BIOMETRICS 44, 1019–1031
December 1988

Markov Regression Models for Time Series: A Quasi-Likelihood Approach

Scott L. Zeger and Bahjat Qaqish
Department of Biostatistics, Johns Hopkins University,
Baltimore, Maryland 21205, U.S.A.

SUMMARY

This paper discusses a quasi-likelihood (QL) approach to regression analysis with time series data. We consider a class of Markov models, referred to by Cox (1981, *Scandinavian Journal of Statistics* **8,** 93–115) as "observation-driven" models in which the conditional means and variances given the past are explicit functions of past outcomes. The class includes autoregressive and Markov chain models for continuous and categorical observations as well as models for counts (e.g., Poisson) and continuous outcomes with constant coefficient of variation (e.g., gamma). We focus on Poisson and gamma data for illustration. Analogous to QL for independent observations, large-sample properties of the regression coefficients depend only on correct specification of the first conditional moment.

1. Introduction

Generalized linear models (McCullagh and Nelder, 1983) and quasi-likelihood (Wedderburn, 1974; McCullagh, 1983; Nelder and Pregibon, 1986) have recently unified regression methods for a variety of discrete and continuous outcomes. This paper considers the quasi-likelihood (QL) analysis of time series data $(\mathbf{x}_t, y_t)$. The dependence of y_t on $\mathbf{x}_t$ is assumed to be of primary interest. However, as serial observations are unlikely to be independent, we propose Markov models in which the expected response at a given time depends not only on the associated covariates but also on past outcomes.

While linear regression methods for Gaussian time series have been studied in detail (e.g., Anderson, 1954; Tsay, 1984), models for non-Gaussian data have received less attention. For binary time series, Cox (1970) proposed a Markov chain that is an extension of logistic regression. Korn and Whittemore (1979) applied this model to panel data. Kalbfleisch and Lawless (1985) and references therein discussed Markov models for regression with categorical outcomes again for panel data. Kaufmann (1987) presented asymptotic theory for these models. A recent paper by West, Harrison, and Migon (1985) discussed regression methods with response variables from the exponential family. Their method is an extension of Kalman filtering; potentially crude approximations are currently needed, however, to compute posterior distributions of the regression coefficients.

Several investigators have developed models for time dependence in non-Gaussian data which as yet are not applicable to the regression problem. For example, Keenan (1982) discussed time dependence in binary series. Azzalini (1982) and references therein considered Markov models for gamma random variables. Lawrance and Lewis (1985) discussed time series models with exponential marginal distributions.

This paper proposes a quasi-likelihood approach to regression with time series data. We consider Markov models that Cox (1981) refers to as being "observation-driven." Here, the

Key words: Generalized linear models; Markov models; Quasi-likelihood; Regression; Time series.

time dependence arises because the conditional expectation of the outcome given the past depends explicitly on past values. For Gaussian and categorical outcomes, this approach corresponds to the Markov models in common use, yet the method naturally extends to other situations. We focus on Poisson and gamma data for illustration. While we discuss specific conditional distributions, the techniques described require specification of only the first and second conditional moments and in this sense are analogues of quasi-likelihood. Godambe (1985) has discussed finite-sample optimality properties of related estimators.

Section 2 describes the observation-driven Markov processes. Estimation is discussed in Section 3. Section 4 presents an analysis of neuron interspike intervals as an example of this approach. Properties of the marginal distributions are given in Section 5.

2. Quasi-Likelihood Markov Models

2.1 *General*

To establish notation, let y_t be an outcome random variable and $\mathbf{x}_t$ an $m \times 1$ vector of covariates for $t = -p + 1, \ldots, 0, 1, \ldots, n$. Let $\mathbf{D}_t$ be the present and past covariates and past observations at time t, i.e.,

$$\mathbf{D}_t = \{\mathbf{x}_t, \mathbf{x}_{t-1}, \ldots, \mathbf{x}_{-p+1}, y_{t-1}, y_{t-2}, \ldots, y_{-p+1}\}.$$

Define

$$\mu_t = \mathrm{E}(y_t \mid \mathbf{D}_t) \qquad \text{and} \qquad v_t = \mathrm{var}(y_t \mid \mathbf{D}_t).$$

We assume

$$h(\mu_t) = \mathbf{x}_t'\boldsymbol{\beta} + \sum_{i=1}^{q} \theta_i f_i(\mathbf{D}_t), \tag{2.1a}$$

where h is a "link" function, f_i's are functions of the past outcomes, and the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)'$ are to be estimated. As in QL, we further assume that

$$v_t = \mathrm{var}(y_t \mid \mathbf{D}_t) = g(\mu_t) \cdot \phi, \tag{2.1b}$$

where g is a variance function and ϕ is an unknown scale parameter. This formulation is identical to that of QL for independent data (McCullagh and Nelder, 1983), except that conditional rather than marginal moments are modelled.

"Observation-driven" models are in contrast to "parameter-driven" models (Cox, 1981). For the latter, correlation arises from an unobservable underlying process that affects y_t rather than from explicit dependence of y_t on past outcomes. An example is given in Section 2.2. Otherwise, this paper focuses on observation-driven models; parameter-driven models are discussed in detail by West et al. (1985).

