Biometrika (2000), **87,** 3, *pp.* 491–505

Printed in Great Britain

On autocorrelation in a Poisson regression model

BY RICHARD A. DAVIS
Department of Statistics, Colorado State University, Fort Collins, Colorado 80523, U.S.A.
rdavis@stat.colostate.edu

WILLIAM T. M. DUNSMUIR
Department of Statistics, University of New South Wales, Sydney, New South Wales 2052, Australia
w.dunsmuir@unsw.edu.au

AND YING WANG
Department of Statistics, Colorado State University, Fort Collins, Colorado 80523, U.S.A.
yingw@stat.colostate.edu

SUMMARY

This paper develops a practical approach to diagnosing the existence of a latent stochastic process in the mean of a Poisson regression model. The asymptotic distribution of standard generalised linear model estimators is derived for the case where an autocorrelated latent process is present. Simple formulae for the effect of autocovariance on standard errors of the regression coefficients are also provided. Methods for adjusting for the severe bias in previously proposed estimators of autocovariance are derived and their behaviour is investigated. Applications of the methods to time series of monthly polio counts in the U.S.A. and daily asthma presentations at a hospital in Sydney are used to illustrate the results and methods.

Some key words: Asymptotic distribution; Autocorrelation; Generalised linear model; Latent process; Poisson count process.

1. INTRODUCTION

In this paper we are concerned with models for a time series of observed counts, $\{Y_t : t = 1, \ldots, n\}$, which have mean function specified by a linear predictor modified by a 'latent' process. Such models have been considered by Zeger (1988), Campbell (1994), Brannas & Johansson (1994) and Chan & Ledolter (1995) for example.

There has been considerable effort in recent years devoted to the development of methods to fit efficiently all the parameters in these types of model. However, existing techniques rely on the specification of a suitable model for the correlation structure in the latent process. As in linear regression with correlated errors, there is a need for model diagnostic and identification techniques to decide if it is necessary to include a latent process in the specification of the mean of the Poisson counts and, if so, to see if there is any evidence of autocorrelation in such a process. This paper follows, to some extent, approaches that have proved successful in linear regression.

To be precise we consider a nonnegative time series ε_t such that

$$Y_t | \varepsilon_t, x_t \sim \text{Po}(\varepsilon_t \mu_t), \tag{1}$$

where

$$\mu_t = e^{x_t^{\mathrm{T}} \beta},$$

in which x_t is a p-vector of observed regressors and $\beta = (\beta_1, \ldots, \beta_p)^{\mathrm{T}}$ is a vector of regression coefficients; $\text{Po}(\lambda)$ refers to a Poisson distribution with mean λ. The first component of x_t is assumed throughout to be unity so that the regression component always includes an intercept term.

It is further assumed that $\{\varepsilon_t\}$ is a nonnegative stationary time series with $E(\varepsilon_t) = 1$, $\text{var}(\varepsilon_t) = \sigma_\varepsilon^2$, autocovariance function $\gamma_\varepsilon(h) = E\{(\varepsilon_t - 1)(\varepsilon_{t+h} - 1)\}$ and autocorrelation function $\rho_\varepsilon(h) = \gamma_\varepsilon(h)/\sigma_\varepsilon^2$. The unit mean condition is imposed for identifiability reasons; otherwise, if $E(\varepsilon_t) \neq 1$, then the mean could be absorbed into the intercept term of the linear regression; see for example Chan & Ledolter (1995, p. 247). The above specification of a latent process is that suggested in Zeger (1988) and used in Campbell (1994). An alternative, used in Chan & Ledolter (1995), is

$$Y_t | \alpha_t, x_t \sim \text{Po}(e^{\alpha_t + x_t^T \beta}),$$

where α_t is a latent process which is assumed to be a stationary Gaussian process with mean μ_α, variance σ_α^2 and autocovariance function given by $\gamma_\alpha(h) = E\{(\alpha_t - \mu_\alpha)(\alpha_{t+h} - \mu_\alpha)\}$. With appropriate choice of mean and variance this is the special case of the Zeger (1988) specification in which the ε_t have a lognormal distribution. In particular, in order to satisfy the identifiability requirement that $E(\varepsilon_t) = E\{\exp(\alpha_t)\} = 1$ it is required that $\alpha_t \sim N(-\sigma_\alpha^2/2, \sigma_\alpha^2)$. Note that, for this choice of mean and variance in the lognormal distribution $\gamma_\varepsilon(h) = \exp\{\gamma_\alpha(h)\} - 1$ for all h.

The regressors x_t may depend on n and so form a triangular array. Without this dependence on n the 'information matrix' associated with the regression estimation procedure may not diverge as $n \to \infty$. Consequently the central limit theorem will not provide a useful approximation in the finite sample case. For example, if a linear time trend is included in the model, then it is necessary to divide time by the sample size n, $x_t = t/n$, to allow for the case where the trend coefficient might be negative, in which case the Poisson mean will eventually become arbitrarily close to zero.

Zeger proposes estimation of β in the above model using a quasilikelihood approach to correct for serial correlation in the latent process $\{\varepsilon_t\}$. Assumptions on the distributional properties of this process are not explicitly stated but for much of this treatment these are not required. However, for the alternative specification in terms of the $\{\alpha_t\}$ process the requirement that these be normally distributed is made quite explicitly in the treatment of Chan & Ledolter (1995).

Detection of autocorrelation in the latent process using the observed count process $\{Y_t\}$ is not straightforward. Using results in Zeger (1988) and Davis, Dunsmuir & Wang (1999) we can show that

$$\rho_Y(h) = \text{corr}(Y_t, Y_{t+h}) = \frac{\rho_\varepsilon(h)}{\{1 + (\sigma_\varepsilon^2 \mu_t)^{-1}\}^{\frac{1}{2}} \{1 + (\sigma_\varepsilon^2 \mu_{t+h})^{-1}\}^{\frac{1}{2}}},$$

where $\rho_\varepsilon(h)$ is the autocorrelation function of the latent process. Note that $\rho_Y(h)$ is not free of the regressors x_t as is to be expected. When ε_t is lognormal and $p = 1$, so that the

regression consists only of a constant mean term, $\{Y_t\}$ is stationary with

$$\rho_Y(h) = \frac{e^{\gamma_\alpha(h)} - 1}{\mu^{-1} + (e^{\sigma_\alpha^2} - 1)},$$

where $\mu = \exp\{\beta_1\}$ and $h > 0$. Since the function $(e^x - 1)/x$ is nondecreasing for $x \geqslant 0$, it follows that, for $\gamma_\alpha(h) \geqslant 0$,

$$0 \leqslant \rho_Y(h) \leqslant \frac{e^{\gamma_\alpha(h)} - 1}{e^{\sigma_\alpha^2} - 1} \leqslant \frac{\gamma_\alpha(h)}{\sigma_\alpha^2} = \rho_\alpha(h).$$

This implies that the autocorrelation function for the Poisson count process is dominated by that for the underlying latent process. This last observation illustrates the difficulty in detecting dependence within the latent process. Little or no autocorrelation observed in the Y_t process may mask significant correlation in the latent process. Typically use of the autocorrelations for Y_t will lead to underestimates of the true size of autocorrelation in the latent process α_t as noted in Zeger (1988).

