Biometrika (2009), **96**, 3, *pp.* 735–749
doi: 10.1093/biomet/asp029
Advance Access publication 1 July 2009

A negative binomial model for time series of counts

BY RICHARD A. DAVIS

Department of Statistics, Columbia University, New York, New York 10027, U.S.A.
rdavis@stat.columbia.edu

AND RONGNING WU

Department of Statistics and Computer Information Systems, Baruch College, The City University of New York, New York, New York 10010, U.S.A.
rongning.wu@baruch.cuny.edu

SUMMARY

We study generalized linear models for time series of counts, where serial dependence is introduced through a dependent latent process in the link function. Conditional on the covariates and the latent process, the observation is modelled by a negative binomial distribution. To estimate the regression coefficients, we maximize the pseudolikelihood that is based on a generalized linear model with the latent process suppressed. We show the consistency and asymptotic normality of the generalized linear model estimator when the latent process is a stationary strongly mixing process. We extend the asymptotic results to generalized linear models for time series, where the observation variable, conditional on covariates and a latent process, is assumed to have a distribution from a one-parameter exponential family. Thus, we unify in a common framework the results for Poisson log-linear regression models of Davis et al. (2000), negative binomial logit regression models and other similarly specified generalized linear models.

Some key words: Generalized linear model; Latent process; Negative binomial distribution; Time series of counts.

1. INTRODUCTION

The analysis of time series of counts, motivated by applications in various fields, is one of the rapidly developing areas in time series modelling. Such applications include monthly polio counts in the USA (Zeger, 1988; Davis et al., 2000), daily asthma presentation at a hospital in Sydney, Australia (Davis et al., 2000), and traffic accidents in the county of Västerbotten, Sweden (Brännäs & Johansson, 1994). In addition, Campbell (1994) investigated the relationship between sudden infant death syndrome and environmental temperature; Johansson (1996) used time series of counts to assess the effect of lowered speed limits on the number of road casualties; Jørgensen et al. (1996) studied the relationship between respiratory morbidity and air pollution; and Cameron & Trivedi (1996) discussed discrete models for financial data.

For time series consisting of counts, classical Gaussian models are inappropriate and it is necessary to consider nonlinear models. Generalized linear models (McCullagh & Nelder, 1989) are widely used for analyzing counts and other types of discrete data, and thus provide a good starting point. Count data are nonnegative, integer-valued and often overdispersed; the variance is larger than the mean. To accommodate overdispersion, many researchers have turned to overdispersed Poisson and binomial regression models. As a natural extension of the Poisson distribution, the negative binomial distribution is more flexible and allows for overdispersion. In this paper, we propose a negative binomial regression model for time series of counts; the model can be classified as a parameter-driven generalized linear model (Cox, 1981), which in turn can be viewed as a special type of state space model. Specifically, let Y_t and α_t denote the observation and unobservable state, or latent, variables at time t, respectively. The observation equation specifies $p(Y_t \mid x_t, \alpha_t)$, where x_t is a vector of covariates. The state comes into the model via $f(u_t) = x_t^{\mathrm{T}}\beta + \alpha_t$, where $f(\cdot)$ is a link function of the standard generalized linear model and $u_t = E(Y_t \mid x_t, \alpha_t)$. On the other hand, the state equation specifies the serial dependence structure of $\{\alpha_t\}$, which evolves independently of the observed data.

The inclusion of a time-dependent latent process provides a more realistic modelling framework. The latent process not only adds serial dependence to the model, but also can be viewed as a proxy for unknown or unavailable covariates. Without adjusting the noise terms, inferences about the covariates included in the model may be misleading. This point is illustrated with the polio data described in § 5. Even if a time series model is specified for the latent process, the likelihood cannot be expressed in a closed form, which can make the theory for maximum likelihood estimation more difficult. Typically, one needs to resort to simulation-based or other approximation-based techniques for computing the likelihood; see Durbin & Koopman (1997) and Davis & Rodriguez-Yam (2005). Often, it is not feasible to even consider a family of time series models for the latent process until the form of the regression function has been established. So, in many situations, it makes sense to consider estimating parameters in the regression function by maximizing the likelihood that excludes the presence of the latent process. This is akin to using ordinary least squares in linear time series models, which in all but some special cases share the same asymptotic efficiency as the maximum likelihood estimator.

An overview of parameter-driven models for time series of counts can be found in Davis et al. (1999). Zeger (1988) studied Poisson log-linear regression models for a time series of counts. An estimating equation approach was used for parameter estimation, and asymptotic results of the quasilikelihood estimator were established. Harvey & Fernandes (1989) studied a structural model for time series of counts and qualitative data using natural conjugate distributions. Jørgensen et al. (1999) proposed a nonstationary state space model for multivariate longitudinal count data driven by a latent gamma Markov process. Blais et al. (2000) extended Zeger's results to the case where each observation, conditional on a stationary and strongly mixing latent process, is assumed to have an exponential family distribution. Davis et al. (2000) developed a practical approach to diagnosing the existence of a latent process in Poisson log-linear regression models and derived the asymptotic properties of the generalized linear model estimator when an autocorrelated latent process is present. In this paper, we extend the asymptotic results of Davis et al. (2000) first to the negative binomial logit regression model and then to a general set-up, assuming that the conditional distribution of the observed variable is from a one-parameter exponential family. A simulation study is presented that illustrates our asymptotic results, and the developed techniques are applied to the polio data.

2. LARGE SAMPLE PROPERTIES OF GENERALIZED LINEAR MODEL ESTIMATORS

2·1. *Set-up*

Let $\{Y_t\}$ be a time series of counts and suppose that for each t, x_t is an observed l-dimensional covariate, which is assumed to be nonrandom and whose first component is one. In some cases, x_t may depend on the sample size n and form a triangular array x_{nt}. We assume that, conditional on a latent process $\{\alpha_t\}$, the random variables $Y_1, \ldots, Y_n$ are independent, and the conditional distribution of Y_t depends only on α_t and is specified by a negative binomial distribution. To be specific, we consider the following parameter-driven model:

$$Y_t \mid \alpha_t \sim \text{NegBin}(r,\, p_t), \tag{1}$$

where r is a positive number and p_t satisfies the logit model

$$-\log\left(\frac{p_t}{1-p_t}\right) = x_{nt}^{\mathrm{T}}\beta + \alpha_t. \tag{2}$$

Here, $\beta = (\beta_1, \ldots, \beta_l)^{\mathrm{T}}$ is the vector of regression coefficients of primary interest. The conditional density function of Y_t is

$$p(Y_t = y_t \mid \alpha_t) = \binom{y_t + r - 1}{r-1} p_t^r (1-p_t)^{y_t}$$

for $y_t = 0, 1, \ldots$. Instead of dealing with $\{\alpha_t\}$ in the derivation, it is more convenient to use the process $\{\epsilon_t = e^{\alpha_t}\}$, which is a strictly stationary nonnegative time series when $\{\alpha_t\}$ is strictly stationary. Furthermore, we centre $\{\alpha_t\}$ such that $\{\epsilon_t\}$ has mean one. The conditional mean of Y_t can be written in terms of ϵ_t,

$$E(Y_t \mid \alpha_t) = \frac{r(1-p_t)}{p_t} = r\exp\left(x_{nt}^{\mathrm{T}}\beta + \alpha_t\right) = r\exp\left(x_{nt}^{\mathrm{T}}\beta\right)\epsilon_t.$$

With $E(\epsilon_t) = 1$, we have $E(Y_t) = r\exp(x_{nt}^{\mathrm{T}}\beta)$, the form for a pure generalized linear model in the negative binomial case.

We assume that $\{\epsilon_t\}$ is a stationary strongly mixing process in the sense that

$$\alpha(m) = \sup_{A \in \mathcal{F}_{-\infty}^{0},\, B \in \mathcal{F}_m^{\infty}} |\text{pr}(AB) - \text{pr}(A)\text{pr}(B)| \to 0$$

as $m \to \infty$, where $\mathcal{F}_{-\infty}^{0}$ and $\mathcal{F}_m^{\infty}$ are σ-fields generated by $\{\epsilon_t, t \leqslant 0\}$ and $\{\epsilon_t, t \geqslant m\}$, respectively. In addition, we assume that $\{\epsilon_t\}$ satisfies the following two assumptions:

Assumption 1. There exists a positive constant λ such that $E(|\epsilon_t|^{\lambda+4}) < \infty$.