2.2 *Examples*

(i) *Gaussian outcomes*: If y_t follows a Gaussian distribution, letting $h(\mu_t) = \mu_t$, $g(\mu_t) = 1$, and $f_i(\mathbf{D}_t) = y_{t-i} - \mathbf{x}_{t-i}'\boldsymbol{\beta}$ gives the autoregressive model of order q (Box and Jenkins, 1976). Alternatively, when $q = 1$, if

$$g(\mu_t) = \left(\frac{\mu_t - \mathbf{x}_t'\boldsymbol{\beta}}{\theta_1}\right)^2 = (y_{t-1} - \mathbf{x}_{t-1}'\boldsymbol{\beta})^2,$$

we have an autoregressive conditionally heteroscedastic (ARCH) model discussed by Engle (1982). ARCH models can account for periods of increased variability as found, for example, in market price series.

(ii) *Binary outcomes*: Letting $f_i(\mathbf{D}_t) = y_{t-i}$ and using a logit link gives

$$\text{logit}(\mu_t) = \mathbf{x}_t'\boldsymbol{\beta} + \theta_1 y_{t-1} + \cdots + \theta_q y_{t-q}.$$

Here $g(\mu_t) = \mu_t(1 - \mu_t)$ and $\phi = 1$. This is the Markov chain of order q suggested by Cox (1970). Here $\exp(\theta_i)$ is the odds of a positive response at time t given $y_{t-i} = 1$ relative to the odds when $y_{t-i} = 0$ with all other components of $\mathbf{D}_t$ being held fixed. When $\theta_i = 0$ $(i = 1, \ldots, q)$, it reduces to the logistic regression model. At the other extreme, a saturated parametrization of the transition matrix can be obtained by letting $q = 2^p - 1$ and defining the f_i's to be all possible interactions of $y_{t-1}, \ldots, y_{t-p}$. For $p = 2$, we have $f_1 = y_{t-1}$, $f_2 = y_{t-2}$, $f_3 = y_{t-1}y_{t-2}$. Note that multivariate logistic regression for categorical outcomes can be similarly extended to obtain Markov models for categorical time series.

(iii) *Counts*: There appear to be a number of reasonable specifications in this case. For example, Wong (1986) considers a model in which $y_t \mid \mathbf{D}_t$ is assumed to be Poisson with

$$\mu_t = \mathrm{E}(y_t \mid y_{t-1}) = \mu[1 + \exp(-\theta_0 - \theta_1 y_{t-1})], \quad \theta_i > 0.$$

Here μ_t is constrained to lie in the interval $[\mu, \mu(1 + e^{-\theta_0})]$ so that the past outcome can alter the current expectation by a factor of at most 2. This constraint facilitates establishing the ergodicity for this process, Wong's focus. We have considered two variations of an alternative model without constraint. In the first, we let

$$\log(\mu_t) = \mathbf{x}_t'\boldsymbol{\beta} + \sum_{i=1}^{q} \theta_i[\log(y_{t-i}^*) - \mathbf{x}_{t-i}'\boldsymbol{\beta}] \tag{2.2}$$

and $v_t = \mu_t \cdot \phi$, where $y_{t-i}^* = \max(y_{t-i}, c)$, $0 < c < 1$. For $q = 1$, this gives

$$\mu_t = \exp(\mathbf{x}_t'\boldsymbol{\beta})\left[\frac{y_{t-1}^*}{\exp(\mathbf{x}_{t-1}'\boldsymbol{\beta})}\right]^{\theta_1}. \tag{2.3}$$

The expected value at time t is modified by the outcome at $t - 1$ relative to $\exp(\mathbf{x}_{t-1}'\boldsymbol{\beta})$. Positive values of θ_1 represent positive autocorrelation; negative values represent negative autocorrelation. The parameter c determines the probability that $y_t > 0$ given $y_{t-1} = 0$. If $c = 0$, the outcome $y_t = 0$ is an absorbing state. Figure 1 displays sample paths of a Poisson time series satisfying (2.3) for a few values of θ_1.

In the alternative, we let

$$\log(\mu_t) = \mathbf{x}_t'\boldsymbol{\beta} + \sum_{i=1}^{q} \theta_i\{\log(y_{t-i} + c) - \log[\exp(\mathbf{x}_{t-i}'\boldsymbol{\beta}) + c]\} \tag{2.4}$$

so that when $q = 1$,

$$\mu_t = \exp(\mathbf{x}_t'\boldsymbol{\beta})\left[\frac{y_{t-1} + c}{\exp(\mathbf{x}_{t-1}'\boldsymbol{\beta}) + c}\right]^{\theta_1}. \tag{2.5}$$

Here c has the interpretation of an "immigration" rate as it is added to each observation rather than only to 0 outcomes.

These models for counts arise in the context of size-dependent branching processes. To illustrate, we consider the model in (2.3) and for simplicity will suppose $\mu_t = \mu$. Suppose y_t represents the number of individuals in a population at generation t. Let $Z_j(y_{t-1})$ be the

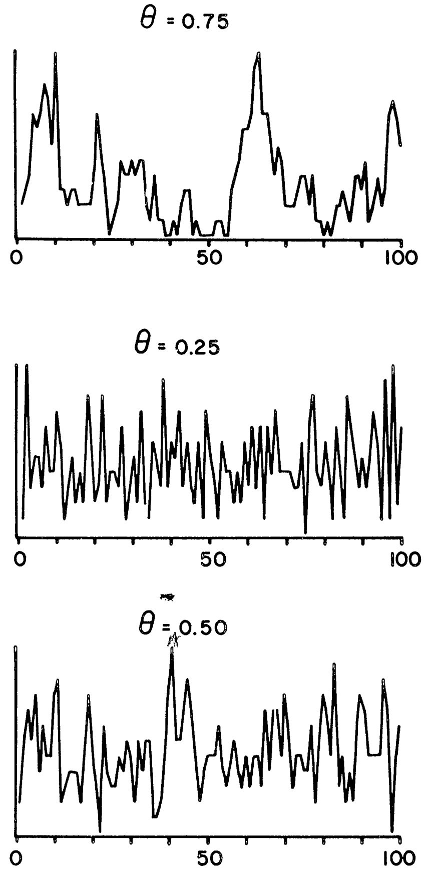

Figure 1. Realizations of conditionally Poisson time series satisfying $\log E(y_t \mid y_{t-1}) = \mu + \theta \log(y^*_{t-1}) - \mu)$, $y^*_{t-1} = \max(y_{t-1}, c)$, $\mu = 5$, $c = .5$, and $\theta = -.5, .25, .75$.