Also, in practice, the regression on x_t will need to be performed before we attempt to estimate this autocorrelation. In order to test for the existence of a latent process and subsequently identify its correlation structure, it is necessary to have a consistent estimation procedure for the regression coefficient vector. A natural and easy way to compute consistent estimators of β is to use a standard generalised linear model analysis. In § 2, we establish the consistency and asymptotic normality of the generalised linear model estimators $\hat{\beta}$ when a stationary autocorrelated process is present in the mean of the Poisson counts. These results are analogous to long standing results for linear regression with autocorrelated errors. They differ in that, while the ordinary least squares estimator is fully efficient for many standard trend and seasonal regressors, the generalised linear model estimators are not efficient. The main theorem of § 2 also provides the formula for standard errors of $\hat{\beta}$, including the correct adjustment for the presence of a latent process.

In § 3·1, the main result of § 2 is illustrated on the polio data in Zeger. Bias adjustments of autocovariance and autocorrelation estimators, the form of which depends critically on the asymptotics derived in § 2 for the generalised linear model estimators, are described in § 3·2. Section 3·3 proposes and discusses some modifications to standard Ljung–Box portmanteau tests for serial dependence. In § 3·4, the techniques developed in earlier sections are applied to the time series of asthma presentations at a Sydney hospital.

2. Large-sample properties of generalised linear model estimators

2·1. *Consistency and asymptotic normality*

Let $Y_1, \ldots, Y_n$ be observations from model (1) with true value $\beta = \beta_0$. The generalised linear model estimators $\hat{\beta}$ are obtained by maximising

$$l(\beta) = -\sum_{t=1}^{n} e^{x_t^{\mathrm{T}}\beta} + \sum_{t=1}^{n} Y_t x_t^{\mathrm{T}}\beta - \log\left\{\prod_{t=1}^{n} (Y_t!)\right\}, \tag{2}$$

the loglikelihood constructed ignoring the presence of a latent process in the model.

We assume there exists a sequence of nonsingular matrices M_n such that the regressors

obey the following conditions:

$$M_n^{\mathrm{T}}\left(\sum_{t=1}^{n} x_t x_t^{\mathrm{T}} e^{x_t^{\mathrm{T}}\beta_0}\right) M_n \to \Omega_I, \tag{3}$$

$$M_n^{\mathrm{T}}\left(\sum_{t=1}^{n} x_t x_{t+h}^{\mathrm{T}} e^{(x_t^{\mathrm{T}}+x_{t+h}^{\mathrm{T}})\beta_0}\right) M_n \to W_h, \tag{4}$$

uniformly in h.

For $h \leqslant 0$,

$$M_n^{\mathrm{T}}\left(\sum_{t=1}^{1-h} x_t x_{t+h}^{\mathrm{T}} e^{(x_t^{\mathrm{T}}+x_{t+h}^{\mathrm{T}})\beta_0}\right) M_n \to 0,$$

the left-hand side being uniformly bounded in h as $n \to \infty$, and, for $h > 0$,

$$M_n^{\mathrm{T}}\left(\sum_{t=n-h}^{n} x_t x_{t+h}^{\mathrm{T}} e^{(x_t^{\mathrm{T}}+x_{t+h}^{\mathrm{T}})\beta_0}\right) M_n \to 0, \tag{5}$$

again with the left-hand side uniformly bounded in h as $n \to \infty$.

Let

$$\Omega_{I,n} = \sum_{t=1}^{n} x_t x_t^{\mathrm{T}} e^{x_t^{\mathrm{T}}\beta_0}, \quad \Omega_{II,n} = \sum_{t=1}^{n}\sum_{s=1}^{n} x_t x_s^{\mathrm{T}} e^{x_t^{\mathrm{T}}\beta_0} e^{x_s^{\mathrm{T}}\beta_0}\gamma_\varepsilon(t-s),$$

and define

$$\Omega_n := \operatorname{cov}\left\{M_n^{\mathrm{T}}\sum_{t=1}^{n} x_t(Y_t - \mu_t)\right\} = M_n^{\mathrm{T}}(\Omega_{I,n} + \Omega_{II,n})M_n.$$

Under conditions (3)–(5) and $\sum_{h=0}^{\infty}|\gamma_\varepsilon(h)| < \infty$, we can show that

$$\Omega_n \to \Omega_I + \Omega_{II}, \tag{6}$$

where $\Omega_{II} = \sum_h W_h\gamma_\varepsilon(h)$. In § 2·2, we will examine convergence of these quantities for a large class of regression functions. The following theorem, whose proof is given in the Appendix, describes the asymptotic properties of the generalised linear model estimators.

THEOREM 1. *Let $Y_1, \ldots, Y_n$ be observations from the model* (1) *with true parameter value $\beta = \beta_0$, where $\{\varepsilon_t = e^{\alpha_t}\}$ and $\{\alpha_t\}$ is a linear Gaussian process with mean $-\sigma_\alpha^2/2$ and variance σ_α^2. Let $\hat{\beta}$ be the generalised linear model estimator of β obtained by maximising $l(\beta)$. Assume that the $\{x_t\}$ satisfy* (3)–(5), $\sum_{h=0}^{\infty}|\gamma_\varepsilon(h)| < \infty$, *and as $n \to \infty$*

$$\sup_{1 \leqslant t \leqslant n} |M_n^{\mathrm{T}} x_t e^{x_t^{\mathrm{T}}\beta_0}| \to 0. \tag{7}$$

Then, for any real vector s,

$$C_n(s) := \sum_{t=1}^{n} s^{\mathrm{T}} M_n^{\mathrm{T}} x_t e^{x_t^{\mathrm{T}}\beta_0}(\varepsilon_t - 1) \to N(0, s^{\mathrm{T}}\Omega_{II}s), \tag{8}$$

in distribution. Moreover, $\hat{\beta} \to \beta_0$, in probability, and

$$M_n^{-1}(\hat{\beta} - \beta_0) \to N(0, \Gamma),$$

in distribution, where

$$\Gamma = \Omega_I^{-1} + \Omega_I^{-1}\Omega_{II}\Omega_I^{-1}. \tag{9}$$

Remark 1. Note that $M_n^{\mathrm{T}}\Omega_I^{-1}M_n$ is the asymptotic covariance matrix obtained from a standard generalised linear model analysis and that $M_n^{\mathrm{T}}\Omega_I^{-1}\Omega_{II}\Omega_I^{-1}M_n$ is the additional contribution to the asymptotic covariance caused by the existence of the latent process.