Assumption 2. The mixing coefficient $\alpha(m)$ satisfies $\sum_{m=1}^{\infty} \alpha(m)^{(\lambda+2)/\lambda} < \infty$.

2·2. *Asymptotic properties when $\{\epsilon_t\}$ is strongly mixing*

The generalized linear model estimator $\hat{\beta}_n$ of β is obtained by ignoring the latent process in the model and maximizing the loglikelihood function of the misspecified model. Before deriving the asymptotic properties of $\hat{\beta}_n$, we evaluate the first two moments of the observed process $\{Y_t\}$. The mean μ_t is given as $r\exp(x_{nt}^{\mathrm{T}}\beta)$. Moreover, it follows from the logit model (2) that

$$p_t = \frac{1}{1 + e^{x_{nt}^{\mathrm{T}}\beta}\epsilon_t}, \qquad q_t = 1 - p_t = \frac{e^{x_{nt}^{\mathrm{T}}\beta}\epsilon_t}{1 + e^{x_{nt}^{\mathrm{T}}\beta}\epsilon_t}.$$

Then the variance of Y_t is

$$\begin{aligned}
\text{var}(Y_t) &= E\{\text{var}(Y_t \mid \alpha_t)\} + \text{var}\{E(Y_t \mid \alpha_t)\} \\
&= E\left\{re^{x_{nt}^{\mathrm{T}}\beta}\epsilon_t\left(1 + e^{x_{nt}^{\mathrm{T}}\beta}\epsilon_t\right)\right\} + r^2 e^{2x_{nt}^{\mathrm{T}}\beta}\text{var}(\epsilon_t) \\
&= re^{x_{nt}^{\mathrm{T}}\beta} + re^{2x_{nt}^{\mathrm{T}}\beta}\{\gamma_\epsilon(0) + 1\} + r^2 e^{2x_{nt}^{\mathrm{T}}\beta}\gamma_\epsilon(0) \\
&= \mu_t + \frac{\mu_t^2}{r} + \frac{\mu_t^2\gamma_\epsilon(0)}{r} + \mu_t^2\gamma_\epsilon(0),
\end{aligned}$$

and the autocovariance function is, for $k \neq 0$,

$$\mathrm{cov}(Y_{t+k}, Y_t) = \mathrm{cov}\{E(Y_{t+k} \mid \alpha_{t+k}), E(Y_t \mid \alpha_t)\} + 0 = r^2 e^{\left(x_{n,t+k}^{\mathrm{T}} + x_{nt}^{\mathrm{T}}\right)\beta} \gamma_\epsilon(k) = \mu_{t+k}\mu_t \gamma_\epsilon(k).$$

Let $Y_1, \ldots, Y_n$ be observations from the model (1)–(2) with true parameter β_0. The estimator $\hat{\beta}_n$ maximizes

$$\begin{aligned} \ell_n(\beta) &= r \sum_{t=1}^n \log p_t + \sum_{t=1}^n Y_t \log(1 - p_t) + \log \prod_{t=1}^n \binom{Y_t + r - 1}{r - 1} \\ &= -r \sum_{t=1}^n \log\left(1 + e^{x_{nt}^{\mathrm{T}}\beta}\right) - \sum_{t=1}^n Y_t \log\left(1 + e^{-x_{nt}^{\mathrm{T}}\beta}\right) + \log \prod_{t=1}^n \binom{Y_t + r - 1}{r - 1}. \end{aligned} \qquad (3)$$

Assumptions on the covariates x_{nt} are needed in order to establish the consistency and asymptotic normality of $\hat{\beta}_n$. We assume that there exists a sequence of nonsingular matrices M_n such that x_{nt} satisfy the following conditions:

$$M_n^{\mathrm{T}} \left\{ \sum_{t=1}^n \frac{x_{nt} x_{nt}^{\mathrm{T}} r e^{x_{nt}^{\mathrm{T}}\beta_0}}{\left(1 + e^{x_{nt}^{\mathrm{T}}\beta_0}\right)^2} \right\} M_n \to \Omega_{11}, \qquad (4)$$

$$M_n^{\mathrm{T}} \left\{ \sum_{t=1}^n \frac{x_{nt} x_{nt}^{\mathrm{T}} r e^{2x_{nt}^{\mathrm{T}}\beta_0}}{\left(1 + e^{x_{nt}^{\mathrm{T}}\beta_0}\right)^2} \right\} M_n \to \Omega_{12}, \qquad (5)$$

and

$$M_n^{\mathrm{T}} \left\{ \sum_{t=1}^n \frac{x_{nt} x_{n,t+k}^{\mathrm{T}} r^2 e^{\left(x_{nt}^{\mathrm{T}} + x_{n,t+k}^{\mathrm{T}}\right)\beta_0}}{\left(1 + e^{x_{nt}^{\mathrm{T}}\beta_0}\right)\left(1 + e^{x_{n,t+k}^{\mathrm{T}}\beta_0}\right)} \right\} M_n \to W_k \qquad (6)$$

uniformly in $|k| < n$ as $n \to \infty$. Furthermore,

$$M_n^{\mathrm{T}} \left\{ \gamma_\epsilon(k) \sum_{t=1}^{-k} \frac{x_{nt} x_{n,t+k}^{\mathrm{T}} e^{\left(x_{nt}^{\mathrm{T}} + x_{n,t+k}^{\mathrm{T}}\right)\beta_0}}{\left(1 + e^{x_{nt}^{\mathrm{T}}\beta_0}\right)\left(1 + e^{x_{n,t+k}^{\mathrm{T}}\beta_0}\right)} \right\} M_n \to 0 \qquad (7)$$

for each $k < 0$ and the left-hand side is uniformly bounded in $k \in (-n, 0)$ as $n \to \infty$; and similarly,

$$M_n^{\mathrm{T}} \left\{ \gamma_\epsilon(k) \sum_{t=n-k+1}^{n} \frac{x_{nt} x_{n,t+k}^{\mathrm{T}} e^{\left(x_{nt}^{\mathrm{T}} + x_{n,t+k}^{\mathrm{T}}\right)\beta_0}}{\left(1 + e^{x_{nt}^{\mathrm{T}}\beta_0}\right)\left(1 + e^{x_{n,t+k}^{\mathrm{T}}\beta_0}\right)} \right\} M_n \to 0 \qquad (8)$$

for each $k > 0$ and the left-hand side is again uniformly bounded in $k \in (0, n)$ as $n \to \infty$.

Remark 1. Conditions (4)–(8) guarantee that the asymptotic variance of $\hat{\beta}_n$ is well defined. We require in the proof of Theorem 1 the existence of

$$\lim_{n \to \infty} M_n^{\mathrm{T}} \left\{ \sum_{t=1}^n \sum_{s=1}^n \frac{x_{nt} x_{ns}^{\mathrm{T}} r^2 e^{\left(x_{nt}^{\mathrm{T}} + x_{ns}^{\mathrm{T}}\right)\beta_0}}{\left(1 + e^{x_{nt}^{\mathrm{T}}\beta_0}\right)\left(1 + e^{x_{ns}^{\mathrm{T}}\beta_0}\right)} \gamma_\epsilon(s - t) \right\} M_n.$$

