using $\hat\beta^{(j)}$ for β.

Given an estimation procedure for $\hat{\theta}(\beta)$, $\hat{\beta}$ is found by alternately solving (8) for $\hat{\beta}^{(j+1)}$ given $\hat{\theta}^{(j)}$, then using the updated $\hat{\beta}^{(j+1)}$ to find $\hat{\theta}^{(j+1)}$ until convergence.

A drawback of (8) is its solution requires inversion of the $n \times n$ covariance matrix, V. We are unaware of an efficient algorithm for inverting matrices with the structure of V. Hence, we consider an approximation to (6) that is computationally simpler and leads to nearly efficient estimators in many practical cases.

Inversion of V is difficult because the parameter-driven process does not have a stationary autocorrelation function. To simplify computation, we approximate the actual autocorrelation matrix, R_ε, by a band diagonal matrix, corresponding to an autoregressive process. Let $D = \text{diag}\,(\mu_t + \sigma^2 \mu_t^2)$. We approximate V with $V_R = D^{\frac{1}{2}} R(\alpha) D^{\frac{1}{2}}$, where $R(\alpha)$ is the autocorrelation matrix for a stationary autoregressive process with an $s \times 1$ vector of parameters, α. Let $\theta_R = (\sigma^2, \alpha)$ and suppose $\hat{\theta}_R$ is a $\sqrt{n}$-consistent estimate of θ_R. Now define $\hat{\beta}_R$ to be the solution of the estimating equation

$$\frac{\partial \mu'}{\partial \beta} V_R(\hat{\theta}_R)^{-1}(y - \mu) = 0. \tag{9}$$

First note that the algorithm for finding $\hat{\beta}_R$ given $\hat{\theta}_R$ is greatly simplified. The inverse of the matrix, V_R^{-1}, ignoring edge effects, satisfies $V_R^{-1} \simeq D^{-\frac{1}{2}} L' L D^{-\frac{1}{2}}$, where L is the matrix which applies the autoregressive filter, i.e. the elements of Ly are

$$y_t - \alpha_1 y_{t-1} - \ldots - \alpha_p y_{t-p} \quad (t > p).$$

The iterative weighted least-squares procedure now has the form

$$\hat{\beta}_R^{(j+1)} = \left\{ \left(L D^{-\frac{1}{2}} \frac{\partial \mu}{\partial \beta} \right)' \left(L D^{-\frac{1}{2}} \frac{\partial \mu}{\partial \beta} \right) \right\}^{-1} \left(L D^{-\frac{1}{2}} \frac{\partial \mu}{\partial \beta} \right)' (L D^{-\frac{1}{2}} Z).$$

We refer to this algorithm as an iterative weighted and filtered least-squares because it involves the following steps:

(i) weight the current values of $\partial \mu / \partial \beta$ and Z by the inverses of the standard deviations, $D^{-\frac{1}{2}}$;

(ii) filter the normalized values, $D^{-\frac{1}{2}} \partial \mu / \partial \beta$ and $D^{-\frac{1}{2}} Z$, with a filter for an autoregressive process of order p;

(iii) solve the least-squares equations;

(iv) iterate (i) to (iii) to convergence.

The following gives the asymptotic properties of $\hat{\beta}_R$.

PROPOSITION 2. *Under the assumptions of Proposition* 1 *and assuming the matrices* I_0 *and* I_1 *exist,* $\sqrt{n}(\hat{\beta}_R - \beta)$ *is asymptotically Gaussian with mean* 0 *and covariance matrix*

$$V_{\hat{\beta}_R} = I_0^{-1} I_1 I_0^{-1},$$

where

$$I_0 = \lim_{n \to \infty} \left(\frac{\partial \mu'}{\partial \beta} V_R^{-1} \frac{\partial \mu}{\partial \beta} \Big/ n \right), \quad I_1 = \lim_{n \to \infty} \left(\frac{\partial \mu'}{\partial \beta} V_R^{-1} V V_R^{-1} \frac{\partial \mu}{\partial \beta} \Big/ n \right).$$

The primary distinction between this result and Proposition 1 is that the matrix, V_R, in the estimating equation is not the actual covariance matrix. This leads to the more complicated form for the asymptotic variance of $\hat{\beta}_R$.