number of offspring of the jth individual in generation $t - 1$. Then for $y_{t-1} > 0$ the total size of the tth generation is

$$y_t = \sum_{j=1}^{y_{t-1}} Z_j(y_{t-1}), \quad y_{t-1} > 0.$$

If $y_{t-1} = 0$, we will assume the population is restarted by Z_0 individuals. Now if we assume $Z_j(y_{t-1})$ is Poisson with $E(Z_j \mid y_{t-1}) = (\mu/y_{t-1})^{1-\theta_1}$, $y_{t-1} > 0$, and $E(Z_0) = (\mu/c)^{1-\theta_1}$, then $E(y_t \mid y_{t-1})$ is as defined in (2.3). For $0 < \theta_1 < 1$, this model describes a crowding effect whereby the expected number of offspring per individual is reduced as the current population increases.

This connection to branching processes can also be used to establish the conditions under which y_t is ergodic. For example, with $\mathbf{x}_t'\boldsymbol{\beta} = \mu$ and $q = 1$, y_t is ergodic for all $\mu > 0$ when $\theta_1 < 1$. Note that $\theta_1 = 1$ gives the Galton–Watson process. Conditions for ergodicity can also be established for this and the other observation-driven processes using standard results from Markov chain theory. See, for example, Tweedie (1976).

An alternative choice for $f_i(\mathbf{D}_t)$ is $f_i(\mathbf{D}_t) = y_{t-i}$ as was used in the logistic model in Example (ii). When $q = 1$, we then have $E(y_t \mid y_{t-1}) = \exp(\mathbf{x}_t'\boldsymbol{\beta})\exp(\theta_1 y_{t-1})$. We prefer $f_i(\mathbf{D}_t) = \log(y^*_{t-i}) - \mathbf{x}_{t-i}'\boldsymbol{\beta}$ as in (2.2) for the following reasons. First, in (2.2),

$E(y_t) \approx \exp(\mathbf{x}_t'\boldsymbol{\beta})$ so that β has interpretation as the proportional change in the marginal expectation of y_t per unit change in $\mathbf{x}_t$. With $f_i(\mathbf{D}_t) = y_{t-i}$, $\exp(\mathbf{x}_t'\boldsymbol{\beta})$ is the rate given the previous outcome is 0. Second, when $\mathbf{x}_t'\boldsymbol{\beta} = \beta_0$ for all t, the alternative model leads to a stationary process only when $\theta_1 \leqslant 0$. Positive association is not possible. Finally, under (2.2), $\hat{\boldsymbol{\beta}}$ and $\hat{\theta}_1$ are approximately orthogonal, making estimation easier.

(iv) *Outcomes with constant coefficient of variation*: The canonical link (McCullagh and Nelder, 1983) for the gamma distribution, which has a constant coefficient of variation, is the inverse function. Hence, we have considered the model

$$\mu_t^{-1} = \mathbf{x}_t'\boldsymbol{\beta} + \sum_{i=1}^{q} \theta_i\left(\frac{1}{y_{t-i}^*} - \mathbf{x}_{t-i}'\boldsymbol{\beta}\right) \tag{2.6}$$

and $v_t = \mu^2 \cdot \phi$. Section 4 discusses this model in detail. Figure 2 shows sample paths of a first-order time series satisfying (2.6) for different values of θ_1. Note the tendency for long troughs and sharp peaks, which is consistent with the conditional variance being proportional to μ_t^2. Figure 3 shows sample paths for a conditionally gamma process with $\theta_1 = .8$ and varying shape parameter ϕ. Note that the process becomes increasingly Gaussian with increasing ϕ.

(v) *Parameter-driven model*: To contrast the observation-driven models in Examples (i)–(iv), we briefly introduce a parameter-driven log-linear model. In parameter-driven models, autocorrelation is introduced through a latent process. Let $\theta_t = \log(\mu_t)$ be the canonical parameter for the log-linear model. Then, in the parameter-driven case, $\theta_t = \theta(\varepsilon_t)$ where ε_t is an unobservable noise process, perhaps representing unmeasured influences on y_t. For example, if y_t given ε_t is Poisson with $E(y_t \mid \varepsilon_t) = \exp(\mathbf{x}_t'\boldsymbol{\beta} + \varepsilon_t)$ and ε_t is a Gaussian autoregressive process of order 1, then y_t will be an autocorrelated series due to the influence of ε_t. Note that $E(y_t \mid \mathbf{D}_t)$ is in general not a simple function of past y_t's as is the case in observation-driven models.