Making a change of variables, $k = s - t$, we have that

$$M_n^{\mathrm{T}} \left\{ \sum_{t=1}^{n} \sum_{s=1}^{n} \frac{x_{nt} x_{ns}^{\mathrm{T}} r^2 e^{\left(x_{nt}^{\mathrm{T}} + x_{ns}^{\mathrm{T}}\right)\beta_0}}{\left(1 + e^{x_{nt}^{\mathrm{T}}\beta_0}\right)\left(1 + e^{x_{ns}^{\mathrm{T}}\beta_0}\right)} \gamma_\epsilon(s-t) \right\} M_n$$

$$= M_n^{\mathrm{T}} \left\{ \sum_{k=-n+1}^{n-1} \gamma_\epsilon(k) \sum_{t=1}^{n} \frac{x_{nt} x_{n,t+k}^{\mathrm{T}} r^2 e^{\left(x_{nt}^{\mathrm{T}} + x_{n,t+k}^{\mathrm{T}}\right)\beta_0}}{\left(1 + e^{x_{nt}^{\mathrm{T}}\beta_0}\right)\left(1 + e^{x_{n,t+k}^{\mathrm{T}}\beta_0}\right)} \right\} M_n$$

$$- M_n^{\mathrm{T}} \left\{ \sum_{k=-n+1}^{-1} \gamma_\epsilon(k) \sum_{t=1}^{-k} \frac{x_{nt} x_{n,t+k}^{\mathrm{T}} r^2 e^{\left(x_{nt}^{\mathrm{T}} + x_{n,t+k}^{\mathrm{T}}\right)\beta_0}}{\left(1 + e^{x_{nt}^{\mathrm{T}}\beta_0}\right)\left(1 + e^{x_{n,t+k}^{\mathrm{T}}\beta_0}\right)} \right\} M_n$$

$$- M_n^{\mathrm{T}} \left\{ \sum_{k=1}^{n-1} \gamma_\epsilon(k) \sum_{t=n-k+1}^{n} \frac{x_{nt} x_{n,t+k}^{\mathrm{T}} r^2 e^{\left(x_{nt}^{\mathrm{T}} + x_{n,t+k}^{\mathrm{T}}\right)\beta_0}}{\left(1 + e^{x_{nt}^{\mathrm{T}}\beta_0}\right)\left(1 + e^{x_{n,t+k}^{\mathrm{T}}\beta_0}\right)} \right\} M_n.$$

By the dominated convergence theorem, it follows from conditions (7) and (8) that the last two terms on the right-hand side go to zero, respectively, while it follows from condition (6) that the first term converges to $\sum_{k=-\infty}^{\infty} W_k \gamma_\epsilon(k)$.

Conditions (4)–(8) hold for a wide range of covariates; for example, a trend function $x_{nt} = f(t/n)$, where $f(\cdot)$ is a continuous vector-valued function on [0, 1]. In this case, take $M_n = n^{-1/2} I_l$, where I_l is an $l \times l$ identity matrix. Then, approximating sums by integrals, we obtain

$$\Omega_{11} = \int_0^1 \frac{f(x) f^{\mathrm{T}}(x) r e^{f^{\mathrm{T}}(x)\beta_0}}{\left(1 + e^{f^{\mathrm{T}}(x)\beta_0}\right)^2} dx, \qquad \Omega_{12} = \int_0^1 \frac{f(x) f^{\mathrm{T}}(x) r e^{2 f^{\mathrm{T}}(x)\beta_0}}{\left(1 + e^{f^{\mathrm{T}}(x)\beta_0}\right)^2} dx,$$

$$W_k = \int_0^1 \frac{f(x) f^{\mathrm{T}}(x) r e^{2 f^{\mathrm{T}}(x)\beta_0}}{\left(1 + e^{f^{\mathrm{T}}(x)\beta_0}\right)^2} dx \qquad (k = 0, \pm 1, \ldots).$$

Other functions satisfying these conditions include harmonic functions that specify seasonal effects and stationary processes; see Davis et al. (2000).

Remark 2. We also give two examples of random covariates that satisfy conditions (7) and (8); here convergence and uniform boundedness are interpreted as in probability.

1. Suppose $x_{nt} = x_t$ is a bounded sequence of random variables. The convergence to zero of the terms in (7) and (8) is immediate. We check the boundedness condition for (7) only; verifying the other is similar. Take $M_n = n^{-1/2}$, then

$$\sup_{-n<k<0} \left| \frac{\gamma_\epsilon(k)}{n} \sum_{t=1}^{-k} \frac{x_t x_{t+k} e^{(x_t + x_{t+k})\beta_0}}{\left(1 + e^{x_t\beta_0}\right)\left(1 + e^{x_{t+k}\beta_0}\right)} \right| \leqslant \frac{\gamma_\epsilon(0)}{n} \sup_{-n<k<0} \left(\sum_{t=1}^{-k} |x_t x_{t+k}| \right) \leqslant C.$$

2. Suppose $x_{nt} = x_t$ is a stationary time series, and $\gamma_\epsilon(k)$ is absolutely summable. As in the previous example, we only check the uniform boundedness part of (7). With $M_n = n^{-1/2}$,

$$\operatorname{pr}\left\{\sup_{-n<k<0}\left|\frac{\gamma_\epsilon(k)}{n}\sum_{t=1}^{-k}\frac{x_t x_{t+k}e^{(x_t+x_{t+k})\beta_0}}{(1+e^{x_t\beta_0})(1+e^{x_{t+k}\beta_0})}\right|>A\right\}$$
$$\leqslant \sum_{-n<k<0}\operatorname{pr}\left\{|\gamma_\epsilon(k)|\sum_{t=1}^{-k}|x_t x_{t+k}|>nA\right\}\leqslant\frac{1}{nA}\sum_{-n<k<0}|\gamma_\epsilon(k)|\sum_{t=1}^{-k}E|x_t x_{t+k}|$$
$$\leqslant\frac{C}{A}\sum_{-n<k<0}|\gamma_\epsilon(k)|$$

provided that $\operatorname{var}(|x_t|)$ exists and $\sum_{k=-\infty}^{\infty}|\gamma_\epsilon(k)|<\infty$. So, for any $\delta>0$, there exists a large A such that the left-hand-side probability is less than δ for all n. That is, the uniform boundedness condition holds.

THEOREM 1. *Let $\hat\beta_n$ be the generalized linear model estimator of the parameter β based on the observations $Y_1,\ldots,Y_n$ coming from the model* (1)–(2). *Assume that the latent process $\{\epsilon_t\}$ is stationary, strongly mixing and satisfies Assumptions* 1 *and* 2. *Also, assume that the covariates x_{nt} satisfy* (4)–(8), *and* $\sup_{1\leqslant t\leqslant n}|M_n^{\mathrm T}x_{nt}|=O(n^{-1/2})$. *Then, $\hat\beta_n\to\beta_0$ in probability and*

$$M_n^{-1}(\hat\beta_n-\beta_0)\to N\left(0,\Omega_1^{-1}+\Omega_1^{-1}\Omega_2\Omega_1^{-1}\right)$$

in distribution as $n\to\infty$, where $\Omega_1=\Omega_{11}+\Omega_{12}$ and $\Omega_2=\Omega_{12}\gamma_\epsilon(0)+\sum_{k=-\infty}^{\infty}W_k\gamma_\epsilon(k)$.

Remark 3. Note that Ω_1^{-1} is the asymptotic covariance matrix from a standard generalized linear model analysis when there is no latent process in the true model, and $\Omega_1^{-1}\Omega_2\Omega_1^{-1}$ is the additional contribution to the asymptotic covariance caused by the presence of the latent process.

Remark 4. In establishing the limiting distribution of the generalized linear model estimator, the pseudo-loglikelihood is decomposed into two pieces: a deterministic quadratic term and a random linear term corresponding to the score function. The quadratic term has the same behaviour as without a latent process. While the random term has asymptotic zero-mean, which gives the consistency of the estimator, its variance requires an adjustment due to the impact of the latent process.

2·3. *Asymptotic properties when $\{\alpha_t\}$ is a Gaussian linear process*

Theorem 1 can be adapted to the case when $\{\alpha_t\}$ is a stationary Gaussian linear process. In this case, in order to satisfy the identifiability condition of $E(e^{\alpha_t})=1$, it is required that $\alpha_t\sim N(-\sigma_\alpha^2/2,\sigma_\alpha^2)$. Moreover, the relationship between the autocovariance functions of $\{\epsilon_t\}$ and $\{\alpha_t\}$ is explicitly given by $\gamma_\epsilon(k)=\exp(\gamma_\alpha(k))-1$ for all k.

The corresponding asymptotic results for $\hat\beta_n$ are given in the following theorem.