3·2. *Estimation of σ^2 and θ_ρ*

The nuisance parameters can be estimated by a method of moments. Note that $\operatorname{var}(y_t) = \mu_t + \sigma^2\mu_t^2$. Hence σ^2 can be estimated by

$$\hat{\sigma}^2 = \sum_{t=1}^{n} \{(y_t - \hat{\mu}_t)^2 - \hat{\mu}_t\} \Big/ \sum_{t=1}^{n} \hat{\mu}_t^2. \tag{10}$$

The autocorrelation function of ε_t can similarly be estimated by

$$\hat{\rho}_\varepsilon(\tau) = \hat{\sigma}^{-2} \sum_{t=\tau+1}^{n} \{(y_t - \hat{\mu}_t)(y_{t-\tau} - \hat{\mu}_{t-\tau})\} \Big/ \sum_{t=\tau+1}^{n} \hat{\mu}_t \hat{\mu}_{t-\tau}. \tag{11}$$

When $\rho_\varepsilon(\tau)$ is fully specified in terms of fewer parameters, θ_ρ, these can be estimated in the traditional way from $\hat{\rho}_\varepsilon(\tau)$. For example, if ε_t is assumed to be a first-order autoregressive process, then $\rho_\varepsilon(\tau) = \theta_\rho^\tau$ and we estimate θ_ρ by $\hat{\rho}_\varepsilon(1)$, the first lag autocorrelation. More generally, the Yule–Walker equations for $\hat{\theta}_\rho$ given $\hat{\rho}_\varepsilon(\tau)$ can be solved. One limitation of moment estimation is that $\hat{\sigma}^2$ can be negative and $\hat{\rho}_\varepsilon(\tau)$ is not constrained to the interval $(-1, 1)$. When the sample size, n, is small and $|\rho_\varepsilon(\tau)|$ large, a different approach may be needed.

3·3. *Relative efficiencies*

Section 3·1 proposes two regression estimators. The first, $\hat{\beta}$, is more efficient but computationally expensive relative to the second, $\hat{\beta}_R$. Here we calculate the large-sample efficiency of $\hat{\beta}_R$ relative to $\hat{\beta}$ for a few simple models. We assume ε_t has autocorrelation function, $\rho_\varepsilon(\tau) = \rho^\tau$ with $\rho = -0{\cdot}4$, $0{\cdot}4$ or $0{\cdot}8$ and variances, σ^2, taking values $0{\cdot}1$, $1{\cdot}0$ or $5{\cdot}0$. Then $\hat{\beta}_R$ is based upon the working assumption that the Pearson residuals, $\hat{v}_t^{-\frac{1}{2}}(y_t - \hat{\mu}_t)$, have the same correlation structure as ε_t. Three 100×2 matrices are used; each includes an intercept and one of the following covariates: trend $(x_t = t - 50$, $t = 1, \ldots, 100)$; random Gaussian variates, x_t from $N(0, 1)$; and a step function $(x_t = -\frac{1}{2}$, $t < 50$; $x_t = \frac{1}{2}$, $t \geqslant 50)$. The true regression coefficients were chosen so that expected counts ranged between about $0{\cdot}5$ and $8{\cdot}0$. The large-sample properties were assumed to hold for sample size $n = 100$. Table 1 lists the ratio of the large-sample variances for the regression coefficient of x_t given in $V_{\hat{\beta}}$ and $V_{\hat{\beta}_R}$.

Note that $\hat{\beta}_R$ is nearly fully efficient relative to $\hat{\beta}$ for the trend and step function and about 90% efficient for the $N(0, 1)$ covariate. The efficiency does not depend strongly on the degree of the extra-Poisson variation, σ^2, nor on the first lag autocorrelation, ρ. Little appears to be lost in these cases for the large computational saving made by using $\hat{\beta}_R$.

Table 1. *Large-sample efficiency of $\hat{\beta}_R$ relative to $\hat{\beta}$ for parameter-driven process* (1)

	Relative efficiency								
	$\sigma^2 = 0{\cdot}1$			$\sigma^2 = 1{\cdot}0$			$\sigma^2 = 5{\cdot}0$		
	ρ			ρ			ρ		
Covariate	−0·4	0·4	0·8	−0·4	0·4	0·8	−0·4	0·4	0·8
Trend	1·0	0·99	0·95	0·99	1·0	0·97	1·0	1·0	0·99
$N(0, 1)$	0·99	0·99	0·91	0·99	0·99	0·87	1·0	0·99	0·90
Step function	1·0	1·0	0·97	1·0	0·93	0·93	1·0	1·0	0·98

It is assumed $\operatorname{corr}(\varepsilon_t, \varepsilon_{t+\tau}) = \rho^\tau$, $\operatorname{var}(\varepsilon_t) = \sigma^2$. True regression coefficients (β_0, β_1): trend, $(1{\cdot}0, 0{\cdot}02)$; $N(0, 1)$, $(1{\cdot}0, 0{\cdot}5)$; step function, $(1{\cdot}0, 1{\cdot}0)$.