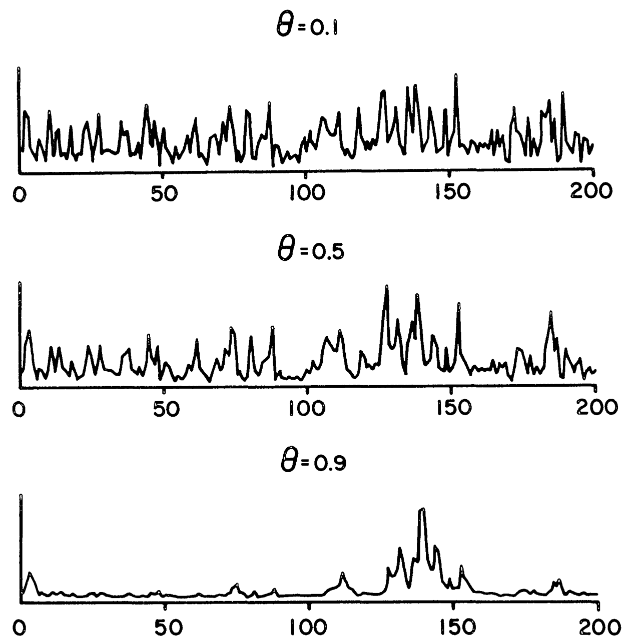

Figure 2. Realizations of conditionally gamma process with $E(y_t \mid y_{t-1})^{-1} = \mu + \theta(y_{t-1}^{*-1} - \mu)$, $y_{t-1}^* = \max(y_{t-1}, c)$, $\mu = 1$, $c = .05$, and $\theta = .1, .5, .9$. The shape parameter is $\phi = 3$.

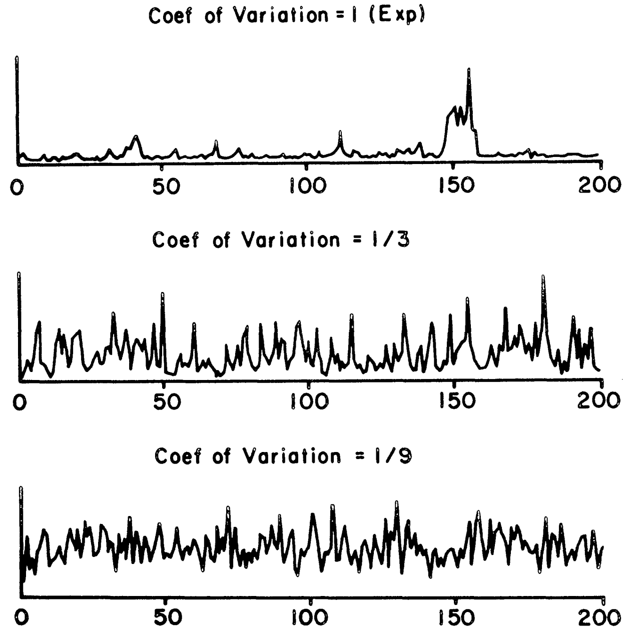

Figure 3. Realizations as in Figure 2 with $\mu = 1$, $c = .05$, $\theta = .8$, and shape parameter $\phi = 1$ (exponential), 3, 9.

3. Estimation

To begin, we assume the parameter c is either unnecessary or known. Let $\boldsymbol{\gamma}' = (\boldsymbol{\beta}', \boldsymbol{\theta}')$. These parameters are to be estimated by a quasi-likelihood (QL) approach (McCullagh and Nelder, 1983) so that $\boldsymbol{\gamma}$ is the root of the log-QL estimating equation

$$U(\boldsymbol{\gamma}) = \sum_{t=1}^{n} \frac{\partial \mu_t}{\partial \boldsymbol{\gamma}} v_t^{-1}(y_t - \mu_t),$$

which, with canonical link, reduces to (3.1)

$$= \sum_{t=1}^{n} \mathbf{Z}_t(y_t - \mu_t) = \mathbf{0},$$

where $\mathbf{Z}_t' = (\mathbf{x}_t', f_1(\mathbf{D}_t), \ldots, f_q(\mathbf{D}_t))$. Equation (3.1) can be solved by iteratively reweighted least squares using, for example, GLIM (Baker and Nelder, 1978) in the case when the $f_i(\mathbf{D}_t)$ do not depend on unknown parameters. An example is the Markov chain for binary data in Section 2.2. For Examples (iii) and (iv), $f(\mathbf{D}_t)$ depends on $\boldsymbol{\beta}$ and a second level of iteration is therefore necessary. Note in these cases that

$$h(\mu_t) = \tilde{\mathbf{x}}_t' \boldsymbol{\beta} + \theta_1 h(y_{t-1}^*) + \cdots + \theta_q h(y_{t-q}^*), \tag{3.2}$$

where $\tilde{\mathbf{x}}_t = \mathbf{x}_t - \theta_1 \mathbf{x}_{t-1} - \cdots - \theta_q \mathbf{x}_{t-q}$. That is, we can fit this model by regressing y_t on $\tilde{\mathbf{x}}_t$—a time-filtered version of $\mathbf{x}_t$—and on $h(y_{t-i}^*)$. The following iteration can be used:

(i) Given $\hat{\theta}^{(k)}$, calculate $\tilde{\mathbf{x}}_t^{(k)}$ and let $\mathbf{D}_t = (h(y_{t-1}^*), \ldots, h(y_{t-q}^*))$.
(ii) Estimate $\hat{\boldsymbol{\beta}}^{(k+1)}$, the coefficients for $\tilde{\mathbf{x}}_t^{(k)}$, and $\hat{\theta}^{(k+1)}$, the coefficients for $\mathbf{D}_t$, using iteratively reweighted least squares.
(iii) Repeat steps (i) and (ii) to convergence of parameter estimates (or deviance).

