Maximum Likelihood Estimation for an Observation Driven Model for Poisson Counts

Richard A. Davis[1], William T.M. Dunsmuir[2] and Sarah B. Streett[3]
[1] Department of Statistics, Colorado State University.
[2] Department of Statistics, University of New South Wales.
[3] Geophysical Statistics Project, National Center for Atmospheric Research.

September 24, 2003

Abstract

This paper is concerned with an observation driven model for time series of counts whose conditional distribution given past observations follows a Poisson distribution. This class of models, called GLARMA, is capable of modeling a wide range of dependence structures and is readily estimated using conditional maximum likelihood. Recursive formulae for carrying out maximum likelihood estimation are provided and the technical components required for establishing a central limit theorem of the maximum likelihood estimates are given in a special case.

1 Introduction

In recent years there has been considerable development of models for non-Gaussian time series. A review of models for the special case of time series of counts is contained in Davis et al. (1999). There, a new class of models, which we will refer to as generalized linear autoregressive moving average (GLARMA) models is introduced. These GLARMA models are developed further in Davis et al. (2003) where, for a simple example, ergodicity of the process is established and asymptotic normality of the maximum likelihood estimates is stated. The primary objective of this paper is to provide some of the technical details required to establish asymptotic normality of the maximum likelihood estimate in the first-order GLARMA model. While equivalent results for the fully general GLARMA models are difficult to establish, the proofs are likely to follow the lines of argument for the cases considered here.

To introduce the general version of our model, assume that the observation Y_t given the past history $\mathcal{F}_{t-1} = \sigma(Y_s, s \leq t-1)$ is Poisson with mean μ_t which will be denoted by

$$Y_t|\mathcal{F}_{t-1} \sim P(\mu_t).$$

It is further assumed that the *state* process $\log(\mu_t)$ is a moving average driven by noise that is a martingale difference sequence generated from the data. Formally, the state process is given by

$$W_t := \log(\mu_t) = \beta + \sum_{j=1}^{q} \tau_j(\boldsymbol{\gamma}) e_{t-j}, \tag{1}$$

where $1 \leq q \leq \infty$,

$$e_t = (Y_t - \mu_t)/\mu_t^{\lambda}, \quad \lambda \geq 0$$

is a martingale difference sequence, and $\boldsymbol{\gamma}$ is a parameter vector. Even if $q = 1$, the conditional mean $E(Y_t|\mathbf{Y}^{(t-1)})$ depends on the whole past and hence is not Markov. On the other hand, the mean process $\log(\mu_t)$ is q^{th} order Markov. For $q = 1$, it was shown in Proposition 2 of Davis et al. (2003) that the process $\{W_t\}$ has a unique stationary distribution and is uniformly ergodic.

One desirable way in which to parameterize the moving average weights $\tau_j(\boldsymbol{\gamma})$ in (1), is to allow them to be the coefficients in an autoregressive-moving average (ARMA) filter. Specifically, set

$$\tau(z) := \sum_{j=1}^{\infty} \tau_j(\boldsymbol{\gamma}) z^j = \theta(z)/\phi(z) - 1,$$

where $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$ and $\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$ are the respective autoregressive and moving average polynomials of the ARMA filter, each having all their zeros outside the unit circle, and $\boldsymbol{\gamma}$ is the parameter vector consisting of the ϕ_i's and θ_j's. Writing $W_t = \beta + Z_t$, where

$$Z_t = \sum_{j=1}^{\infty} \tau_j(\boldsymbol{\gamma}) e_{t-j}, \tag{2}$$

it follows that $\{Z_t\}$ satisfies the ARMA-like recursions,

$$Z_t = \sum_{i=1}^{p} \phi_i (Z_{t-i} + e_{t-i}) + \sum_{i=1}^{q} \theta_i e_{t-i}. \tag{3}$$

When this condition holds for the $\{W_t\}$ process, the model is referred to as a generalized ARMA or GLARMA (see Davis et al. (2003)).

2 Estimation and Inference for the Model

2.1 Maximum Likelihood Estimation

The likelihood and its first and second derivatives can easily be computed recursively and used in a Newton-Raphson update procedure for the GLARMA model. Standard errors for the parameter estimates that properly account for serial dependence are also readily available. The details follow.