Remark 2. The theorem can be extended to cover other forms of latent processes including the cases when $\{\varepsilon_t\}$ is a stationary nonnegative linear process and $\{\varepsilon_t\}$ is a stationary mixing process. The key to the proof is establishing (8).

Remark 3. The theorem is also valid when the regression variables depend on n and form a triangular array x_{nt}.

2·2. *Conditions on the regressors*

A wide range of regression functions satisfy the conditions of Theorem 1. These include a myriad of trend functions that form triangular arrays, harmonic functions and stationary processes. For triangular arrays of a particular type, the asymptotic covariance matrix of the generalised linear model estimator has a simple form.

Trend functions. Suppose that the trend function x_t is a triangular array that can be expressed as

$$x_{nt} = f(t/n)$$

for some continuous vector-valued function $f(.)$ defined on the unit interval $[0, 1]$. In this case, take $M_n = n^{-\frac{1}{2}}I_p$, where I_p is a $p \times p$ identity matrix. Then

$$\Omega_n = \frac{1}{n}(\Omega_{I,n} + \Omega_{II,n}) = \frac{1}{n}\sum_{t=1}^{n} x_{nt}x_{nt}^{\mathrm{T}}\mu_t + \frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{n} x_{nt}x_{ns}^{\mathrm{T}}\mu_t\mu_s\gamma_\varepsilon(s-t),$$

where, using an integral approximation to the sums, we obtain that

$$\frac{1}{n}\sum_{t=1}^{n} x_{nt}x_{nt}^{\mathrm{T}}\mu_t \to \int_0^1 f(x)f^{\mathrm{T}}(x)e^{f^{\mathrm{T}}(x)\beta_0}\,dx,$$

$$\frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{n} x_{nt}x_{ns}^{\mathrm{T}}\mu_t\mu_s\gamma_\varepsilon(s-t) \to \sum_h \left(\int_0^1 f(x)f^{\mathrm{T}}(x)e^{2f^{\mathrm{T}}(x)\beta_0}\,dx\right)\gamma_\varepsilon(h),$$

provided $\sum_h |\gamma_\varepsilon(h)| < \infty$. Thus,

$$\Omega_n \to \int_0^1 f(x)f^{\mathrm{T}}(x)e^{f^{\mathrm{T}}(x)\beta_0}\,dx + \left(\int_0^1 f(x)f^{\mathrm{T}}(x)e^{2f^{\mathrm{T}}(x)\beta_0}\,dx\right)\sum_h \gamma_\varepsilon(h).$$

Note that, if $f(x) = (1, x)^{\mathrm{T}}$, which corresponds to the linear regression function of the form $x_{nt} = (1, t/n)^{\mathrm{T}}$, then the generalised linear model estimators are consistent and have the asymptotic normal distribution specified in Theorem 1 for all values of β_0. On the other hand, if the sample size scaling n in the linear component is omitted, that is $x_t = (1, t)^{\mathrm{T}}$, then the generalised linear model estimator would not be consistent for negative values of the slope parameter. This is because the Poisson mean is converging to 0 rapidly.

Harmonic functions. Suppose that the regression function includes harmonic terms to specify annual effects or day-of-the-week effects. This is an example of an asymptotically stationary process and the convergence of the matrix Ω_n can be established using elementary properties of trigonometric functions.

Stationary processes. Stationary processes can arise in the regression function, as for example with seasonally adjusted temperature series; see Campbell (1994) for an example of this. Ergodic properties of the process can be used to establish convergence of the covariance matrix Ω_n.

3. APPLICATIONS

3·1. *Analysis of Zeger's polio data*

We now apply the results of Theorem 1 to the polio data example from Zeger (1988). These data consist of the monthly numbers of cases of poliomyelitis in the U.S.A. for the years 1970–1983 as reported by the Center for Disease Control. The data reveal some seasonality and the possibility of a slight decreasing trend. Detection of the decreasing trend is one of the main objectives in modelling this series. Here we use the same regression variables as in Zeger (1988) consisting of an intercept term, a linear trend and harmonics at periods of 6 and 12 months. Specifically,

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

where $t' = (t - 73)$ is used to locate the intercept term at January 1976 as in Zeger's analysis. Table 1 summarises Zeger's estimates of β, based on an estimating equation approach, and generalised linear model estimators arising from a standard generalised linear model fit.

Table 1. *Estimates and their standard errors* (SE) *from an analysis of the Zeger* (1988) *polio data*

	Zeger's method		GLM fit		Asym.	Simulations	
Covariate	$\hat{\beta}_Z$	SE	$\hat{\beta}_{\mathrm{GLM}}$	SE	SE	ave$(\hat{\beta}_{\mathrm{GLM}})$	SD $(\hat{\beta}_{\mathrm{GLM}})$
Intercept	0·17	0·13	0·207	0·075	0·205	0·150	0·213
Trend $\times 10^{-3}$	−4·35	2·68	−4·799	1·399	4·115	−4·887	3·937
$\cos(2\pi t/12)$	−0·11	0·16	−0·149	0·097	0·157	−0·145	0·144
$\sin(2\pi t/12)$	−0·48	0·17	−0·532	0·109	0·168	−0·531	0·168
$\cos(4\pi t/12)$	0·20	0·14	0·169	0·098	0·122	0·167	0·123
$\sin(4\pi t/12)$	−0·41	0·14	−0·432	0·101	0·125	−0·440	0·125

GLM, generalised linear model; asym. SE, asymptotic standard error; SD, standard deviation.

The asymptotic standard errors for the generalised linear model estimators given in Theorem 1 are estimated using the values of $\hat{\sigma}_\varepsilon^2 = 0{\cdot}77$ and $\hat{\rho}_\varepsilon(h) = (0{\cdot}77)^h$ reported in Table 3 of Zeger (1988). These were obtained using the formulae

$$\mathrm{var}(\hat{\beta}_{\mathrm{GLM}}) = \hat{\Omega}_{I,n}^{-1} + \hat{\Omega}_{I,n}^{-1}\hat{\Omega}_{II,n}\hat{\Omega}_{I,n}^{-1},$$

$$\hat{\Omega}_{I,n} = \sum_{t=1}^{n} x_t x_t^{\mathrm{T}} e^{x_t^{\mathrm{T}}\hat{\beta}_{\mathrm{GLM}}}, \quad \hat{\Omega}_{II,n} = \sum_{t=1}^{n}\sum_{s=1}^{n} x_t x_s^{\mathrm{T}} e^{(x_t^{\mathrm{T}} + x_s^{\mathrm{T}})\hat{\beta}_{\mathrm{GLM}}}\hat{\gamma}_\varepsilon(s-t).$$

Use of the correct standard errors for the trend term would lead to the conclusion that the trend is not significant whereas use of the standard errors produced by the generalised linear model analysis would falsely lead to declaring the trend to be significant.