Now consider a case such as the log-linear example where the constant c is an unknown parameter to be estimated. The parameter c is identifiable only if $\theta_i \neq 0$ for some i. Redefine $\boldsymbol{\gamma}' = (\boldsymbol{\beta}', \boldsymbol{\theta}', c)$. The estimating equation for c is

$$U_c(\boldsymbol{\gamma}) = \sum_{t=1}^{n} w_t(c)[y_t - \mu_t(c)] = 0, \tag{3.3}$$

where $w_t(c) = \sum_{i=1}^{q} \theta_i I(y_{t-i} < c)$ and $I(\cdot)$ is the indicator function. If we redefine

$$\mathbf{Z}_t' = (\tilde{\mathbf{x}}_t', f_1(\mathbf{D}_t), \ldots, f_q(\mathbf{D}_t), w_t(c)),$$

(3.3) is the last equation in the system given by (3.1). To simultaneously estimate c, the previous algorithm requires one additional step after (ii) in which (3.3) is solved for $\hat{c}^{(k+1)}$ given $\hat{\boldsymbol{\beta}}^{(k+1)}$ and $\hat{\theta}^{(k+1)}$.

We now have the following result on the asymptotic distribution of $\hat{\boldsymbol{\gamma}}$.

Proposition Let y_t $(t = -p+1, \ldots, n)$ be a time series satisfying (2.1) for known canonical link and variance function. Let $\hat{\boldsymbol{\gamma}}' = (\hat{\boldsymbol{\beta}}', \hat{\theta}', c)$ be the solution of the QL estimating equation (3.1). Suppose the limiting matrix, $\mathbf{V}$, exists and is positive definite. Then under regularity conditions for the conditional densities of $y_t \mid \mathbf{D}_t$ and conditional on $\mathbf{D}_1$ we have that $\sqrt{n}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})$ converges to a Gaussian random vector with mean $\mathbf{0}$ and variance $\mathbf{V}^{-1}$, where

$$\mathbf{V} = \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \mathbf{Z}_t \mathbf{Z}_t' v_t.$$

The proof follows from Theorem 3 of Kaufmann (1987). The asymptotic normality is established by expanding $U(\boldsymbol{\gamma})$ in a Taylor series about the true $\boldsymbol{\gamma}$. We then have, conditioned on $y_{-p+1}, \ldots, y_0$, that

$$\sqrt{n}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) = \mathbf{V}_n^{-1} U(\boldsymbol{\gamma})/\sqrt{n} + o(1),$$

where

$$\mathbf{V}_n = \sum_{t=1}^{n} \mathbf{Z}_t \mathbf{Z}_t' v_t / n.$$

The result follows by applying a central limit theorem to $U(\boldsymbol{\gamma})$.

The definition of the QL carries over to the time series case again by treating $\mathbf{Z}_t$'s as fixed covariates and conditioning on the first q observations.

The scale parameter ϕ can be estimated from the Pearson residuals, which are defined by

$$r_t = (y_t - \hat{\mu}_t)/g(\hat{\mu}_t)^{1/2}.$$

We use the moment estimator

$$\hat{\phi} = \frac{1}{n} \sum_{t=1}^{n} r_t^2.$$

4. Example

The quasi-likelihood modelling of time-dependent data is illustrated with an analysis of interspike times collected from neurons in the motor cortex of a monkey. Neurons are cells that experience momentary potential changes or "spikes" comprising a point process or "spike train" over time. Information in the brain is encoded in the intensity of the spike train. Figure 4 displays the interspike times for a motor cortex neuron studied

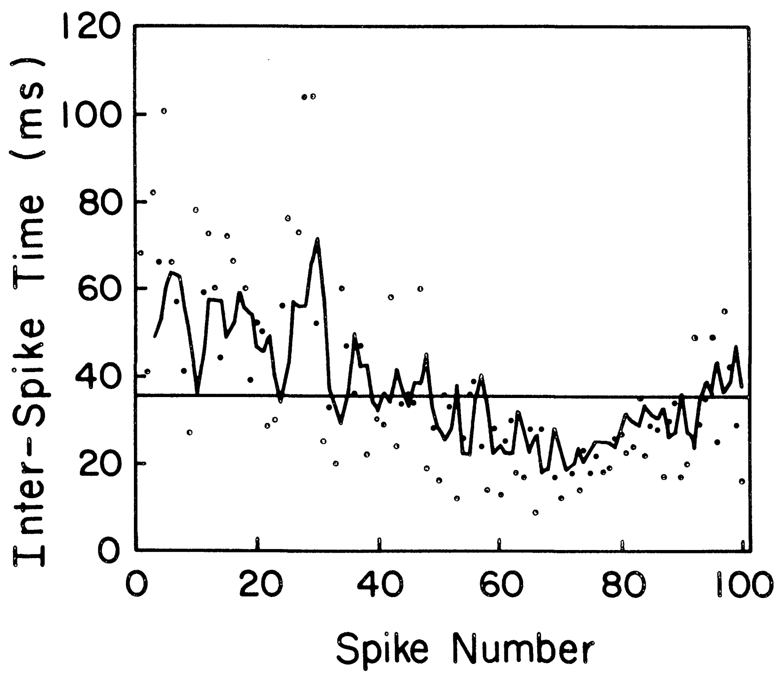

Figure 4. Observed (•) and predicted (———) interspike times from second-order Markov model. Horizontal line is the predicted level from a generalized linear model without time dependence.

by Dr A. Georgopoulos of Johns Hopkins University. The monkey was unstimulated during recording. The objectives of the analysis are to estimate the firing rate prior to stimulation and to characterize the time dependence.