4. EXAMPLE AND SIMULATION

We have applied the parameter-driven model to disease incidence data as an illustration. Table 2 lists the monthly number of cases of poliomyelitis reported by the U.S. Centers for Disease Control for the years 1970 to 1983. Of interest is whether this record provides evidence of a long-term decrease in the rate of U.S. polio infection. As there is evidence of seasonality, we have regressed the monthly number of cases on a linear trend as well as sine, cosine pairs at the annual and semi-annual frequencies. Table 3 reports the results from fitting three models: I, parameter-driven model using $\hat{\beta}_R$; II, log linear model with var $(y_t) = \mu_t\phi$; and III, log linear model with var $(y_t) = \mu_t + \sigma^2\mu_t^2$. Model I is capable of accounting for autocorrelation; Models II and III assume repeated observations are independent. The November 1972 observation is an outlier but was left in the analysis since it had a minor affect on the findings.

Table 2. *Monthly number of U.S. cases of poliomyelitis for* 1970 *to* 1983

	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
1970	0	1	0	0	1	3	9	2	3	5	3	5
1971	2	2	0	1	0	1	3	3	2	1	1	5
1972	0	3	1	0	1	4	0	0	1	6	14	1
1973	1	0	0	1	1	1	1	0	1	0	1	0
1974	1	0	1	0	1	0	1	0	1	0	0	2
1975	0	1	0	1	0	0	1	2	0	0	1	2
1976	0	3	1	1	0	2	0	4	0	2	1	1
1977	1	1	0	1	1	0	2	1	3	1	2	4
1978	0	0	0	1	0	1	0	2	2	4	2	3
1979	3	0	0	2	7	8	2	4	1	1	2	4
1980	0	1	1	1	3	0	0	0	0	1	0	1
1981	1	0	0	0	0	0	1	2	0	2	0	0
1982	0	1	0	1	0	1	0	2	0	0	1	2
1983	0	1	0	0	0	1	2	1	0	1	3	6

Reported to the U.S. Centers for Disease Control and published in Morbidity and Mortality Weekly Report Annual Summary (1970–1983).

The coefficients are very similar across models and indicate a decrease in the rate of polio cases per month from about 1·7 to 0·8 over the period 1970 to 1983. Model I which accounts for autocorrelation has standard errors about 40% larger than Models II and III which assume observations are independent. Under the incorrect assumption that repeated observations are independent, these data are interpreted as providing strong evidence of decreasing polio incidence. Under the parameter-driven model, the evidence is seen to be weaker.

The lag-one correlation coefficient for the Pearson residuals from Model I is 0·25. This corresponds to a lag-one correlation of 0·77 in ε_t. The residual autocorrelation is statistically significant as determined by a Monte Carlo experiment in which the correlation coefficients for 100 time series, each being a random reordering of the Pearson residuals, all fell in the interval $(-0{\cdot}15, 0{\cdot}19)$. The increase in the estimated standard error of about 40% when correlation is taken into account is slightly larger than would be expected for Gaussian data. In a linear regression model with the same design matrix and Gaussian first order autoregressive errors with lag-one correlation of 0·25, the expected increase is 28%.

Table 3. *Coefficients and standard errors from regressing polio incidence on a linear trend and* (sin, cos) *pairs at the annual and semi-annual frequencies using: the parameter-driven model with first-order autoregressive errors; log linear model with* $\text{var}(y_t) = \mu_t \phi$; *and log linear model with* $\text{var}(y_t) = \mu_t + \sigma^2 \mu_t^2$.

	Parameter-driven model		Log linear model $\text{var}(y_t)=\mu_t\phi$		Log linear model $\text{var}(y_t)=\mu_t\phi+\mu_t^2\sigma^2$	
	$\hat\beta$	Std err.	$\hat\beta$	Std err.	$\hat\beta$	Std err.
Intercept (log rate in Jan. 1976)	0·17	0·13	0·15	0·11	0·15	0·10
Trend $\times 10^{-3}$	−4·35	2·68	−4·80	1·94	−4·28	2·06
$\cos(2\pi t/12)$	−0·11	0·16	−0·15	0·13	−0·14	0·14
$\sin(2\pi t/12)$	−0·48	0·17	−0·53	0·15	−0·49	0·15
$\cos(2\pi t/6)$	0·20	0·14	0·18	0·14	0·18	0·14
$\sin(2\pi t/6)$	−0·41	0·14	−0·43	0·14	−0·42	0·14
$\hat\phi$	1·0		1·91		1·0	
$\hat\sigma^2$	0·77		0·0		0·80	
$\hat\rho_y(1)$	0·25		0·0		0·0	
$\hat\rho_\varepsilon(1)$	0·77					