THEOREM 2. *Suppose $\{\alpha_t\}$ is a stationary Gaussian linear process, and $\sum_{k=0}^{\infty}|\gamma_\epsilon(k)|<\infty$, where $\{\epsilon_t=e^{\alpha_t}\}$. If the conditions in Theorem* 1 *on the covariates x_{nt} are met, then the asymptotic results in Theorem* 1 *hold for the generalized linear model estimator $\hat\beta_n$.*

3. EXTENSION TO THE ONE-PARAMETER EXPONENTIAL FAMILY

3·1. *Asymptotic properties of generalized linear model estimators*

In this section, we study time series regression models in a more general set-up. Let $Y_1,\ldots,Y_n$ denote the observed time series that are independent conditional on a latent process $\{\alpha_t\}$. The conditional distribution, depending only on α_t, belongs to the one-parameter exponential family and is given by

$$p(Y_t \mid \alpha_t) = \exp\{\theta_t Y_t - b(\theta_t) + c(Y_t)\}, \tag{9}$$

where $\theta_t = g(x_{nt}^{\mathrm{T}}\beta + \alpha_t)$ for a real function g that is uniquely determined by the chosen link function f. To see this, assume f is monotone and differentiable, and let $\eta_t = x_{nt}^{\mathrm{T}}\beta + \alpha_t$. Then

$$E(Y_t \mid \alpha_t) = h(\eta_t) = h\left(x_{nt}^{\mathrm{T}}\beta + \alpha_t\right), \tag{10}$$

where h is the inverse function of f. On the other hand, it is well known that $E(Y_t \mid \alpha_t) = b'(\theta_t) = (b' \circ g)(x_{nt}^{\mathrm{T}}\beta + \alpha_t)$. Therefore, $(b' \circ g)(\cdot) = h(\cdot)$.

Suppose $Y_1, \ldots, Y_n$ are observations from the true model with the parameter β_0. The generalized linear model estimator $\hat{\beta}_n$ of β is defined as the maximizer of the pseudo-loglikelihood function

$$\begin{aligned} \ell_n(\beta) &= \sum_{t=1}^{n} \{g(x_{nt}^{\mathrm{T}}\beta)Y_t - (b \circ g)(x_{nt}^{\mathrm{T}}\beta) + c(Y_t)\} \\ &= -\sum_{t=1}^{n}(b \circ g)(x_{nt}^{\mathrm{T}}\beta) + \sum_{t=1}^{n} Y_t g(x_{nt}^{\mathrm{T}}\beta) + \sum_{t=1}^{n} c(Y_t), \end{aligned} \tag{11}$$

which ignores the latent process $\{\alpha_t\}$ in the model.

In order to establish the asymptotic properties of $\hat{\beta}_n$, we take a link function f such that the loglikelihood function (11) is concave and $E\{h(x_{nt}^{\mathrm{T}}\beta + \alpha_t)\} = h(x_{nt}^{\mathrm{T}}\beta)$. Such a link function exists for Poisson, negative binomial, Gaussian, among other cases; see examples in § 3·2. Then, $\mu_t = E(Y_t) = h(x_{nt}^{\mathrm{T}}\beta_0)$. Moreover, the variance of Y_t is given by

$$\operatorname{var}(Y_t) = E\{(b'' \circ g)(x_{nt}^{\mathrm{T}}\beta_0 + \alpha_t)\} + \operatorname{var}\{h(x_{nt}^{\mathrm{T}}\beta_0 + \alpha_t)\}$$

since $\operatorname{var}(Y_t \mid \alpha_t) = b''(\theta_t) = (b'' \circ g)(x_{nt}^{\mathrm{T}}\beta_0 + \alpha_t)$ and the autocovariance function is

$$\operatorname{cov}(Y_{t+k}, Y_t) = \operatorname{cov}\{h(x_{n,t+k}^{\mathrm{T}}\beta_0 + \alpha_{t+k}), h(x_{nt}^{\mathrm{T}}\beta_0 + \alpha_t)\} \quad (k \neq 0).$$

Suppose there exists a sequence of nonsingular matrices M_n such that the covariates x_{nt} satisfy the conditions $\sup_{1\leqslant t\leqslant n} |M_n^{\mathrm{T}} x_{nt}| = O(n^{-1/2})$ and

$$\begin{aligned} &M_n^{\mathrm{T}} \sum_{t=1}^{n} x_{nt}x_{nt}^{\mathrm{T}}(b'' \circ g)(x_{nt}^{\mathrm{T}}\beta_0)\{g'(x_{nt}^{\mathrm{T}}\beta_0)\}^2 M_n \to \Omega_1, \\ &M_n^{\mathrm{T}} \sum_{t=1}^{n} x_{nt}x_{nt}^{\mathrm{T}}\mathrm{E}\{(b'' \circ g)(x_{nt}^{\mathrm{T}}\beta_0 + \alpha_t)\}\{g'(x_{nt}^{\mathrm{T}}\beta_0)\}^2 M_n \to \Omega_1^{\dagger}, \\ &M_n^{\mathrm{T}} \sum_{j,t=1}^{n} x_{nj}x_{nt}^{\mathrm{T}}\operatorname{cov}\{h(x_{nj}^{\mathrm{T}}\beta_0 + \alpha_j), h(x_{nt}^{\mathrm{T}}\beta_0 + \alpha_t)\}g'(x_{nj}^{\mathrm{T}}\beta_0)g'(x_{nt}^{\mathrm{T}}\beta_0)M_n \to \Omega_2^{\dagger}. \end{aligned}$$

The asymptotic results for $\hat{\beta}_n$ are stated in Theorem 3, whose proof follows similar lines to the proof of Theorem 1 in the Appendix and hence is omitted.

THEOREM 3. *Let $\hat{\beta}_n$ be the generalized linear model estimator of β for the parameter-driven model* (9)–(10). *If*

$$C_n(s) = \sum_{t=1}^{n} s^{\mathrm{T}} M_n^{\mathrm{T}} x_{nt} g'(x_{nt}^{\mathrm{T}}\beta_0)\{h(x_{nt}^{\mathrm{T}}\beta_0 + \alpha_t) - h(x_{nt}^{\mathrm{T}}\beta_0)\} \to V$$

in distribution, where $V \sim N(0, s^{\mathrm{T}}\Omega_2^{\dagger}s)$*, then under suitable assumptions* $\hat{\beta}_n \to \beta_0$ *in probability, and*

$$M_n^{-1}(\hat{\beta}_n - \beta_0) \to N\{0, \Omega_1^{-1}(\Omega_1^{\dagger} + \Omega_2^{\dagger})\Omega_1^{-1}\}$$

in distribution.

Remark 5. When there is no latent process in the model, $\Omega_1^{\dagger}$ reduces to Ω_1, and Ω_1^{-1} is the asymptotic covariance matrix of $\hat{\beta}_n$ from a standard generalized linear model estimation.

3·2. *Examples*

Example 1. *Poisson case.* Suppose the random variables $Y_1, \ldots, Y_n$ are modelled by $Y_t \mid \alpha_t \sim \mathrm{Po}(\lambda_t)$ and $\log \lambda_t = x_{nt}^{\mathrm{T}}\beta + \alpha_t$, where $\{\alpha_t\}$ is a stationary latent process. Writing the conditional density function in the canonical form (9), we have $\theta_t = \log \lambda_t$ and $b(\theta_t) = e^{\theta_t}$. If we take a canonical link, namely $f(z) = \log(z)$, then its inverse function $h(z) = \exp(z)$ and the function $g(z) = z$.

In the Poisson set-up, $E(Y_t \mid \alpha_t) = \lambda_t = \exp(x_{nt}^{\mathrm{T}}\beta + \alpha_t)$. With ϵ_t defined as e^{α_t} and the condition $E(\epsilon_t) = 1$, we have $E(Y_t) = \exp(x_{nt}^{\mathrm{T}}\beta)$. It is easy to verify that $\Omega_1^{\dagger} = \Omega_1$. The asymptotic results for $\hat{\beta}_n$ reduce to Theorem 1 of Davis et al. (2000).

Example 2. *Negative binomial case*. In the negative binomial case, the model is specified by $Y_t \mid \alpha_t \sim \mathrm{NegBin}(r, p_t)$, where $-\log\{p_t/(1-p_t)\} = x_{nt}^{\mathrm{T}}\beta + \alpha_t$ with some known positive number r. Writing the conditional density function in the form (9), we obtain $\theta_t = \log(1-p_t)$ and $b(\theta_t) = -r\log(1-e^{\theta_t})$. If we take a link such that its inverse function $h(z) = r\exp(-z)$, then $g(z) = -\log(1+e^z)$.