This is clearly not a realization of a Gaussian process. The variance of the time series increases with the level and the observations are skewed to higher values. Note also the low frequency variation indicative of positive autocorrelation. One effective approach in this case is to transform the data to the logarithmic scale and then to fit an autoregressive model. While the log transform makes the Gaussian assumption reasonable here, this is often not the case with spike-train data, even for neurons from the same experimental animal [see Poggio and Viernstein (1963) and references therein]. In a typical experiment, several neurons are studied; using different scales for analysis is not feasible. An alternative is to use the quasi-likelihood model in Example (iv). In this study, we wish to estimate the mean interspike time since the animal was unstimulated during the trial; the degree of dependence among successive interspike times is also of interest. We have chosen an inverse link function and have assumed that the conditional variance is proportional to the square of the conditional mean. The inverse link is desirable because the intercept is then the firing rate in spikes per unit time. In addition, there is evidence in the neuroscience literature that the standard deviation of the interspike times is nearly proportional to the mean (Poggio and Viernstein, 1963). Note that our quasi-likelihood model assumes the conditional, rather than marginal, coefficient of variation (CV) to be constant. However, this assumption implies that the marginal CV is nearly constant as well.

Table 1 lists the interspike times. Table 2 presents the estimated firing rate and parameters for the time dependence with their associated standard errors obtained by fitting a second-order Markov model with $c = 0$. Figure 4 also presents the observed and fitted values from the model. A second-order model was chosen as it appears to account for the observed autocorrelation. Adding additional lags failed to reduce the deviance substantially. Because

Table 1
Listing of motor cortex neuron interspike intervals in msec
(*Read across rows.*)

68	41	82	66	101	66	57	41	27	78
59	73	6	44	72	66	59	60	39	52
50	29	30	56	76	55	73	104	104	52
25	33	20	60	47	6	47	22	35	30
29	58	24	34	36	34	6	19	28	16
36	33	12	26	36	39	24	14	28	13
2	30	18	17	28	9	28	20	17	12
19	18	14	23	18	22	18	19	26	27
23	24	35	22	29	28	17	30	34	17
20	49	29	35	49	25	55	42	29	16

Source: Dr A. Georgopoulos, Johns Hopkins University Department of Neuroscience.

Table 2
Modelling results for data listed in Table 1;
Model A: *Intercept only*; *Model B*: *Intercept and spike number*

	Model A			Model B		
Variable	Estimate	St. error	t	Estimate	St. error	t
Intercept	.0217	.00264	8.22	.0268	.00199	13.5
Spike number				.000263	.0000494	5.32
Lag-1	.336	.0845	3.97	.218	.0853	2.56
Lag-2	.283	.0804	3.51	.146	.0837	1.74
Deviance	17.20			15.47		
df	96			95		
Scale parameter	.174			.160		

we expect the residuals to be non-Gaussian, standard tests for autocorrelation among the residuals are not appropriate. We have generated 95% test regions under the null model of no autocorrelation by: scrambling the order of the residuals; calculating the autocorrelation function of the randomly reordered data; repeating this process a large number of times. The empirical distribution of correlations at each lag can be used to obtain an approximate 95% null interval. In our example, the estimated correlation function lies within the 95% limits at each lag.

The 95% confidence interval for β_0, determined from the asymptotic distribution of $\hat{\gamma}$, is (16.1, 26.9) spikes per second. In our conditional model, $\hat{\beta}_0$ is approximately the expected firing rate given that the previous two intervals were of typical length. An alternative statistic is

$$\hat{\alpha} = \frac{1}{n} \sum_t \hat{\mu}_t^{-1} = \frac{1}{n}\left[\hat{\beta}_0 + \hat{\theta}_1 \sum_t \left(\frac{1}{y_{t-1}} - \hat{\beta}_0\right) + \hat{\theta}_2 \sum_t \left(\frac{1}{y_{t-2}} - \hat{\beta}_0\right)\right],$$

which is the average expected firing rate over this realization of about 100 spikes. A 95% interval obtained using a delta-method approximation for $\hat{\alpha}$ is (16.6, 26.4) spikes per second, which is similar to the interval for $\hat{\beta}_0$. The analogous interval when the correlation is ignored is only half as wide.

In this example, there is a priori no reason to expect a trend in the firing rate since the monkey was not being stimulated. For illustration, however, we have refit the model including spike sequence number as a covariate. Its coefficient is the change in the firing rate with each additional spike. The results of fitting a second-order model are also in

Table 2. It is apparent in Figure 4 that there is a decreasing trend in interspike interval length and hence increasing firing rate. The 95% interval for $\hat{\beta}_2$ is (1.64, 3.62) $\times$ 10^{-4}, corresponding to an increase in firing rate in 100 spikes of between 16.4 and 36.2 spikes per second. Note that the coefficients for spike length at times $t - 1$ and $t - 2$ decrease when a linear trend is included. This is because longer-term fluctuations previously taken into account by the autoregressive terms are now attributed to the systematic linear trend.

5. Marginal Distributions

Here, we briefly discuss properties of the marginal distributions of the first-order Markov processes above for the case $\mathbf{x}_t'\boldsymbol{\beta} = \mu$. The marginal distributions for the Gaussian and binary cases are well known; we focus on the Poisson and gamma cases.