A simulation was conducted to examine whether these findings hold on average over many trials. One hundred realizations on the parameter-driven model fitted to the polio data were generated. To guarantee ε_t was positive, $\delta_t = \log \varepsilon_t$ was assumed to be a Gaussian autoregressive process of order 1 with lag-one correlation $\rho_\delta = 0{\cdot}82$ and variance $\sigma_\delta^2 = 0{\cdot}57$ which gives the moments for ε_t estimated in the example. For each realization, Models I–III were fitted and the coefficients and their estimated variances stored. When fitting Model I, the computationally faster estimator, $\hat\beta_R$ was used. We consider only the trend coefficient, β.

For all three models, $\hat\beta$ was estimated to have positive bias of about 10%. The average of the variance estimates for $\hat\beta$ under Models I–III were: 1·4, 0·34 and $0{\cdot}36 \times 10^{-5}$ respectively. The variance of the parameter-driven coefficient was estimated by Monte Carlo to be $1{\cdot}67 \times 10^{-5}$ corresponding to a 20% underestimate of the variance of $\hat\beta$ in the parameter-driven model. The efficiency of the three methods was about the same for this design. Hence, while the variance estimates from Model I are on average 20% too small; those from Models II and III are too small by a factor of about 4.

The estimator of σ^2 in (10) was approximately 15% biased in this simulation with a mean of 0·49 and interquartile range of 0·32 to 0·61 when the true value was 0·57. The estimator of $\rho_\varepsilon(1)$ in equation (11) was also negatively biased. The true value was 0·82 while the average estimate was 0·70 with interquartile range 0·59 to 0·90.

For the polio design matrix, the benefit of accounting for autocorrelation is correct inferences rather than increased efficiency. However, in other designs, the parameter-driven model gives more efficient estimates as well. To illustrate, the relative efficiency of regression coefficients from Models I and III have been compared by simulation for the three design matrices in § 3·3. Time series of length 100 were generated as described above. Table 4 presents this relative efficiency as well as the ratio of the average variance estimate to the actual variance for each model.

For the linear trend and step function, ignoring correlation leads to estimators as efficient as obtained by the parameter driven model. This is not surprising since, in the Gaussian case, least-squares is nearly efficient for these designs. Note, however, that with the Gaussian x's, Model III is inefficient relative to Model I, the inefficiency increasing

Table 4. *Efficiency, $E = \mathrm{var}(\hat{\beta}_R)/\mathrm{var}(\hat{\beta}_{\mathrm{III}})$, of regression coefficient from Model* III *relative to $\hat{\beta}_R$ from Model* I *and ratios, R_{I} and R_{III}, of average model based estimate of variance of $\hat{\beta}$ to actual variance, estimated by Monte Carlo, for Models* I *and* III *respectively*

			Trend			$N(0,1)$			Step function	
σ_δ^2	ρ_δ	E	R_{I}	R_{III}	E	R_{I}	R_{III}	E	R_{I}	R_{III}
0·25	0·00	1·07	0·74	0·78	1·08	1·09	1·20	1·00	1·15	1·04
	0·25	0·98	0·92	0·71	0·97	0·75	0·73	1·00	1·03	0·84
	0·50	1·04	1·24	0·64	0·97	0·78	0·79	1·00	0·99	0·56
	0·75	1·09	0·98	0·37	0·82	1·18	1·06	1·00	0·78	0·28
0·50	0·00	1·05	0·88	0·94	1·04	0·87	0·92	0·99	0·97	0·93
	0·25	1·07	0·75	0·63	0·94	0·98	0·93	1·02	1·00	0·75
	0·50	1·03	0·89	0·46	0·93	1·12	1·14	0·97	1·14	0·56
	0·75	1·02	0·84	0·27	0·79	1·02	1·01	0·95	0·80	0·22
1·00	0·00	1·05	0·83	0·90	1·04	0·87	0·97	1·00	0·94	0·94
	0·25	1·03	0·90	0·69	0·94	0·84	0·83	0·99	0·85	0·63
	0·50	1·03	0·86	0·48	0·81	0·88	0·82	0·94	0·67	0·32
	0·75	1·05	0·60	0·20	0·57	1·19	0·90	0·95	0·80	0·23

$\varepsilon_t = \exp(\delta_t)$, where δ_t is Gaussian autoregressive process of order 1 with variance σ_δ^2, lag-one correlation ρ_δ. Design matrices as in Table 1.

with σ_δ^2 and ρ_δ. For example, when $\sigma_\delta^2 = 1\cdot0$ and $\rho_\delta = 0\cdot75$, Model I is estimated to be nearly twice as efficient as Model III.