This is the set-up in § 2, where $E(Y_t \mid \alpha_t) = r\exp(x_{nt}^{\mathrm{T}}\beta)\epsilon_t$ with $\epsilon_t = \exp(\alpha_t)$. Therefore, $E(Y_t) = r\exp(x_{nt}^{\mathrm{T}}\beta)$ under the condition $E(\epsilon_t) = 1$. Further derivations show that the asymptotic results of $\hat{\beta}_n$ reduce to Theorem 1 in § 2·2, where $\Omega_1^{\dagger}$ and $\Omega_2^{\dagger}$ correspond to $\Omega_1 + \Omega_{12}\gamma_{\epsilon}(0)$ and $\Omega_2 - \Omega_{12}\gamma_{\epsilon}(0)$, respectively.

Example 3. *Gaussian case*. Consider the model $Y_t = x_{nt}^{\mathrm{T}}\beta + \alpha_t + Z_t$, where the noise is described by a stationary process plus an independent Gaussian observation error with known variance τ^2. That is, conditional on the latent process $\{\alpha_t\}$, $Y_1, \ldots, Y_n$ are independent normal; and the conditional distribution is specified by $Y_t \mid \alpha_t \sim N(\nu_t, \tau^2)$, where $\nu_t = x_{nt}^{\mathrm{T}}\beta + \alpha_t$. Writing the conditional density function in the form (9), we obtain

$$p(Y_t \mid \alpha_t) = \exp\left[Y_t\left(\frac{\nu_t}{\tau^2}\right) - \frac{\nu_t^2}{2\tau^2} - \frac{1}{2}\left\{\frac{Y_t^2}{\tau^2} + \log(2\pi\tau^2)\right\}\right].$$

Then, $\theta_t = \nu_t/\tau^2$ and $b(\theta_t) = \theta_t^2\tau^2/2$. It follows that $b'(\theta_t) = \tau^2\theta_t$ and $b''(\theta_t) = \tau^2$. If we take a link such that its inverse function $h(z) = z$, then $g(z) = z/\tau^2$.

In this Gaussian case, $E(Y_t \mid \alpha_t) = x_{nt}^{\mathrm{T}}\beta + \alpha_t$, which yields an additive model instead of a multiplicative model as in the Poisson and negative binomial set-ups. It is straightforward to check that $\Omega_1^{\dagger} = \Omega_1$. Then $M_n^{-1}(\hat{\beta}_n - \beta_0) \to N(0, \Omega_1^{-1} + \Omega_1^{-1}\Omega_2^{\dagger}\Omega_1^{-1})$ in distribution.

4. NUMERICAL STUDY

A simulation study was conducted to evaluate the finite sample performance of the generalized linear model estimator. We considered two experiments: one with a negative binomial and the other

Table 1. *Simulation results for generalized linear model estimates: empirical means and standard deviations*

Experiment 1

ϕ	0·5				0·8			
n	500		1000		500		1000	
	$\hat{\beta}_2$	$\hat{\beta}_4$	$\hat{\beta}_2$	$\hat{\beta}_4$	$\hat{\beta}_2$	$\hat{\beta}_4$	$\hat{\beta}_2$	$\hat{\beta}_4$
Mean	0·305	0·701	0·295	0·701	0·256	0·699	0·292	0·698
SD	0·342	0·109	0·253	0·078	0·542	0·087	0·381	0·064
ASD	0·358	0·108	0·253	0·077	0·555	0·092	0·394	0·065

Experiment 2

ϕ	0·1			0·5		
n	100	200	500	100	200	500
Mean	0·718	0·709	0·705	0·744	0·723	0·708
SD	0·205	0·138	0·082	0·344	0·234	0·141
ASD	0·189	0·134	0·085	0·316	0·223	0·141

SD, standard deviation; ASD, asymptotic standard deviation.

with a Bernoulli density for the conditional distribution of observations given the latent process. For each case, we simulated 1000 replications and estimated the parameters of interest, reporting the empirical means and standard deviations of the estimates together with the asymptotic standard deviation.

Experiment 1: *Negative binomial.* Suppose that the random variables $Y_1, \ldots, Y_n$ are independent, conditional on the stationary latent process $\{\alpha_t\}$ and the covariate x_{nt}. The conditional distribution is specified by $Y_t \mid \alpha_t, x_{nt} \sim \text{NegBin}(4, p_t)$, where $\log\{p_t/(1-p_t)\} = x_{nt}^{\mathrm{T}}\beta + \alpha_t$.

We used a covariate sequence defined by $x_{nt} = \{1, t/n, \cos(2\pi t/6), \sin(2\pi t/6)\}^{\mathrm{T}}$, which includes a standardized trend and two harmonic function components. The value of the true parameter vector β_0 was taken to be $(0{\cdot}1, 0{\cdot}3, 0{\cdot}5, 0{\cdot}7)^{\mathrm{T}}$. The latent process is specified by an AR(1) model $\alpha_t = \phi\alpha_{t-1} + Z_t$, where Z_t are independent identically distributed $N(0, \sigma^2)$ variables, where the value of ϕ was taken to be 0·5 and 0·8, and σ was chosen such that $\text{var}(\alpha_t) = 1$. Results are based on samples of size 500 and 1000, respectively.

Searching for the maximizer of the loglikelihood function is implemented using R (R Development Core Team, 2008). A summary of the simulation results is given in Table 1; here we only report the results for coefficients of the linear trend and one trigonometric term. In all cases, the empirical standard deviation is rather close to the asymptotic standard deviation. The generalized linear model estimates are approximately unbiased except for the case where $\phi = 0{\cdot}8$ and $n = 500$, which is due to slow convergence to the limit distribution when ignoring a more strongly autocorrelated latent process. In all cases, normal probability plots, not included, support the asymptotic normality of the β estimates.

Experiment 2: *Bernoulli.* In this case the conditional distribution of $Y_1, \ldots, Y_n$ given the latent process and covariate is specified by $\text{Ber}(p_t)$, where $-\log p_t = x_{nt}^{\mathrm{T}}\beta + \alpha_t$ is required to be nonnegative. For simplicity, we took a constant covariate $x_{nt} = 1$ and set $\beta_0 = 0{\cdot}7$. The latent process is specified by an AR(1) model $\alpha_t = \phi\alpha_{t-1} + Z_t$, where we used a sequence of independent and exponentially distributed random variables with parameter $\lambda = 1$ as the innovation process $\{Z_t\}$, and took ϕ to be 0·1 and 0·5. We considered sample sizes of 100, 200 and 500.

In order that the condition $E(\epsilon_t) = 1$ holds, we define $\epsilon_t = e^{-\alpha_t}/E(e^{-\alpha_t})$, where $E(e^{-\alpha_t})$ is obtained from the formula $\prod_{j=0}^{\infty} 1/(1 + \phi^j/\lambda)$. Then,

$$n^{-1/2}(\hat{\beta}_n - \beta_0) \to N\{0, (\Omega_1^{\dagger} + \Omega_2^{\dagger})/\Omega_1^2\}$$

in distribution, where

$$\Omega_1 = \frac{1}{e^{\beta_0} - 1}, \quad \Omega_1^{\dagger} = \frac{1}{(e^{\beta_0} - 1)^2}\{e^{\beta_0} - 1 - \gamma_\epsilon(0)\}, \quad \Omega_2^{\dagger} = \frac{1}{(e^{\beta_0} - 1)^2}\sum_{k=-\infty}^{\infty} \gamma_\epsilon(k).$$

Simulation results are reported in Table 1. In all cases, the empirical standard deviation is close to the asymptotic standard deviation. But there is some bias in the generalized linear model estimates when $\phi = 0{\cdot}5$. Further studies show that the generalized linear model estimation does not perform well when $\phi > 0{\cdot}6$. This is due to the fact that when the latent process becomes large, the probabilities $p_t = e^{-(\beta_0+\alpha_t)}$ become extremely small so that most of the simulated values of $\{Y_t\}$ are zero. This, in turn, makes estimation virtually impossible.