The final two columns of Table 1 report the results of 1000 simulations of a time series of length $n = 168$ using $\hat{\beta}_{\mathrm{GLM}}$ as the true values. The latent process in this simulation was assumed to be a lognormal autoregression of order 1 with mean $-\sigma_\alpha^2/2 = -0{\cdot}285$, $\phi = 0{\cdot}82$ and $\sigma_\alpha^2 = 0{\cdot}57$. These values of the parameters were chosen in order to match

the second-order properties of the fitted $\{\varepsilon_t\}$ process. The average over the 1000 simulated values of the intercept estimate is observed to be 0·150, which is significantly biased downwards from the true value of 0·207 used in the simulations. The other parameters are estimated without substantial bias. The standard deviations of the generalised linear model estimators observed over the 1000 replications are reported in the last column of Table 1. These are in good agreement with the standard errors obtained from the asymptotic theory reported in the fifth column.

3·2. *Estimation of autocovariances and autocorrelations in the latent process*

Various estimators of the autocovariances from the literature are reviewed in Davis et al. (1999). Here we will focus on the estimators proposed in Zeger (1988). The variance σ_ε^2 is estimated by

$$\hat{\sigma}_\varepsilon^2 = \hat{\gamma}_\varepsilon(0) = \frac{\sum_{t=1}^n \{(Y_t - \hat{\mu}_t)^2 - \hat{\mu}_t\}}{\sum_{t=1}^n \hat{\mu}_t^2},$$

while the estimator of the autocovariance function at positive lag τ is

$$\hat{\gamma}_\varepsilon(\tau) = \sum_{t=\tau+1}^{n} \{(Y_t - \hat{\mu}_t)(Y_{t-\tau} - \hat{\mu}_{t-\tau})\} \Big/ \sum_{t=\tau+1}^{n} \hat{\mu}_t \hat{\mu}_{t-\tau},$$

where $\hat{\mu}_t = e^{x_t^T \hat{\beta}}$. Combining these gives estimators of the autocorrelation function of the latent process:

$$\hat{\rho}_Z(\tau) = \hat{\gamma}_\varepsilon(\tau)/\hat{\sigma}_\varepsilon^2. \tag{10}$$

Note that the estimators of variance and autocovariances are exactly unbiased, if $\hat{\mu}_t$ is replaced by the true value μ_t and, although one might expect $\hat{\sigma}_\varepsilon^2$ to be approximately unbiased, the use of the estimated values, $\hat{\mu}_t$, leads to substantial bias. Exploratory simulations showed that for the Zeger estimator of variance the numerator seriously underestimates $\sum_{t=1}^n \{(Y_t - \mu_t)^2 - \mu_t\}$ while the denominator overestimates $\sum_{t=1}^n \mu_t^2$. Clearly, both directions of bias in these terms contribute to the bias in the ratio used to define $\hat{\sigma}_\varepsilon^2$.

To correct for the bias due to estimation of μ_t, we use the asymptotics from Theorem 1. First note that we may write

$$\hat{\mu}_t = \mu_t \exp\{x_t^T(\hat{\beta} - \beta_0)\}.$$

Since $(\hat{\beta} - \beta_0)$ is approximately normally distributed with mean 0 and covariance matrix $G_n = \Omega_{I,n}^{-1} + \Omega_{I,n}^{-1}\Omega_{II,n}\Omega_{I,n}^{-1}$, $\hat{\mu}_t$ has an approximate lognormal distribution with mean

$$E(\hat{\mu}_t) = \mu_t E[\exp\{x_t^T(\hat{\beta} - \beta_0)\}] \simeq \mu_t \exp(x_t^T G_n x_t/2) \tag{11}$$

and second moment

$$E(\hat{\mu}_t^2) = \mu_t^2 E[\exp\{2x_t^T(\hat{\beta} - \beta_0)\}] \simeq \mu_t^2 \exp(2x_t^T G_n x_t). \tag{12}$$

Hence $\hat{\mu}_t$ and $\hat{\mu}_t^2$ will have positive bias. Now consider the term

$$\frac{1}{n}\sum_{t=1}^{n}(Y_t - \mu_t)^2 = \frac{1}{n}\sum_{t=1}^{n}(Y_t - \hat{\mu}_t)^2 + \frac{1}{n}\sum_{t=1}^{n}(\hat{\mu}_t - \mu_t)^2 + \frac{2}{n}\sum_{t=1}^{n}(Y_t - \hat{\mu}_t)(\hat{\mu}_t - \mu_t).$$

By analogy with standard regression theory, the last term is negligible. If we use (11) and

(12), the second term has approximate expectation given by

$$E\left\{\frac{1}{n}\sum_{t=1}^{n}(\hat{\mu}_t-\mu_t)^2\right\}\simeq\frac{1}{n}\sum_{t=1}^{n}\mu_t^2(e^{2x_t^{\mathrm{T}}G_nx_t}-2e^{x_t^{\mathrm{T}}G_nx_t/2}+1). \tag{13}$$

Since an approximately unbiased estimator of μ_t^2 is $\hat{\mu}_t^2\exp(-2x_t^{\mathrm{T}}G_nx_t)$, approximate adjustments for the biases in the numerator and denominator of Zeger's estimators suggest an estimator of variance given by

$$\hat{\sigma}^2_{\varepsilon,\mathrm{UB}}=\frac{\sum_{t=1}^{n}\{(Y_t-\hat{\mu}_t)^2+\hat{\mu}_t^2e^{-2x_t^{\mathrm{T}}\hat{G}_nx_t}(e^{2x_t^{\mathrm{T}}\hat{G}_nx_t}-2e^{x_t^{\mathrm{T}}\hat{G}_nx_t/2}+1)-\hat{\mu}_t\}}{\sum_{t=1}^{n}\hat{\mu}_t^2\exp(-2x_t^{\mathrm{T}}\hat{G}_nx_t)},$$

in which the estimator of the asymptotic covariance matrix is given by

$$\hat{G}_n=\hat{\Omega}_{I,n}^{-1}+\hat{\Omega}_{I,n}^{-1}\hat{\Omega}_{II,n}\hat{\Omega}_{I,n}^{-1}, \tag{14}$$

where

$$\hat{\Omega}_{I,n}=\sum_{t=1}^{n}x_tx_t^{\mathrm{T}}\hat{\mu}_t,\quad \hat{\Omega}_{II,n}=\sum_{h=-L}^{L}\sum_{t=\max(1-h,1)}^{\min(n-h,n)}x_tx_{t+h}^{\mathrm{T}}\hat{\mu}_t\hat{\mu}_{t+h}\hat{\gamma}_{\varepsilon}(h)$$

and the maximum lag L is chosen to give a good approximation to the infinite series.