Let $\boldsymbol{\delta} = (\beta, \boldsymbol{\gamma}^T)^T$ and define $L_t(\boldsymbol{\delta}) = \log f(y_t|\mathcal{F}_{t-1})$, where f is the conditional Poisson density of Y_t given $\mathcal{F}_{t-1}$. The log-likelihood can then be written as $\sum_{t=1}^{n} L_t(\boldsymbol{\delta})$ which, upon ignoring terms which do not involve the parameters, becomes

$$L(\boldsymbol{\delta}) = \sum_{t=1}^{n} \left(Y_t W_t(\boldsymbol{\delta}) - e^{W_t(\boldsymbol{\delta})} \right), \tag{4}$$

where

$$\log(\mu_t) = W_t(\boldsymbol{\delta}) = \beta + \sum_{i=1}^{q} \tau_i(\boldsymbol{\gamma}) e_{t-i}(\boldsymbol{\delta})$$

and

$$e_t(\boldsymbol{\delta}) = (Y_t - \mu_t)/\mu_t^{\lambda}.$$

For brevity, we will often suppress the dependence of W_t and e_t on $\boldsymbol{\delta}$. The first and second derivatives of L are given by the following expressions

$$\frac{\partial L}{\partial \boldsymbol{\delta}} = \sum_{t=1}^{n} (Y_t - \mu_t) \frac{\partial W_t}{\partial \boldsymbol{\delta}} = \sum_{t=1}^{n} e_t \mu_t^{\lambda} \frac{\partial W_t}{\partial \boldsymbol{\delta}}$$

and

$$\begin{aligned} \frac{\partial^2 L}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}^T} &= \sum_{t=1}^{n} \left[(Y_t - \mu_t) \frac{\partial^2 W_t}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}^T} - \mu_t \frac{\partial W_t}{\partial \boldsymbol{\delta}} \frac{\partial W_t}{\partial \boldsymbol{\delta}^T} \right] \\ &= \sum_{t=1}^{n} \left[e_t \mu_t^{\lambda} \frac{\partial^2 W_t}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}^T} - \mu_t \frac{\partial W_t}{\partial \boldsymbol{\delta}} \frac{\partial W_t}{\partial \boldsymbol{\delta}^T} \right]. \end{aligned}$$

The remaining recursive expressions needed to calculate these derivatives are given below. Asymptotic results for these estimates are given in Section 2.2 for the case where $\lambda = 1$ and $q = 1$. Under these conditions, the asymptotic distribution of the maximum likelihood estimates is $N(0, V^{-1})$, where

$$V = \lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} e^{W_t} \dot{W}_t \dot{W}_t^T, \tag{5}$$

with $\dot{W}_t = \frac{\partial W_t(\delta)}{\partial \delta}$.

To initialize the Newton Raphson recursions we have found that using the GLM estimates without the autoregressive moving average terms together with zero initial values for e_t, $t \leq 0$, gives reasonable starting values. Convergence in the majority of cases that we have considered occurred within 10 iterations from these starting conditions.

The remaining expressions needed to calculate the derivatives of the likelihood are derived below. These can be readily programmed and implementation in the S-language is available from the second author upon request. First we note that

$$\frac{\partial e_t}{\partial \boldsymbol{\delta}} = -[e^{(1-\lambda)W_t} + \lambda e_t] \frac{\partial W_t}{\partial \boldsymbol{\delta}}.$$

and

$$\frac{\partial W_t}{\partial \boldsymbol{\delta}} = \frac{\beta}{\partial \boldsymbol{\delta}} + \frac{\partial Z_t}{\partial \boldsymbol{\delta}},$$

where $\{Z_t\}$ is defined in (2). From the recursion (3), it follows that

$$\begin{aligned} \frac{\partial Z_t}{\partial \boldsymbol{\delta}} &= \sum_{i=1}^{p} \frac{\partial \phi_i}{\partial \boldsymbol{\delta}} (Z_{t-i} + e_{t-i}) + \sum_{i=1}^{p} \phi_i \left(\frac{\partial Z_{t-i}}{\partial \boldsymbol{\delta}} + \frac{\partial e_{t-i}}{\partial \boldsymbol{\delta}} \right) \\ &+ \sum_{i=1}^{q} \frac{\partial \theta_i}{\partial \boldsymbol{\delta}} e_{t-i} + \sum_{i=1}^{q} \theta_i \frac{\partial e_{t-i}}{\partial \boldsymbol{\delta}}. \end{aligned}$$