5. APPLICATION TO REAL DATA

We applied the results of § 2 to the polio data studied by Zeger (1988). The data consist of monthly counts of poliomyelitis cases in the USA from the year 1970 to 1983 as reported by the Centers for Disease Control. The time series nature of the data is well indicated by the autocorrelation function plot. Davis et al. (2000) modelled the counts by the Poisson distribution given a latent process. We instead used a negative binomial conditional distribution as described in § 2. We specified the same covariates as in Davis et al. (2000), namely

$$x_t = \{1, t'/1000, \cos(2\pi t'/12), \sin(2\pi t'/12), \cos(2\pi t'/6), \cos(2\pi t'/6)\}^{\mathrm{T}},$$

where $t' = t - 73$ is used to locate the intercept term at January 1976. In order to compare our results with those of Davis et al. (2000), we used the link function

$$\log\left\{\frac{r(1-p_t)}{p_t}\right\} = x_t^{\mathrm{T}}\beta + \alpha_t,$$

where $\{\alpha_t\}$ was assumed to be an AR(1) latent process with Gaussian innovations. This specification is slightly different than the one used in § 2, with which we obtain

$$\mu_t = E\left\{\frac{r(1-p_t)}{p_t}\right\} = E\left(e^{x_t^{\mathrm{T}}\beta+\alpha_t}\right) = e^{x_t^{\mathrm{T}}\beta}$$

under the condition $E(e^{\alpha_t}) = 1$. Then, the regression coefficients have the same interpretation as those in Davis et al. (2000).

The data were fitted by a standard negative binomial generalized linear model. For parameter estimation we adopted the approach of Benjamin et al. (2003), maximizing the likelihood $\mathcal{L}(\beta, r)$ with respect to β for different r values. The estimate $\hat{r} = 2$ was determined by the r value that yielded the smallest AIC. The corresponding estimate $\hat{\beta}_{\mathrm{NB}}$ of β and its standard error are reported in Table 2, columns 5–6. The AIC value for this model fit is 519·857. For comparison, the results from a standard Poisson generalized linear model fit of Davis et al. (2000) are also included in Table 2, columns 2–3; the AIC value for their model fit is 557·898. Based on AIC, the standard negative binomial generalized linear model fit is better.

The estimates and standard errors in the Poisson and negative binomial cases are comparable. These standard error calculations ignore the possibility of the presence of a latent process. We examined the Pearson residuals for the existence of a latent process and a suitable model if such a process exists. Since the serial dependence among observations is ignored when fitting a standard generalized linear model, we would expect that Pearson residuals display the same dependence structure as that of the polio counts. In both cases, the partial autocorrelation function plots of Pearson residuals support the assumption of an AR(1) latent process.

Table 2. *Estimates and their standard errors from analyses of polio data by standard negative binomial and Poisson generalized linear models*

	Poisson			Negative binomial			Simulations	
Covariate	$\hat\beta_{\text{Po}}$	SE	ASE	$\hat\beta_{\text{NB}}$	SE	ASE	Mean	SD
Intercept	0·207	0·075	0·205	0·209	0·100	0·167	0·162	0·173
Trend	−4·799	1·399	4·115	−4·354	1·970	3·311	−4·381	3·190
$\cos(2\pi t'/12)$	−0·149	0·097	0·157	−0·143	0·134	0·156	−0·153	0·158
$\sin(2\pi t'/12)$	−0·532	0·109	0·168	−0·504	0·144	0·165	−0·512	0·163
$\cos(2\pi t'/6)$	0·169	0·098	0·122	0·168	0·136	0·144	0·179	0·149
$\sin(2\pi t'/6)$	−0·432	0·101	0·125	−0·422	0·138	0·146	−0·424	0·145

ASE, asymptotic standard error; SD, standard deviation; SE, standard error.

Suppose $\alpha_t = \phi\alpha_{t-1} + Z_t$, where Z_t are independent identically distributed $N(0, \sigma^2)$ variables. The asymptotic standard errors in the seventh column of Table 2 were obtained using Theorem 2. To be specific, we first estimated ϕ and σ^2 by the method of moments. Recall that

$$\text{var}(Y_t) = \mu_t + \mu_t^2 \frac{(r+1)\sigma_\epsilon^2 + 1}{r},$$

where $\epsilon_t = e^{\alpha_t}$. So, $\hat\sigma_\epsilon^2 = 0{\cdot}3586$ and $\hat\rho_\epsilon(1) = 0{\cdot}7719$ were obtained using an ordinary least squares type of estimators suggested by Brännäs & Johansson (1994):

$$\hat\sigma_\epsilon^2 = \frac{1}{\hat r + 1}\left[\frac{\hat r \sum_{t=1}^n \hat\mu_t^2\{(Y_t - \hat\mu_t)^2 - \hat\mu_t\}}{\sum_{t=1}^n \hat\mu_t^4} - 1\right],$$

$$\hat\rho_\epsilon(1) = \frac{\hat\sigma_\epsilon^{-2}\sum_{t=2}^n \hat\mu_t\hat\mu_{t-1}(Y_t - \hat\mu_t)(Y_{t-1} - \hat\mu_{t-1})}{\sum_{t=2}^n \hat\mu_t^2\hat\mu_{t-1}^2}.$$

Then $\hat\sigma_\alpha^2 = 0{\cdot}3065$ and $\hat\rho_\alpha(1) = 0{\cdot}7973$ were obtained through the identity $\gamma_\epsilon(k) = e^{\gamma_\alpha(k)} - 1$. It followed that $\hat\phi = 0{\cdot}7973$ and $\hat\sigma^2 = 0{\cdot}1117$. Moreover, $\hat\gamma_\alpha(k)$ and $\hat\gamma_\epsilon(k)$ were readily computed. With these estimates, we approximated the asymptotic standard errors of $\hat\beta_{\text{NB}}$ using the formula

$$\text{var}(\hat\beta_{\text{NB}}) = (\hat\Omega_{11,n} + \hat\Omega_{12,n})^{-1}(\hat\Omega_{11,n} + \hat\Omega_{12,n} + \hat\Omega_{12,n}\hat\sigma_\epsilon^2 + \hat W_n)(\hat\Omega_{11,n} + \hat\Omega_{12,n})^{-1},$$

where

$$\hat\Omega_{11,n} = \sum_{t=1}^n \frac{x_t x_t^{\mathrm{T}}\hat\mu_t}{(1 + \hat\mu_t/\hat r)^2}, \quad \hat\Omega_{12,n} = \frac{1}{\hat r}\sum_{t=1}^n \frac{x_t x_t^{\mathrm{T}}\hat\mu_t^2}{(1 + \hat\mu_t/\hat r)^2},$$

$$\hat W_n = \sum_{t=1}^n\sum_{s=1}^n \frac{x_t x_s^{\mathrm{T}}\hat\mu_t\hat\mu_s}{(1 + \hat\mu_t/\hat r)(1 + \hat\mu_s/\hat r)}\hat\gamma_\epsilon(t-s),$$

with $\hat\mu_t = \exp(x_t^T\hat\beta_{\text{NB}})$.

One of the main objectives in modelling the polio data is to investigate whether or not the incidence of polio has been decreasing since 1970. The results showed that the negative trend was not significant using the standard error that includes a latent process. A false significance would be obtained if using the standard error of 1·970 produced by the standard negative binomial generalized linear model estimation. This is in agreement with the conclusion of Davis et al. (2000).

To further check the generalized linear model estimates and asymptotic standard errors, we simulated 1000 replications of a time series of length 168 using $\hat r$ and $\hat\beta_{\text{NB}}$ as true parameter

Table 3. *Estimates and their standard errors from analyses of polio data by parameter-driven negative binomial and Poisson generalized linear models*

	Poisson		Negative binomial	
Covariate	$\hat{\beta}_{\text{Po}}$	SE	$\hat{\beta}_{\text{NB}}$	SE
Intercept	0·090	0·141	0·106	0·177
Trend	−3·600	2·751	−3·467	3·375
$\cos(2\pi t'/12)$	−0·098	0·143	−0·109	0·129
$\sin(2\pi t'/12)$	−0·478	0·154	−0·488	0·140
$\cos(2\pi t'/6)$	0·190	0·121	0·182	0·122
$\sin(2\pi t'/6)$	−0·355	0·122	−0·365	0·123

SE, standard error.

values. The latent process $\{\alpha_t\}$ was assumed to be a Gaussian AR(1) with $\phi = 0{\cdot}7973$ and marginal distribution $N(-\sigma_\alpha^2/2, \sigma_\alpha^2)$, where $\sigma_\alpha^2 = 0{\cdot}3065$. The empirical means and standard deviations of estimates from simulations are reported in the last two columns of Table 2. The estimates were approximately unbiased, except for the intercept term where the empirical mean of 0·162 was significantly different from 0·209 used for simulating data. The empirical standard deviations were in good agreement with the asymptotic standard errors.