For the first-order Poisson Markov chain, the transition matrix has the form

$$P_{ij} = \Pr(y_t = j \mid y_{t-1} = i) = \frac{\exp(-\lambda_i)\lambda_i^j}{j!},$$

where $\lambda_i = \mu(i^*/\mu)^\theta$ and $i^* = \max(i, c)$. The limiting stationary distribution, Π, is the left eigenvector of the transition matrix and satisfies $\Pi = \Pi\mathbf{P}$. We have not found an explicit solution to this equation. We have numerically estimated Π by powering the transition matrix, however. Table 3 presents the mean, variance, and first-lag autocorrelation of the stationary distribution for $\mu = 2, 5$; $c = .1, .5$; and for a variety of values of θ.

Table 3
Marginal means, variances, and first-lag autocorrelation as a function of μ, c, and θ for a conditionally Poisson process satisfying equation (2.5)

		θ					
μ	c	$-.4$	$-.2$	0	.2	.4	.8
			Mean				
2	.1	2.59	2.18	2.00	1.84	1.65	1.21
	.5	2.19	2.08	2.00	1.93	1.87	1.96
5	.1	5.43	5.15	5.00	4.86	4.67	3.31
	.5	5.33	5.13	5.00	4.88	4.72	4.10
			Variance				
2	.1	5.48	2.54	2.00	1.99	2.17	3.18
	.5	2.63	2.17	2.00	2.00	2.16	3.84
5	.1	10.19	5.68	5.00	5.17	5.97	11.49
	.5	7.41	5.49	5.00	5.14	5.83	11.14
			Autocorrelation				
2	.1	$-.47$	$-.29$	.00	.25	.47	.78
	.5	$-.37$	$-.19$	.00	.18	.36	.70
5	.1	$-.41$	$-.24$	.00	.23	.45	.84
	.5	$-.44$	$-.23$	.00	.22	.43	.79

As θ increases, the stationary marginal mean decreases. The marginal variance increases with θ as it does in the Gaussian case but not with the same functional dependence, $1/(1 - \theta^2)$. Sample paths tend to linger at smaller values. This is because the conditional variance is proportional to the mean. The pointed peaks and flat valleys in Figure 1 illustrate this fact. It is also apparent in the marginal distributions, which are increasingly squeezed toward 0 on the left and have longer tails on the right as θ increases. The marginal first-lag correlation is similar to θ, with the similarity increasing as the mean increases and the process becomes more nearly Gaussian. The autocorrelation function, although not shown, decreases more slowly than the geometric decrease found for Gaussian autoregressive time series of order 1. Finally, note that the choice of c has little impact on the marginal moments even in the case $\mu = 5$. This is desirable so that estimation of μ will not depend strongly on estimation of c.

The analogous properties of the marginal distribution for the first-order gamma Markov chains are presented in Table 4. These were determined in a small Monte Carlo study in which 2,000 realizations of length 50 were generated. The process was found to be approximately stationary by time 30 in each case. Hence, the moments were estimated by the average over times 40 to 50 and over all realizations. We have considered gamma processes with mean 1 and shape parameters: 1 (exponential) and 5 with coefficients of variation of 1 and $\frac{1}{5}$, respectively. For the exponential case, we let $c = .05$ and $.10$ with probabilities of .049 and .095 of an observation falling below c given the previous value equalled μ.

Table 4
Marginal means, variances, and first-lag autocorrelation as a function of shape parameter ϕ, c, and θ for conditionally gamma processes satisfying equation (2.6)

		θ				
ϕ	c	0	.2	.4	.6	.8
			Mean			
1	.05	1.00	.74	.56	.41	.28
	.10	1.00	.78	.63	.50	.40
5	.05	1.00	.95	.90	.81	.59
			Variance			
1	.05	1.00	.71	.60	.51	.45
	.10	1.00	.73	.64	.60	.63
5	.05	.20	.19	.21	.26	.33
			Autocorrelation			
1	.05	.00	.26	.42	.52	.61
	.10	.00	.24	.39	.49	.57
5	.05	.00	.22	.43	.62	.80

The marginal mean for the gamma process decreased with θ, more dramatically than for the Poisson case. The gamma conditional variance is proportional to the square of the mean and hence much longer periods of small values are possible. The variance is relatively stable as a function of θ. For the exponential case, the first-lag autocorrelation is higher than θ for small θ and smaller for θ greater than about .5. With shape parameter 5, however, the marginal first-lag autocorrelation is approximately equal to θ. The effect of the time dependence is to squeeze the marginal distributions toward smaller values. This again reflects the tendency for the sample paths to remain at small values, and to return quickly after excursions to larger values.

6. Discussion

Quasi-likelihood estimators with independent observations have the desirable property of being consistent given correct specification of the mean structure and regularity conditions. If a robust variance estimate is used (Royall, 1986), inferences about the regression parameters are also consistent as long as the mean structure is correct. In the time series case, consistency is assured as long as the specification of the conditional mean as a function of the covariates and of the past values is correct. That is, consistent estimation of $\boldsymbol{\beta}$ will in general require proper modelling of the time dependence. This is unnecessary for autoregressive models with a linear link.

Residual analysis is important for the proposed methods because inferences about the regression parameters depend on the choice of model for the time dependence. When the model is correct, the Pearson residuals are approximately uncorrelated. A simple model check can be made by calculating the autocorrelation function of the residuals. As the distribution of the residuals is unknown, we have used a resampling scheme to obtain a null distribution for comparison as illustrated in the example.