Employing similar approximations we suggest the following bias-corrected estimators for the autocovariances:

$$\hat{\gamma}_{\varepsilon,\mathrm{UB}}(h)=\left(\sum_{t=1}^{n-h}\hat{\mu}_t\hat{\mu}_{t+h}g_{t,h}\right)^{-1}\sum_{t=1}^{n-h}\{(Y_t-\hat{\mu}_t)(Y_{t+h}-\hat{\mu}_{t+h})$$
$$+\hat{\mu}_t\hat{\mu}_{t+h}g_{t,h}(1-e^{x_t^{\mathrm{T}}\hat{G}_nx_t/2}-e^{x_{t+h}^{\mathrm{T}}\hat{G}_nx_{t+h}/2}+1/g_{t,h})\},$$

where

$$g_{t,h}=\exp\{-(x_t+x_{t+h})^{\mathrm{T}}\hat{G}_n(x_t+x_{t+h})/2\}.$$

Bias-adjusted autocorrelation estimators are then defined as

$$\hat{\rho}_{\mathrm{Z,UB}}(h)=\hat{\gamma}_{\varepsilon,\mathrm{UB}}(h)/\hat{\sigma}^2_{\varepsilon,\mathrm{UB}}. \tag{15}$$

It is straightforward to show that these adjusted estimators are consistent. In particular $\hat{G}_n$ will converge to zero as $n\to\infty$. This implies that the adjustment to the denominator will tend to unity while that of the numerator will tend to zero, as is required for the adjustments to be asymptotically negligible.

Prior to examining the autocovariance structure of the latent process, we recommend testing for the existence of the latent process. There have been a number of tests proposed in the literature for this task; see Davis et al. (1999) for a review. We shall use the test based on the statistic S_a described by Dean & Lawless (1989) which is a variant of a test proposed by Brannas & Johansson (1994).

Tables 2(a) and (b) compare the Zeger and bias-adjusted Zeger estimators of autocovariance and autocorrelation for two Poisson regression models. The sample size is 100 and the latent process in both cases is assumed to be a lognormal AR(1) process with $\phi=0{\cdot}9$ and $\sigma_\alpha^2=0{\cdot}6931$. This choice of parameter values implies that the autocovariance and autocorrelation functions of the latent process are the same. The summarised results are based on 1000 replications. In the computation of the bias-adjustment factors, we used $L=15$ in (14) for all the realisations. The autocorrelation function results are conditional

Table 2. *Autocovariance and autocorrelation estimators of a lognormal* AR(1) *latent process in model* (1): (a) *with regression function* $1 + t/100$, *where in* 3 *out of* 1000 *replications* S_a *was less than* 1·645; *and* (b) *with regression function* $1 + \cos(2\pi t/12)$, *where in* 5 *out of* 1000 *replications* S_a *was less than* 1·645

(a) *Regression function* $1 + t/100$

		ACVF mean		ACVF SD		ACF mean		ACF SD	
Lag	True	Z	Z,UB	Z	Z,UB	Z	Z,UB	Z	Z,UB
0	1·00	0·49	0·71	0·29	0·59				
1	0·87	0·39	0·59	0·26	0·53	0·79	0·81	0·18	0·16
2	0·75	0·31	0·49	0·23	0·47	0·61	0·65	0·20	0·19
3	0·66	0·24	0·41	0·20	0·42	0·46	0·52	0·22	0·21
4	0·58	0·19	0·34	0·18	0·38	0·35	0·42	0·23	0·22
5	0·51	0·14	0·27	0·16	0·34	0·25	0·32	0·24	0·24
6	0·45	0·10	0·22	0·15	0·31	0·17	0·25	0·26	0·25

(b) *Regression function* $1 + \cos(2\pi t/12)$

		ACVF mean		ACVF SD		ACF mean		ACF SD	
Lag	True	Z	Z,UB	Z	Z,UB	Z	Z,UB	Z	Z,UB
0	1·00	0·60	0·80	0·38	0·62				
1	0·87	0·49	0·67	0·35	0·57	0·81	0·82	0·21	0·19
2	0·75	0·42	0·58	0·31	0·50	0·68	0·70	0·22	0·21
3	0·66	0·36	0·50	0·28	0·46	0·57	0·59	0·26	0·25
4	0·58	0·30	0·42	0·26	0·42	0·48	0·50	0·30	0·29
5	0·51	0·25	0·37	0·23	0·38	0·41	0·44	0·31	0·31
6	0·45	0·21	0·31	0·21	0·34	0·34	0·37	0·31	0·31

ACVF mean, mean of the autocovariance function; ACVF SD, standard deviation of the autocovariance function; ACF mean, mean of the autocorrelation function; ACF SD, standard deviation of the autocorrelation function; Z, estimates based on Zeger's (1988) method; Z,UB, estimates based on a bias-adjusted version of Zeger's (1988) method.

on $S_a > 1{\cdot}645$, so that the null hypothesis of no latent process is rejected at the 5% level. This was done to eliminate values of autocorrelations badly affected by zero or near zero variance estimators.

Note that the bias-improved versions of the estimators of the autocovariances do indeed have better bias properties but at the expense of higher variance. Autocovariances are important in properly estimating the variances of the generalised linear model estimators. Since the unadjusted estimators are seriously biased towards zero, at least when there is positive autocovariance, the effect of failing to adjust for bias will be to use excessively small standard errors, leading to declaration of significant regressor effects too frequently. The bias-adjusted versions of the autocorrelation function are slightly better in bias and standard deviation than the unadjusted estimators. This means that, for correlation estimation and identification purposes, the bias-adjusted estimators are preferable to the unadjusted versions.

However, for purposes in which an unbiased estimator of scale is required, even the bias-adjusted estimator of σ_ε^2 is biased towards 0. The effect of the bias of $\hat{\sigma}_\varepsilon^2$ on the correction factors was examined using the true value of σ_ε^2 in the correction factor (14). Under this scenario, the bias of the bias-corrected autocovariance function estimators was

substantially reduced. Thus the bias in estimating σ_ε^2 has a significant impact on the resulting correction factors.

Although the estimators $\hat{\gamma}_\varepsilon(h)$ are biased we do not recommend iterated estimators obtained by replacing $\hat{\gamma}_\varepsilon(h)$ with updated $\hat{\gamma}_{\varepsilon,\mathrm{UB}}(h)$. Such a replacement method, if iterated, would lead to divergent estimators because the adjustment term in the denominator would get progressively closer to zero. An alternative procedure is to model the $\hat{\gamma}_\varepsilon(h)$ by use of an autoregressive process or other suitable model. This does not appear to lead to better estimators than those based on the nonparametric forms proposed above.