As in the polio example, the parameter-driven model leads to more valid inferences, particularly for the trend and step-function designs. The estimated variance of $\hat{\beta}$ under Model I has acceptable finite sample bias in nearly every case. Model III underestimates the variance of $\hat{\beta}$ by as much as a factor of five.

Note that our simulation does not include seasonal effects.

5. Discussion

In the regression context, it is desirable to formulate parameter-driven models so that the marginal expectation, μ_t, depends on regression parameters but not on unknown parameters of the latent process. In this way, the interpretation of regression coefficients does not depend on assumptions about the unobservable process, ε_t, which are difficult to verify. Generalized linear models with linear and log links can be extended to parameter-driven models with this property; for other link functions, it may not be possible.

The variance function in the parameter-driven model (1) has the negative binomial form rather than being proportional to the mean. Hence the degree of overdispersion relative to the Poisson depends on μ_t. These two variance-mean relations give similar regression inferences in problems where the fitted values vary over a relatively narrow range such as was the case in the polio example above. An interesting question is how to generate the constant overdispersion discussed by Cox (1983) while also introducing dependence in time series of counts.

This paper does not address theoretical properties of the iterative weighted and filtered least-squares algorithm. However, in the simulation studies, the algorithm converged in every realization using Poisson regression to obtain starting values. The algorithm does not appear to be sensitive to the choice of starting values.

The parameter-driven model discussed here is a time series analogue of Morton's (1987) model for counts with strata of variation. In Morton's problem, observations are obtained in clusters. Those within a cluster share a common multiplicative error and are correlated; those from distinct clusters are independent. For data sets with many clusters, inferences about the regression can be made robust to misspecification of the time dependence following the approach of Zeger, Liang & Self (1985). With single time series, independent subsets of data from which robust variances can be derived are not available. Hence inferences depend on correct specification of the time series model.

Acknowledgements

The author acknowledges support from the Australian Commonwealth Scientific and Industrial Research Organization, Division of Mathematics and Statistics, the U.S. National Institute of Mental Health, and the U.S. National Institute of Health. Thanks are due to a referee whose comments substantially improved an earlier draft.

References

ANDERSON, D. A. & AITKIN, M. (1985). Variance component models with binary response: interviewer variability, *J. R. Statist. Soc.* B **47**, 203–10.
AZZALINI, A. (1982). Approximate filtering of parameter driven processes. *J. Time Series Anal.* **3**, 219–23.
BRESLOW, N. E. (1984). Extra-poisson variation in log-linear models. *Appl. Statist.* **33**, 38–44.
CENTERS FOR DISEASE CONTROL (1970–1983). Morbidity and Mortality Weekly Report Annual Summary: Reported Morbidity and Mortality in the U.S. Vol. 18–31, No. 54.
COX, D. R. (1981). Statistical analysis of time series, some recent developments. *Scand. J. Statist.* **8**, 93–115.
COX, D. R. (1983). Some remarks on over-dispersion. *Biometrika* **70**, 267–74.
KAUFMANN, H. (1987). Regression methods for non-stationary categorical time series: Asymptotic estimation theory. *Ann. Statist.* **17**, 79–98.
LIANG, K.-Y. & ZEGER, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22.
MCCULLAGH, P. (1983). Quasi-likelihood functions. *Ann. Statist.* **11**, 59–67.
MCCULLAGH, P. & NELDER, J. A. (1983). *Generalized Linear Models.* London: Chapman and Hall.
MORTON, R. (1987). A generalized linear model with nested strata of extra-Poisson variation. *Biometrika* **74**, 247–58.
STIRATELLI, R., LAIRD, N. & WARE, J. H. (1985). Random effects models for serial observations with binary response. *Biometrics* **40**, 961–71.
WEDDERBURN, R. W. M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss–Newton method. *Biometrika* **61**, 439–47.
WEST, M., HARRISON, J. P. & MIGON, H. S. (1985). Dynamic generalized linear models and Bayesian forecasting. *J. Am. Statist. Assoc.* **80**, 73–96.
ZEGER, S. L., LIANG, K.-Y. & SELF, S. G. (1985). The analysis of binary longitudinal data with time independent covariates. *Biometrika* **72**, 31–8.
ZEGER, S. L. & QAQISH, B. (1988). Markov regression models for time series; a quasi-likelihood approach. *Biometrics.* To appear.

[*Received October* 1986. *Revised June* 1988]