Godambe (1985) has discussed optimality properties of estimators in the context of stochastic processes. His main result is an analogue of the Gauss–Markov theorem. The optimality is based on only the first two moments of the conditional distributions. In cases where the scale parameter ϕ is known [e.g., Example (ii)] the quasi-likelihood estimators proposed here are members of the class of estimators considered by Godambe and hence share their finite-sample optimality properties.

Finally, many questions need to be addressed for effective use of the proposed regression methods. Likelihood-based techniques for choosing the order of the Markov process have been discussed in detail for the Gaussian (e.g., Akaike, 1970) and binary (e.g., Katz, 1981) cases. Generalizations to this quasi-likelihood setting are needed. Adaptation of the data-driven choice of the link and variance functions (Nelder and Pregibon, 1986) and modelling of the mean and variance as separate functions of the covariates (Efron, 1986) have application for time series data as well. Effective techniques for assessing the robustness of inferences about $\boldsymbol{\beta}$ to various aspects of the time-dependent model also require study.

Acknowledgements

The authors thank Dr Subhash Lele for pointing out the connection of the models for counts with branching processes and Dr Apostolos Georgopoulos for use of his data. Thanks also to the editors and referee whose comments led to an improved version. This work was supported by Grant 1-R29-AI25529-01 from the National Institutes of Health.

Résumé

Ce papier traite une approche de quasi-vraisemblance, pour une analyse de régression, avec des données de séries temporelles. Nous considérons une classe de modèles de Markov, rapportée par Cox (1981, *Scandinavian Journal of Statistics* **8,** 93–115), comme modèles "d'observations courantes," dans lesquels les espérances, et variances a posteriori, sont des fonctions explicites des expériences passées. La classe contient aussi bien des modèles autorégressifs, et de chaîne de Markov, pour des observations continues, et en catégories, que des modèles pour des observations de comptage (loi de Poisson), et continues avec un coefficient de variation constant (loi gamma). A titre d'exemple, nous nous focalisons sur des données de Poisson, et de loi gamma. Par analogie avec la quasi-vraisemblance pour des observations indépendantes, les propriétés des grands échantillons, des coefficients de régression, dépend seulement de la spécification correcte du premier moment conditionnel.

References

Akaike, H. (1970). Statistical predictor identification. *Annals of the Institute of Statistical Mathematics* **22,** 203–217.

Anderson, R. L. (1954). The problem of autocorrelated data in regression analysis. *Journal of the American Statistical Association* **49,** 113–129.

Azzalini, A. (1982). Approximate filtering of parameter-driven processes. *Journal of Time Series Analysis* **3,** 219–223.

Baker, R. J. and Nelder, J. A. (1978). *The GLIM System, Release 3*. Oxford: Numerical Algorithms Group.

Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control.* San Francisco: Holden-Day.

Cox, D. R. (1970). *The Analysis of Binary Data.* London: Chapman and Hall.

Cox, D. R. (1981). Statistical analysis of time series: Some recent developments. *Scandinavian Journal of Statistics* **8,** 93–115.

Efron, B. (1986). Double exponential families and their use in generalized linear regression. *Journal of the American Statistical Association* **81,** 709–721.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of United Kingdom inflation. *Econometrica* **50,** 987–1007.

Godambe, V. P. (1985). The foundations of finite-sample estimation in stochastic processes. *Biometrika* **72,** 419–428.

Kalbfleisch, J. D. and Lawless, J. F. (1985). The analysis of panel data under a Markov assumption. *Journal of the American Statistical Association* **80,** 863–871.

Katz, R. W. (1981). On some criteria for estimating the order of a Markov chain. *Technometrics* **23,** 243–249.

Kaufmann, H. (1987). Regression models for nonstationary categorical time series: Asymptotic estimation theory. *Annals of Statistics* **15,** 79–98.

Keenan, D. M. (1982). A time series analysis of binary data. *Journal of the American Statistical Association* **77,** 816–821.

Korn, E. L. and Whittemore, A. S. (1979). Methods for analyzing panel studies of acute health effects of air pollution. *Biometrics* **35,** 795–802.

Lawrance, A. J. and Lewis, P. A. W. (1985). Modelling and residual analysis of nonlinear autoregressive time series in exponential variables. *Journal of the Royal Statistical Society, Series B* **47,** 165–202.

McCullagh, P. (1983). Quasi-likelihood functions. *Annals of Statistics* **11,** 59–67.

McCullagh, P. and Nelder, J. A. (1983). *Generalized Linear Models.* London: Chapman and Hall.

Nelder, J. A. and Pregibon, D. (1986). Quasi-likelihood models and data analysis. *Biometrika* **74,** 221–233.

Poggio, G. F. and Viernstein, L. J. (1963). Time series analysis of impulse sequences of thalamic somatic sensory neurons. *Journal of Neurophysiology* **26,** 517–545.

Royall, R. M. (1986). Model robust inference using maximum likelihood estimators. *International Statistical Review* **54,** 221–226.

Tsay, R. S. (1984). Regression models with time series errors. *Journal of the American Statistical Association* **79,** 118–124.

Tweedie, R. L. (1976). Criteria for classifying general Markov chains. *Advances in Applied Probability* **8,** 737–771.

Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalized linear models and the Gauss–Newton model. *Biometrika* **61,** 439–447.

West, M., Harrison, P. J., and Migon, H. S. (1985). Dynamic generalized linear models and Bayesian forecasting. *Journal of the American Statistical Association* **80,** 73–96.

Wong, W. H. (1986). Theory of partial likelihood. *Annals of Statistics* **14,** 88–123.

Received March 1987; *revised February* 1988.