We also fitted the polio data by parameter-driven negative binomial and Poisson generalized linear models using the procedure GLIMMIX in SAS, assuming a Gaussian AR(1) latent process. The estimates of coefficients and their standard errors are reported in Table 3, in the negative binomial case $\hat{r} = 4{\cdot}146$. Because the AIC generated by GLIMMIX is based on pseudolikelihood estimation, it is not useful for comparing models. Instead, we performed model diagnostics via Pearson residuals. Since now the serial dependence among the polio counts had been taken into account, we should expect Pearson residuals to be independent. The autocorrelation function plot showed that the Pearson residuals in the negative binomial case appeared independent. In addition, a p-value of 0·14550 from the Ljung–Box test of randomness implied no evidence against independence, whereas in the Poisson case the p-value 0·00097 indicated strong evidence to reject independence of the Pearson residuals. Therefore, the parameter-driven negative binomial generalized linear model appears more appropriate for the polio data.

We further simulated time series of length 168 from the Poisson model using the estimates from Davis et al. (2000) as true parameter values and fitted the simulated data by both parameter-driven negative binomial and Poisson generalized linear models. It turned out that both model-fittings yielded the same estimates for β. The very large value of $\hat{r}$ in the negative binomial case indicated that the model was actually Poisson. On the contrary, when we simulated time series from the negative binomial model using $\hat{\beta}_{\text{NB}}$ as true parameter values, the negative binomial model-fitting distinguished itself from the Poisson model-fitting in most cases. This again supported our conclusion that the parameter-driven negative binomial generalized linear model is more appropriate for the polio data.

ACKNOWLEDGEMENT

We thank the Editor, an Associate Editor and two anonymous referees for comments that led to improvement of the paper. The research of R. A. Davis was supported in part by the National Science Foundation.

Appendix

Proof of Theorem 1

We first state a central limit theorem for strongly mixing processes.

Proposition 1 (Davidson, 1992). *Let $\{X_{nt}, t = 1, \ldots, n, n \geqslant 1\}$ denote a triangular array of random variables defined on the probability space $(\Omega, \mathcal{F}, \mathcal{P})$ such that*

1. $E(X_{nt}) = 0$ *and* $E(\sum_{t=1}^{n} X_{nt})^2 = 1$*;*
2. *There exists a positive constant array $\{c_{nt}\}$ and a constant $\gamma > 2$ such that $\{E(|X_{nt}/C_{nt}|^{\gamma})\}^{1/\gamma}$ is uniformly bounded in t and n;*
3. *For each n, the sequence $\{X_{nt}\}$ is strongly mixing with mixing coefficient $\alpha(m)$ such that $\sum_{m=1}^{\infty} \alpha(m)^{\gamma/(\gamma-2)} < \infty$; and*
4. $\sup_n\{n(\max_{1\leqslant t\leqslant n} c_{nt})^2\} < \infty$.

Then, $\sum_{t=1}^{n} X_{nt} \xrightarrow{d} N(0, 1)$ *in distribution.*

Proof of Theorem 1. Let $u = M_n^{-1}(\beta - \beta_0)$, then maximizing $\ell_n(\beta)$ in (3) with respect to β is equivalent to minimizing $g_n(u) = -\ell_n(\beta_0 + M_n u) + \ell_n(\beta_0)$ with respect to u. We write $g_n(u) = g_{n,1}(u) - g_{n,2}(u)$, where

$$g_{n,1}(u) = -r\sum_{t=1}^{n} e^{x_{nt}^{\mathrm{T}}\beta_0} x_{nt}^{\mathrm{T}} M_n u + r\sum_{t=1}^{n}\left(1 + e^{x_{nt}^{\mathrm{T}}\beta_0}\right)\left\{\log\left(1 + e^{x_{nt}^{\mathrm{T}}(\beta_0 + M_n u)}\right) - \log\left(1 + e^{x_{nt}^{\mathrm{T}}\beta_0}\right)\right\},$$

$$g_{n,2}(u) = -\sum_{t=1}^{n}\left(Y_t - re^{x_{nt}^{\mathrm{T}}\beta_0}\right)\left\{\log\left(1 + e^{-x_{nt}^{\mathrm{T}}(\beta_0 + M_n u)}\right) - \log\left(1 + e^{-x_{nt}^{\mathrm{T}}\beta_0}\right)\right\}.$$

For any fixed u, $g_{n,1}(u) \to u^{\mathrm{T}}\Omega_1 u/2$. To show this, we express by applying a Taylor series

$$g_{n,1}(u) = \frac{r}{2}\sum_{t=1}^{n} \frac{e^{x_{nt}^{\mathrm{T}}\beta_0}}{1 + e^{x_{nt}^{\mathrm{T}}\beta_0}}\left(x_{nt}^{\mathrm{T}} M_n u\right)^2 + E_n^1(u),$$

where $E_n^1(u)$ is the remainder in the expansion. It is easily shown that $E_n^1(u) \to 0$. Thus, the result follows from conditions (4)–(5). Moreover, we show that, for any fixed u, $g_{n,2}(u) \to u^{\mathrm{T}} N(0, \Omega_1 + \Omega_2)$ in distribution. We can similarly write $g_{n,2}(u) = u^{\mathrm{T}} U_n - E_n^2(u)$, where

$$U_n = \sum_{t=1}^{n}\left(Y_t - re^{x_{nt}^{\mathrm{T}}\beta_0}\right)\frac{M_n^{\mathrm{T}} x_{nt}}{1 + e^{x_{nt}^{\mathrm{T}}\beta_0}}.$$

As the remainder $E_n^2(u) \to 0$ in probability, it suffices to show

$$U_n \to N(0, \Omega_1 + \Omega_2) \tag{A1}$$

in distribution. For any real vector s, it can be shown that

$$E\left(e^{is^{\mathrm{T}} U_n} \mid \alpha_t\right) \approx \exp(D_n + F_n), \tag{A2}$$

where $A \approx B$ means $A - B \to 0$ in probability,

$$D_n = r\sum_{t=1}^{n} e^{x_{nt}^{\mathrm{T}}\beta_0}\left\{\exp\left(\frac{is^{\mathrm{T}} M_n^{\mathrm{T}} x_{nt}}{1 + e^{x_{nt}^{\mathrm{T}}\beta_0}}\right) - 1 - \frac{is^{\mathrm{T}} M_n^{\mathrm{T}} x_{nt}}{1 + e^{x_{nt}^{\mathrm{T}}\beta_0}}\right\} - \frac{(\gamma_\epsilon(0) + 1)r}{2}\sum_{t=1}^{n}\frac{(s^{\mathrm{T}} M_n^{\mathrm{T}} x_{nt})^2 e^{2x_{nt}^{\mathrm{T}}\beta_0}}{\left(1 + e^{x_{nt}^{\mathrm{T}}\beta_0}\right)^2},$$

$$F_n = r\sum_{t=1}^{n}(\epsilon_t - 1)e^{x_{nt}^{\mathrm{T}}\beta_0}\left\{\exp\left(\frac{is^{\mathrm{T}} M_n^{\mathrm{T}} x_{nt}}{1 + e^{x_{nt}^{\mathrm{T}}\beta_0}}\right) - 1\right\}.$$

It is easy to show

$$D_n \to -\frac{1}{2}s^{\mathrm{T}}\{\Omega_1 + \Omega_{12}\gamma_\epsilon(0)\}s. \tag{A3}$$

Moreover, we can show that $|F_n - iC_n(s)| \to 0$ in probability, where

$$C_n(s) = r\sum_{t=1}^{n} \frac{s^{\mathrm{T}}M_n^{\mathrm{T}}x_{nt}e^{x_{nt}^{\mathrm{T}}\beta_0}}{1+e^{x_{nt}^{\mathrm{T}}\beta_0}}(\epsilon_t - 1).$$