3·3. *Testing for autocorrelation in the latent process*

Brannas & Johansson (1994) studied the performance of the Box–Pierce and Ljung–Box portmanteau test statistics for detection of overall serial correlation in the latent process. They considered several types of residual as the basis for these tests, including the Pearson residuals,

$$\varepsilon_t = (Y_t - \hat{\mu}_t)/\hat{\mu}_t^{1/2}.$$

Based on these, the Ljung–Box statistic can be formed as

$$H_{\mathrm{LB}}^2 = \sum_{h=1}^{L} \{\hat{\rho}_{\mathrm{P}}(h)\}^2/\tilde{V}_h,$$

where $\tilde{V}_h = (n-h)/\{n(n+2)\}$. There are two potential difficulties with this test statistic for detecting serial dependence in the latent process of the Poisson regression model. First, and contrary to Brannas & Johansson's claim, neither these residuals nor the others considered by them possess the correlation structure of the latent process. To illustrate this, consider the sample autocorrelation function based on the Pearson residuals, defined by

$$\hat{\rho}_{\mathrm{P}}(h) = \frac{\sum_{t=1}^{n-h} e_t e_{t+h}}{\sum_{t=1}^{n} e_t^2}.$$

Mean correction of the e_t is not utilised since their sample mean is near 0. Approximating the mean of the ratio by the ratio of the means and using results from § 2·2, we obtain

$$E\{\hat{\rho}_{\mathrm{P}}(h)\} \simeq \frac{n^{-1}\sum_{t=1}^{n-h} \mu_t^{1/2}\mu_{t+h}^{1/2}}{n^{-1}\sum_{t=1}^{n}(\sigma_\varepsilon^{-2}+\mu_t)}\,\rho_\varepsilon(h) \simeq \frac{\int_0^1 e^{f(x)\beta_0}\,dx}{\sigma_\varepsilon^{-2}+\int_0^1 e^{f(x)\beta_0}\,dx}\,\rho_\varepsilon(h), \tag{16}$$

under the assumption that $x_t = f(t/n)$. As $\sigma_\varepsilon^2 \to 0$, we see that the mean value of $\hat{\rho}_{\mathrm{P}}(h)$ becomes arbitrarily small. Thus, for alternatives consisting of a highly correlated latent process for which σ_ε^2 is small, the sample autocorrelation function of the Pearson residuals may not provide any evidence of correlation, and tests based on Pearson residuals will often fail to reject the null hypothesis of white noise. This is precisely the problem encountered in the analysis of the asthma data discussed later in § 3·4.

A second potential problem with using the standard definition of the Ljung–Box statistic, H_{LB}^2, is that, under the null hypothesis of a white noise latent process for our Poisson model, the process of Pearson residuals $\{e_t\}$ is heteroscedastic and hence nonstationary. Since the variance and covariances of the process have different forms of dependence on the mean function μ_t, there is no single normalisation of the residuals that will simultaneously eliminate this dependence from the variance and the covariance, and so it is not possible to define residuals which overcome all problems associated with heteroscedas-

ticity. However, it is possible to use alternative estimators of the variance of the autocorrelation estimators in place of the $\tilde{V}_h$ used in H^2_{LB}. By the same lines of reasoning as in Lo & MacKinlay (1989) and in a University of Iowa technical report by I. Lobato, J. Nankervis and N. E. Savin for other heteroscedastic residuals, the variance of the correlations under the white noise latent process assumption can be approximated by

$$\begin{aligned}\mathrm{var}\{\hat{\rho}_{\mathrm{P}}(h)\} &\simeq \left\{\frac{1}{n}\sum_{t=1}^{n}\mu_t^{-1}E(Y_t-\mu_t)^2\right\}^{-2}\\ &\quad\times\left\{\frac{1}{n^2}\sum_{t=1}^{n-h}\mu_t^{-1}\mu_{t+h}^{-1}E(Y_t-\mu_t)^2E(Y_{t+h}-\mu_{t+h})^2\right\}\\ &= \left\{\frac{1}{n}\sum_{t=1}^{n}(1+\mu_t\sigma_\varepsilon^2)\right\}^{-2}\left\{\frac{1}{n^2}\sum_{t=1}^{n-h}(1+\mu_t\sigma_\varepsilon^2)(1+\mu_{t+h}\sigma_\varepsilon^2)\right\}\\ &= V_h,\end{aligned}$$

say. Davis et al. (1999) give some comparisons of V_h with $\tilde{V}_h$. For some of the lags there are considerable differences. For example, in the cosine regression introduced in § 3·2, $V_h = 1{\cdot}19/n$ and $\tilde{V}_h = 0{\cdot}97/n$ for $h = 1$, and $V_h = 0{\cdot}72/n$ and $\tilde{V}_h = 0{\cdot}92/n$ for $h = 6$. This means that care should be taken when testing a particular lag autocorrelation using the standard estimation technique based on the Pearson residuals.

In light of these observations we propose a modification, which we will label $H^2_{\mathrm{LB,M}}$, to the Ljung–Box statistic in which $\tilde{V}_h$ is replaced by an estimator $\hat{V}_h$ of V_h obtained by using estimators $\hat{\mu}_t$ and $\hat{\sigma}^2_\varepsilon$ based on $\hat{\beta}_{\mathrm{GLM}}$.

We also propose here an alternative test statistic based on the bias-adjusted estimators of autocorrelation discussed above. For a given maximum lag L and the bias-adjusted estimators, define

$$H^2_{\mathrm{Z,UB}} = \sum_{h=1}^{L}\hat{\rho}^2_{\mathrm{Z,UB}}(h)/\hat{V}\{\hat{\rho}_{\mathrm{Z,UB}}(h)\}, \tag{17}$$

where, under the assumption that the latent process is independent, the estimated variance of $\hat{\rho}_{\mathrm{Z,UB}}(h)$ is given by

$$\hat{V}\{\hat{\rho}_{\mathrm{Z,UB}}(h)\} = \left\{\sum_{t=1}^{n-h}(\hat{\mu}_t\hat{\mu}_{t+h})\right\}^{-2}\sum_{t=1}^{n-h}\hat{\mu}_t^2\hat{\mu}_{t+h}^2(1+\hat{\mu}_t^{-1}\hat{\sigma}^{-2}_{\varepsilon,\mathrm{UB}})(1+\hat{\mu}_{t+h}^{-1}\hat{\sigma}^{-2}_{\varepsilon,\mathrm{UB}});$$

see Davis et al. (1999). Under the hypothesis of independence of the latent process $H^2_{\mathrm{Z,UB}}$ will have an approximate χ^2 distribution on L degrees of freedom.