In particular:

$$\frac{\partial Z_t}{\partial \beta} = \sum_{i=1}^{p} \phi_i \left(\frac{\partial Z_{t-i}}{\partial \beta} + \frac{\partial e_{t-i}}{\partial \beta} \right) + \sum_{i=1}^{q} \theta_i \frac{\partial e_{t-i}}{\partial \beta},$$

$$\frac{\partial Z_t}{\partial \phi_a} = Z_{t-a} + e_{t-a} + \sum_{i=1}^{p} \phi_i \left(\frac{\partial Z_{t-i}}{\partial \phi_a} + \frac{\partial e_{t-i}}{\partial \phi_a} \right) + \sum_{i=1}^{q} \theta_i \frac{\partial e_{t-i}}{\partial \phi_a}$$

and

$$\frac{\partial Z_t}{\partial \theta_a} = \sum_{i=1}^{p} \phi_i \left(\frac{\partial Z_{t-i}}{\partial \theta_a} + \frac{\partial e_{t-i}}{\partial \theta_a} \right) + e_{t-a} + \sum_{i=1}^{q} \theta_i \frac{\partial e_{t-i}}{\partial \theta_a}.$$

The second derivatives are then

$$\begin{aligned} \frac{\partial^2 e_t}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}^T} &= -[e^{(1-\lambda)W_t} + \lambda e_t] \frac{\partial^2 W_t}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}^T} \\ &\quad - \left[\frac{\partial W_t}{\partial \boldsymbol{\delta}} (1-\lambda) e^{(1-\lambda)W_t} + \lambda \frac{\partial e_t}{\partial \boldsymbol{\delta}} \right] \frac{\partial W_t}{\partial \boldsymbol{\delta}^T} \end{aligned}$$

and

$$\frac{\partial^2 W_t}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}^T} = \frac{\partial^2 \beta}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}^T} + \frac{\partial \boldsymbol{\delta}^2 Z_t}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}^T} = \frac{\partial \boldsymbol{\delta}^2 Z_t}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}^T},$$

in which

$$\begin{aligned} \frac{\partial^2 Z_t}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}^T} &= \sum_{i=1}^{p} \left[\frac{\partial \phi_i}{\partial \boldsymbol{\delta}} \left(\frac{\partial Z_{t-i}}{\partial \boldsymbol{\delta}^T} + \frac{\partial e_{t-i}}{\partial \boldsymbol{\delta}^T} \right) + \left(\frac{\partial Z_{t-i}}{\partial \boldsymbol{\delta}} + \frac{\partial e_{t-i}}{\partial \boldsymbol{\delta}} \right) \frac{\partial \phi_i}{\partial \boldsymbol{\delta}^T} \right] \\ &\quad + \sum_{i=1}^{p} \phi_i \left(\frac{\partial^2 Z_{t-i}}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}^T} + \frac{\partial^2 e_{t-i}}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}^T} \right) + \sum_{i=1}^{q} \left[\frac{\partial \theta_i}{\partial \boldsymbol{\delta}} \frac{\partial e_{t-i}}{\partial \boldsymbol{\delta}^T} + \frac{\partial e_{t-i}}{\partial \boldsymbol{\delta}} \frac{\partial \theta_i}{\partial \boldsymbol{\delta}^T} \right] \\ &\quad + \sum_{i=1}^{q} \theta_i \frac{\partial^2 e_{t-i}}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}^T}. \end{aligned}$$

2.2 Asymptotic Distribution of MLE

In this section we establish asymptotic properties of the MLEs given in Section 2.1 for the first order model with $\lambda = 1$:

$$W_t = \beta + \gamma \left(Y_{t-1} - e^{W_{t-1}} \right) e^{-W_{t-1}}. \tag{6}$$

Uniform ergodicity and stationarity of $\{W_t\}$ are the key ingredients of the argument.