On the other hand, by applying Proposition 1,

$$C_n(s) \to V \tag{A4}$$

in distribution, where $V \sim N(0, s^{\mathrm{T}}Ws)$ with $W = \sum_{k=-\infty}^{\infty} W_k\gamma_\epsilon(k)$. To be specific, let

$$\begin{aligned}\tau_n^2(s) &= \mathrm{var}\{C_n(s)\}\\ &= r^2 s^{\mathrm{T}}M_n^{\mathrm{T}}\sum_{j=1}^{n}\sum_{t=1}^{n}\frac{x_{nj}e^{x_{nj}^{\mathrm{T}}\beta_0}x_{nt}^{\mathrm{T}}e^{x_{nt}^{\mathrm{T}}\beta_0}}{\left(1+e^{x_{nj}^{\mathrm{T}}\beta_0}\right)\left(1+e^{x_{nt}^{\mathrm{T}}\beta_0}\right)}\mathrm{cov}(\epsilon_j, \epsilon_t)M_n s.\end{aligned}$$

Then $\tau_n^2(s) \to s^{\mathrm{T}}Ws$. Defining

$$Z_{nt} = \frac{r}{\tau_n(s)}\cdot\frac{s^{\mathrm{T}}M_n^{\mathrm{T}}x_{nt}e^{x_{nt}^{\mathrm{T}}\beta_0}}{1+e^{x_{nt}^{\mathrm{T}}\beta_0}}(\epsilon_t - 1),$$

we show $\sum_{t=1}^{n} Z_{nt} \to N(0,1)$ in distribution by verifying the conditions in Proposition 1. Firstly, it is clear that

$$E(Z_{nt}) = 0, \quad E\left(\sum_{t=1}^{n} Z_{nt}\right)^2 = E\left\{\frac{C_n(s)}{\tau_n(s)}\right\}^2 = 1.$$

Next, put $c_{nt} = n^{-1/2}$, then $\sup_n\{n(\max_{1\leqslant t\leqslant n} c_{nt})^2\} = 1 < \infty$. For $\lambda > 0$,

$$\begin{aligned}\left\{E\left(\frac{Z_{nt}}{c_{nt}}\right)^{\lambda+2}\right\}^{\frac{1}{\lambda+2}} &= \left[E\left\{\frac{r}{\tau_n(s)}\cdot\frac{s^{\mathrm{T}}M_n^{\mathrm{T}}x_{nt}e^{x_{nt}^{\mathrm{T}}\beta_0}}{c_{nt}\left(1+e^{x_{nt}^{\mathrm{T}}\beta_0}\right)}(\epsilon_t - 1)\right\}^{\lambda+2}\right]^{1/(\lambda+2)},\\ &= \frac{r}{\tau_n(s)}\cdot\frac{s^{\mathrm{T}}M_n^{\mathrm{T}}x_{nt}e^{x_{nt}^{\mathrm{T}}\beta_0}}{c_{nt}\left(1+e^{x_{nt}^{\mathrm{T}}\beta_0}\right)}\left\{E(\epsilon_t - 1)^{\lambda+2}\right\}^{1/(\lambda+2)},\end{aligned}$$

which is bounded uniformly in t and n by the assumption $\sup_{1\leqslant t\leqslant n}|M_n^{\mathrm{T}}x_{nt}| = O(n^{-1/2})$ and Assumption 1. Finally, the third condition of Proposition 1 is assured by Assumption 2. Therefore, (A4) holds, and it follows that $F_n \to iV$ in distribution. This together with (A2) and (A3) yields

$$E(e^{is^{\mathrm{T}}U_n}) \to E\left(\exp\left[-\frac{1}{2}s^{\mathrm{T}}\{\Omega_1 + \Omega_{12}\gamma_\epsilon(0)\}s + iV\right]\right) = \exp\left\{-\frac{1}{2}s^{\mathrm{T}}(\Omega_1 + \Omega_2)s\right\},$$

from which (A1) follows.

Since g_n has convex sample paths, the convergence of $g_n(u)$ in distribution to

$$g(u) = \frac{1}{2}u^{\mathrm{T}}\Omega_1 u - u^{\mathrm{T}}N(0, \Omega_1 + \Omega_2)$$

can be extended to finite dimensional convergence, and further extended to convergence in distribution on the space $C(\mathbb{R}^l)$; see Rockafellar (1970) and Pollard (1991). The limit process $g(u)$ has a unique minimizer, say $\hat{u}$. Therefore, the estimator $\hat{u}_n = M_n^{-1}(\hat{\beta}_n - \beta_0)$ converges in distribution to $\hat{u}$. Note that $\hat{u} \sim N(0, \Omega_1^{-1} + \Omega_1^{-1}\Omega_2\Omega_1^{-1})$. The proof is complete. □

Proof of Theorem 2

It suffices to show that (A4) holds true. This can be done by adapting the technique for proving the central limit theorem of Gaussian linear processes; see Brockwell & Davis (1991) for the theoretical results required. □

References

Benjamin, M. A., Rigby, R. A. & Stasinopoulos, D. M. (2003). Generalized autoregressive moving average models. *J. Am. Statist. Assoc.* **98**, 214–23.

Blais, M., MacGibbon, B. & Roy, R. (2000). Limit theorem for regression models of time series of counts. *Statist. Prob. Lett.* **46**, 161–8.

Brännäs, K. & Johansson, P. (1994). Time series count data regression. *Commun. Statist.* A **23**, 2907–25.

Brockwell, P. J. & Davis, R. A. (1991). *Time Series: Theory and Methods*, 2nd ed. New York: Springer.

Cameron, A. C. & Trivedi, P. K. (1996). Count data models for financial data. In *Handbook of Statistics, Volume* 14*: Statistical Methods in Finance*, Ed. G. S. Maddala & C. R. Rao, pp. 363–92. Amsterdam: North-Holland.

Campbell, M. J. (1994). Time series regression for counts: an investigation into the relationship between sudden infant death syndrome and environmental temperature. *J. R. Statist. Soc.* A **157**, 191–208.

Cox, D. R. (1981). Statistical analysis of time series: some recent developments. *Scand. J. Statist.* **8**, 93–115.

Davidson, J. (1992). A central limit theorem for globally nonstationary near-epoch dependent functions of mixing processes. *Economet. Theory* **8**, 313–29.

Davis, R. A., Dunsmuir, W. T. M. & Wang, Y. (1999). Modeling time series of count data. In *Asymptotics, Nonparametrics and Time Series*, Ed. S. Ghosh, pp. 63–114. New York: Marcel Dekker.

Davis, R. A., Dunsmuir, W. T. M. & Wang, Y. (2000). On autocorrelation in a Poisson regression model. *Biometrika* **87**, 491–505.

Davis, R. A. & Rodriguez-Yam, G. (2005). Estimation for state-space models: an approximate likelihood approach. *Statist. Sinica* **15**, 381–406.

Durbin, J. & Koopman, S. J. (1997). Monte Carlo maximum likelihood estimation for non-Gaussian state space models. *Biometrika* **84**, 669–84.

Harvey, A. C. & Fernandes, C. (1989). Time series models for count or qualitative observations. *J. Bus. Econ. Statist.* **7**, 407–17.

Johansson, P. (1996). Speed limitation and motorway casualties: a time series count data regression approach. *Accident Anal. Prev.* **28**, 73–87.

Jørgensen, B., Lundbye-Christensen, S., Song, X.-K. & Sun, L. (1996). A longitudinal study of emergency room visits and air pollution for Prince George, British Columbia. *Statist. Med.* **15**, 823–36.

Jørgensen, B., Lundbye-Christensen, S., Song, X.-K. & Sun, L. (1999). A state space model for multivariate longitudinal count data. *Biometrika* **86**, 169–81.

McCullagh, P. & Nelder, J. A. (1989). *Generalized Linear Models*, 2nd ed. London: Chapman and Hall.

Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators. *Economet. Theory* **7**, 186–99.

R Development Core Team (2008). *R: A Language and Environment for Statistical Computing.* Austria: R Foundation for Statistical Computing, Vienna. ISBN 3-900051-07-0, URL http://www.R-project.org.

Rockafellar, R. T. (1970). *Convex Analysis*. Princeton, NJ: Princeton University Press.

Zeger, S. L. (1988). A regression model for time series of counts. *Biometrika* **75**, 621–29.

[*Received February* 2008. *Revised December* 2008]