A simulation comparison of the empirical size and power of several tests of autocorrelation in the latent process was reported in Davis et al. (1999). The results were mixed and highly dependent on the form of the regression function. In a limited simulation study the performance of the test statistics $H^2_{\mathrm{Z,UB}}$ and H^2_{Z}, based on $\hat{\rho}^2_{\mathrm{Z}}$ in (10), were compared for both the linear and cosine regression functions. It was found that $H_{\mathrm{Z,UB}}$ was uniformly more powerful than H^2_{Z} after the sizes of the two tests had been calibrated to the nominal level.

3·4. *Application to Sydney asthma count data*

We illustrate the above methods with an application to counts of daily admissions for asthma to Campbelltown Hospital in the Sydney metropolitan area. Figure 1 shows the

daily number of asthma presentations from 1 January 1990–31 December 1993. These data are described in greater detail in Davis et al. (1999). An analysis of temporally related effects identified the following: (i) no upward or downward trend in counts at this location; (ii) a triple peaked annual cycle modelled by pairs of the form $\cos(2\pi kt/365)$, $\sin(2\pi kt/365)$, where t is the day number and $k = 1, 2, 3, 4, 5, 8$; and (iii) a day of the week effect best characterised by separate indicator variables for Sundays and Mondays to model the increased level of admittance for these days compared to that for Tuesday to Saturday.

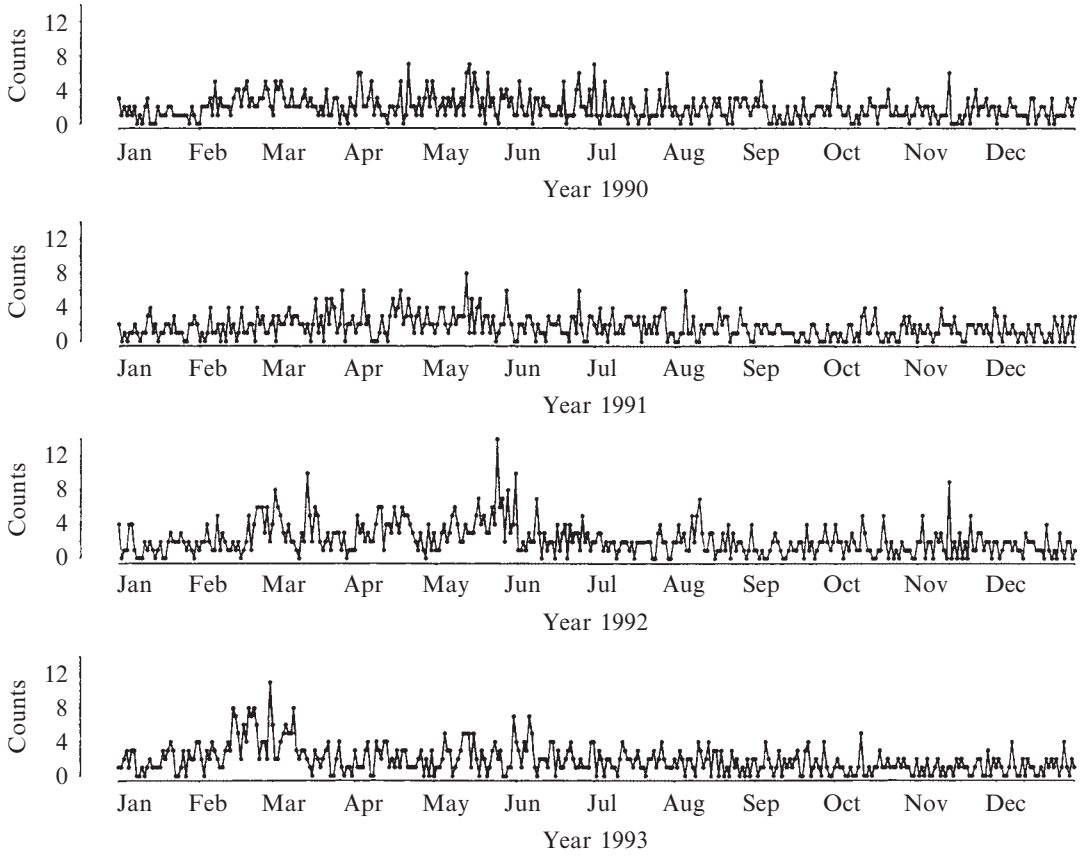

Fig. 1. Asthma presentations at Campbelltown hospital, Sydney, from 1 January 1990 to 31 December 1993.

A detailed preliminary analysis of the possible effects of meteorological variables, daily maximum and minimum temperatures and humidity, and pollution variables ozone, NO and NO_2 identified humidity at lags of approximately 12 to 18 days as the only variable that appears to have an association with asthma presentations. A humidity variable, H_t, was constructed as

$$H_t = \frac{1}{7}\sum_{i=0}^{6} h_{t-12-i},$$

where h_t is the residual from an annual-cycle harmonic model fitted to the daily average value of humidity at 0900 and 1500 hours.

In Tables 3, 4 and 5, the results for the coefficients of the harmonic components are not given because they are of no interest to the analysis. Results of the ordinary generalised

linear model fit as well as various standard errors and diagnostic procedures for the presence of a latent process are provided in Tables 3 and 4.

Table 3. *Generalised linear model regression estimators for Campbelltown, Sydney asthma count series. The two standard errors are computed using a standard generalised linear model analysis without a latent process and using Theorem* 1, *respectively*

Effect	$\hat{\beta}$	GLM SE $\{\hat{\beta}\}$	$\hat{G}$ SE $\{\hat{\beta}\}$
Sunday	0·230	0·051	0·055
Monday	0·236	0·051	0·055
H_t	0·210	0·048	0·066

GLM SE $\{\hat{\beta}\}$, standard error of $\hat{\beta}$ based on a generalised linear model fit assuming independence; $\hat{G}$ SE $\{\hat{\beta}\}$, standard error of $\hat{\beta}$ based on the asymptotic covariance matrix given in (14).

Table 4. *Tests of correlation in the latent process for the Campbelltown, Sydney asthma data with p-values in parentheses*

	Degrees of freedom		
Test statistic	5	10	15
$H^2_{Z,UB}$	44·63 ($1{\cdot}72 \times 10^{-8}$)	74·86 ($5{\cdot}08 \times 10^{-12}$)	81·32 ($4{\cdot}00 \times 10^{-11}$)
$H^2_{LB,M}$	10·78 (0·056)	25·60 (0·004)	26·83 (0·030)

Table 5. *Autocovariance and autocorrelation estimates for the Campbelltown, Sydney asthma data*

lag h	$\hat{\gamma}_Z(h)$	SE $\{\hat{\gamma}_Z(h)\}$	$\hat{\gamma}_{Z,UB}(h)$	$\hat{\rho}_Z(h)$	$\hat{\rho}_{Z,UB}(h)$	SE $\{\hat{\rho}_{Z,UB}(h)\}$	$\hat{\rho}_P(h)$
0	0·054	0·024	0·067	1·0	1·0		1·0
1	0·041	0·014	0·053	0·76	0·79	0·209	0·047
2	0·030	0·015	0·041	0·56	0·62	0·224	0·021
3	0·038	0·015	0·050	0·71	0·74	0·224	0·055
4	0·023	0·015	0·033	0·42	0·50	0·224	0·033
5	0·025	0·015	0·036	0·47	0·54	0·224	0·026
6	0·020	0·015	0·030	0·37	0·45	0·224	0·025
7	0·046	0·014	0·057	0·85	0·85	0·209	0·072
8	0·024	0·015	0·033	0·44	0·50	0·224	0·035

The t-ratios for the coefficient of humidity for the two estimators of standard error in Table 3 are 4·41 and 3·19, respectively. It is also clear from Table 3 that the effect of lagged seven-day average humidity is highly significant when we use the proper standard error based on the asymptotic covariance results of Theorem 1.