We first note that by the ergodic theorem,

$$n^{-1} L(\boldsymbol{\delta}) := n^{-1} \sum_{t=1}^{n} (Y_t W_t(\boldsymbol{\delta}) - e^{W_t(\boldsymbol{\delta})}) \to S(\boldsymbol{\delta}) := E_{\boldsymbol{\delta}_0}(Y_1 W_1(\boldsymbol{\delta}) - e^{W_1(\boldsymbol{\delta})})$$

a.s., where $E_{\boldsymbol{\delta}_0}$ represents the expectation operator when the true parameter value is equal to $\boldsymbol{\delta}_0$. The following result establishes identifiability of the model parameterization.

Proposition 2.1 *The function $S(\boldsymbol{\delta})$ has a unique maximum at the true parameter value $\boldsymbol{\delta} = \boldsymbol{\delta}_0$.*

Proof: We have

$$\begin{aligned} S(\boldsymbol{\delta}) &= E_{\boldsymbol{\delta}_0}(e^{W_1(\boldsymbol{\delta}_0)}(W_1(\boldsymbol{\delta}) - W_1(\boldsymbol{\delta}_0) + W_1(\boldsymbol{\delta}_0) - e^{W_1(\boldsymbol{\delta})-W_1(\boldsymbol{\delta}_0)}) \\ &\leq E_{\boldsymbol{\delta}_0}(e^{W_1(\boldsymbol{\delta}_0)}(W_1(\boldsymbol{\delta}_0) - 1)) \\ &= S(\boldsymbol{\delta}_0)\,, \end{aligned}$$

where the inequality follows from the relation $x - e^x \leq -1$. The inequality is an equality if and only if $W_1(\boldsymbol{\delta}) = W_1(\boldsymbol{\delta}_0)$ a.s. However, in this case, we have

$$0 = \beta_0 - \beta + (\gamma_0 e^{-W_0(\boldsymbol{\delta}_0)} - \gamma e^{-W_0(\boldsymbol{\delta})})Y_0 - (\gamma_0 - \gamma)$$

for $Y_0 = 0, 1, \ldots$. It follows that $\beta = \beta_0$ and $\gamma = \gamma_0$ which completes the proof. □

We next consider the normalized score function evaluated at $\boldsymbol{\delta}_0$,

$$H_n := n^{-1/2}\frac{\partial L(\boldsymbol{\delta}_0)}{\partial \boldsymbol{\delta}} = n^{-1/2}\sum_{t=1}^{n} e_t e^{W_t}\dot{W}_t\,, \tag{7}$$

where $W_t = W_t(\boldsymbol{\delta}_0)$, $\dot{W}_t = \frac{\partial W_t(\boldsymbol{\delta}_0)}{\partial \boldsymbol{\delta}}$, and $e_t = \left(Y_t - e^{W_t}\right)/e^{W_t}$. The following result establishes the asymptotic normality of the score function, which is typically the primary factor in establishing the asymptotic normality of maximum likelihood estimates.

Proposition 2.2 *If $\gamma > 0$ and $\gamma(1 + e^{\gamma-\beta})^{1/2} < 1$, then the normalized score function H_n is asymptotically normal with mean $\mathbf{0}$ and covariance matrix $V(\boldsymbol{\delta}_0)$ given by*

$$\frac{1}{n}\sum_{t=1}^{n} e^{W_t}\dot{W}_t\dot{W}_t^T \xrightarrow{a.s.} V(\boldsymbol{\delta}_0) := E(e^{W_1}\dot{W}_1\dot{W}_1^T)$$