The statistic S_a has an observed value of 3·30, which is highly significant and clearly indicates the presence of a latent process. Table 4 reports values of test statistics for autocorrelation in the latent process. The modified Ljung–Box statistic based on Pearson residuals indicates significant autocorrelation. The portmanteau test based on bias

adjusted autocorrelations, $\hat{\rho}_{Z,UB}(h)$, provides stronger evidence of serial correlation in the latent process than does the modified Ljung–Box test based on Pearson residuals.

Table 5 provides details of various autocovariance and autocorrelation estimates. Based on the estimates $\hat{\rho}_Z$, we would conclude that the autocorrelations are significant at lags 1, 2, 3 and 7 days. The autocorrelation estimates using either $\hat{\rho}_Z$ or $\hat{\rho}_{Z,UB}$ demonstrate the need for an autoregressive latent process with nonzero coefficients at lags 1, 2, 3 and 7. The autocorrelation function based on the Pearson residuals is completely misleading. If we use the calculation in (16), the expected value of $\hat{\rho}_P(1)$ is

$$E\{\hat{\rho}_P(1)\} \doteq \frac{1{\cdot}934}{(0{\cdot}054)^{-1}+1{\cdot}939}(0{\cdot}76) = 0{\cdot}0718,$$

which explains the small observed values of the $\hat{\rho}_P$ in Table 5.

Acknowledgement

We gratefully acknowledge the assistance of Louise Helby for providing the daily asthma admissions data at Sydney hospitals and of Elly Spark for providing the meteorological data used in the analysis of the asthma data. The research of R. A. Davis was supported in part by the National Science Foundation.

Appendix

Proof of Theorem 1

Here we provide a sketch of the proof in the case when the regression variables form a triangular array with $x_{nt} = f(t/n)$, and f is a continuous function; see § 2·2. If $u = n^{\frac{1}{2}}(\beta - \beta_0)$, maximising $l(\beta)$ is equivalent to minimising

$$g_n(u) := -\{l(\beta) - l(\beta_0)\} = \sum_{t=1}^{n} e^{x_{nt}^{\mathrm{T}}\beta_0}(e^{x_{nt}^{\mathrm{T}}un^{1/2}} - 1) - \sum_{t=1}^{n} Y_t x_{nt}^{\mathrm{T}} u/n^{1/2},$$

where $\beta = u/n^{1/2} + \beta_0$. If $\hat{u}_n$ minimises g_n, then $\hat{u}_n = n^{1/2}(\hat{\beta} - \beta_0)$ and it suffices to show that $\hat{u}_n \to \hat{u}$ in distribution, where $\hat{u} \sim N(0, \Omega_I^{-1} + \Omega_I^{-1}\Omega_{II}\Omega_I^{-1})$.

Write

$$g_n(u) = B_n(u) - A_n(u),$$

where

$$B_n(u) := \sum_{t=1}^{n} e^{x_{nt}^{\mathrm{T}}\beta_0}(e^{x_{nt}^{\mathrm{T}}u/n^{1/2}} - 1 - x_{nt}^{\mathrm{T}}un^{1/2}),$$

$$A_n(u) := \frac{1}{n^{1/2}} \sum_{t=1}^{n} (Y_t - e^{x_{nt}^{\mathrm{T}}\beta_0}) x_{nt}^{\mathrm{T}} u.$$

Note that $g_n(u)$ is a convex function of u and so we can apply a standard result for functional limit theorems to establish that $\hat{u}_n \to \hat{u}$ in distribution.

First note that

$$B_n(u) = \sum_{t=1}^{n} e^{x_{nt}^{\mathrm{T}}\beta_0} \frac{(x_{nt}^{\mathrm{T}}un^{1/2})^2}{2} + E_n(u) = \frac{1}{2} u^{\mathrm{T}} \left(\frac{1}{n} \sum_{t=1}^{n} x_{nt} x_{nt}^{\mathrm{T}} e^{x_{nt}^{\mathrm{T}}\beta_0} \right) u + E_n(u).$$

It follows from the assumptions that

$$B_n(u) \to \frac{1}{2} u^{\mathrm{T}} \Omega_I u$$

in probability for any fixed u.

Next consider

$$A_n(u) = U_n^{\mathrm{T}} u,$$

where

$$U_n := \frac{1}{n^{1/2}} \sum_{t=1}^{n} (Y_t - e^{x_{nt}^{\mathrm{T}} \beta_0}) x_{nt}.$$

If we adapt proofs of central limit theorems for mixing sequences or for linear processes, it is straightforward to show that

$$U_n \to N(0, \Omega_I + \Omega_{II}),$$

in distribution. Since g_n has convex sample paths, we conclude that

$$g_n(u) \to g(u) = \tfrac{1}{2} u^{\mathrm{T}} \Omega_I u - N(0, \Omega_I + \Omega_{II})$$

in distribution on the space $C(\mathbb{R}^p)$; see Rockafellar (1970) and Pollard (1991) for the theoretical results required for this proof. In view of this convergence and the convexity of g_n, it now follows that $\hat{u}_n$ converges in distribution to the minimiser of g, which is easily seen to be

$$\hat{u} = \Omega_I^{-1} N(0, \Omega_I + \Omega_{II}) \sim N(0, \Omega_I^{-1} + \Omega_I^{-1} \Omega_{II} \Omega_I^{-1}).$$

This completes the proof of the result. □

References

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

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.

Chan, K. S. & Ledolter, J. (1995). Monte Carlo EM estimation for time series models involving counts. *J. Am. Statist. Assoc.* **90**, 242–52.

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

Dean, C. & Lawless J. F. (1989). Tests for detecting overdispersion in Poisson regression models. *J. Am. Statist. Assoc.* **84**, 467–72.

Lo, A. W. & MacKinlay A. C. (1989). The size and power of the variance ratio test in finite samples: a Monte Carlo investigation. *J. Economet.* **40**, 203–38.

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

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**, 621–9.

[*Received August* 1998. *Revised February* 2000]