Proof: It is easy to see that H_n is a sum of a triangular array of vector martingale differences,

$$\eta_{nt} = n^{-1/2}e_t b_t,$$

where

$$b_t = \dot{W}_t e^{W_t} = \dot{W}_t\mu_t.$$

In order to apply a martingale central limit theorem, it suffices to show (see Corollary 3.1 of Hall and Heyde [4]) that

$$\sum_{t=1}^{n} E(\eta_{nt}\eta_{nt}^T \mid \mathcal{F}_{t-1}) \xrightarrow{P} V(\boldsymbol{\delta}_0), \tag{8}$$

where $\mathcal{F}_t = \sigma(Y_s, s \leq t)$, and, for all $\epsilon > 0$,

$$\sum_{t=1}^{n} E\left(\eta_{nt}\eta_{nt}^T I[|\eta_{nt}| > \epsilon] \;\middle|\; \mathcal{F}_{t-1}\right) \xrightarrow{P} 0. \tag{9}$$

Under these two conditions, we have

$$H_n \xrightarrow{d} N(0, V).$$

To establish conditions (8) and (9), we see from (6) that

$$\begin{aligned}
\dot{W}_t &= \begin{bmatrix} \frac{\partial W_t}{\partial \gamma} \\ \frac{\partial W_t}{\partial \beta} \end{bmatrix} = \begin{bmatrix} \dot{W}_{t,1} \\ \dot{W}_{t,2} \end{bmatrix} \\
&= \begin{bmatrix} Y_{t-1}e^{-W_{t-1}} - 1 - \gamma Y_{t-1}e^{-W_{t-1}}\dot{W}_{t-1,1} \\ 1 - \gamma Y_{t-1}e^{-W_{t-1}}\dot{W}_{t-1,2} \end{bmatrix} \\
&= \begin{bmatrix} U_t + A_t\dot{W}_{t-1,1} \\ 1 + A_t\dot{W}_{t-1,2} \end{bmatrix} = \begin{bmatrix} U_t + \sum_{i=1}^{\infty} A_t \cdots A_{t-i+1}U_{t-i} \\ 1 + \sum_{i=1}^{\infty} A_t \cdots A_{t-i+1} \end{bmatrix}, \qquad (10)
\end{aligned}$$

where $U_t = Y_{t-1}e^{-W_{t-1}} - 1$ and $A_t = -\gamma Y_{t-1}e^{-W_{t-1}}$. Since $\dot{W}_t$ is a function of $\{W_s, s \leq t\}$, it also is a strictly stationary ergodic process. Now,

$$\sum_{t=1}^{n} E(\eta_{nt}\eta_{nt}^T \mid \mathcal{F}_{t-1}) = \frac{1}{n}\sum_{t=1}^{n} e^{W_t}\dot{W}_t\dot{W}_t^T,$$

which is a function of two stationary ergodic processes, $\{W_t\}$ and $\{\dot{W}_t\}$. By the ergodic theorem we then have

$$\frac{1}{n}\sum_{t=1}^{n} e^{W_t}\dot{W}_t\dot{W}_t^T \xrightarrow{a.s.} V = E(e^{W_1}\dot{W}_1\dot{W}_1^T)$$

if $E|e^{W_t(\boldsymbol{\delta}_0)}\dot{W}_t\dot{W}_t^T| < \infty$. Conditions under which this holds will now be derived for a particular choice of parameter values of β and γ. It suffices to show $E|e^{W_t}\dot{W}_{t,i}^2| < \infty,\ i = 1, 2$. First we will consider the case $i = 1$. Using $\|\cdot\|_2$ to denote the L_2 norm, we have from (10),

$$\|e^{W_t/2}\dot{W}_{t,1}\|_2 \leq \|e^{W_t/2}U_t\|_2 + \sum_{i=1}^{\infty} \|e^{W_t/2}A_t \cdots A_{t-i+1}U_{t-i}\|_2.$$

Using properties of the moment generating function for a Poisson distributed random variable and the fact that the process W_t is bounded below by $\beta - \gamma$, we have

$$\begin{aligned}
\|e^{W_t/2}U_t\|_2^2 &= E\left[e^{\beta-\gamma}e^{\gamma Y_{t-1}e^{-W_{t-1}}}(Y_{t-1}e^{-W_{t-1}} - 1)^2\right] \\
&= e^{\beta-\gamma}E\left[E\left(e^{\gamma Y_{t-1}e^{-W_{t-1}}}(Y_{t-1}^2e^{-2W_{t-1}} - 2Y_{t-1}e^{-W_{t-1}} + 1) \mid W_{t-1}\right)\right] \\
&= e^{\beta-\gamma}E\Big[e^{-W_{t-1}}e^{\gamma e^{-W_{t-1}}}e^{e^{W_{t-1}}(e^{\gamma e^{-W_{t-1}}}-1)} + e^{2\gamma e^{-W_{t-1}}}e^{e^{W_{t-1}}(e^{\gamma e^{-W_{t-1}}}-1)} \\
&\quad -2e^{\gamma e^{-W_{t-1}}}e^{e^{W_{t-1}}(e^{\gamma e^{-W_{t-1}}}-1)} + e^{e^{W_{t-1}}(e^{\gamma e^{-W_{t-1}}}-1)}\Big] \\
&= e^{\beta-\gamma}E\left[e^{e^{W_{t-1}}(e^{\gamma e^{-W_{t-1}}}-1)}\left[1 + e^{\gamma e^{-W_{t-1}}}\left(e^{\gamma e^{-W_{t-1}}} + e^{-W_{t-1}} - 2\right)\right]\right] \\
&\leq e^{\beta-\gamma}e^{e^{\beta-\gamma}(e^{\gamma e^{-(\beta-\gamma)}}-1)}\left[1 + e^{\gamma e^{-(\beta-\gamma)}}\left(e^{\gamma e^{-(\beta-\gamma)}} + e^{-(\beta-\gamma)} - 2\right)\right] := c_1^2, \qquad (11)
\end{aligned}$$

where the last inequality follow from the fact that the function $x(\exp(\gamma x^{-1})-1)$ is decreasing in $x>0$.

$$\begin{aligned}
E\left[e^{W_t}A_t^2 \mid \mathcal{F}_{t-1}\right] &= E\left[\gamma^2 e^{\beta-\gamma}Y_{t-1}^2 e^{-2W_{t-1}} e^{\gamma Y_{t-1}e^{-W_{t-1}}} \mid W_{t-1}\right] \\
&= \gamma^2 e^{\beta-\gamma}\left[e^{-W_{t-1}}e^{\gamma e^{-W_{t-1}}}e^{e^{W_{t-1}}(e^{\gamma e^{-W_{t-1}}}-1)} + e^{2\gamma e^{-W_{t-1}}}e^{e^{W_{t-1}}(e^{\gamma e^{-W_{t-1}}}-1)}\right] \\
&\le \gamma^2 e^{\beta-\gamma}e^{e^{\beta-\gamma}(e^{\gamma e^{-(\beta-\gamma)}}-1)}e^{\gamma e^{-(\beta-\gamma)}}\left(e^{\gamma e^{-(\beta-\gamma)}}+e^{-(\beta-\gamma)}\right) := b_1^2, \\
E\left[A_t^2 \mid \mathcal{F}_{t-1}\right] &= E\left[\gamma^2 Y_{t-1}^2 e^{-2W_{t-1}} \mid W_{t-1}\right] \\
&= \gamma^2\left(1+e^{-W_{t-1}}\right) \\
&\le \gamma^2(1+e^{-(\beta-\gamma)})
\end{aligned}$$

and

$$\begin{aligned}
E\left[U_t^2 \mid \mathcal{F}_{t-1}\right] &= E\left[Y_{t-1}^2 e^{-2W_{t-1}} - 2Y_{t-1}e^{-W_{t-1}} + 1 \mid W_{t-1}\right] \\
&= e^{-W_{t-1}} \\
&\le e^{-(\beta-\gamma)} := b_2^2.
\end{aligned}$$

Applying these results, $\|e^{W_t/2}A_t\cdots A_{t-i+1}U_{t-i}\|_2^2$ may be calculated recursively:

$$\begin{aligned}
\|e^{W_t/2}A_t\cdots A_{t-i+1}U_{t-i}\|_2^2 &= E\left(e^{W_t}A_t^2\cdots A_{t-i+1}^2 U_{t-i}^2\right) \\
&= E\left[E\left(e^{W_t}A_t^2\cdots A_{t-i+1}^2 U_{t-i}^2 \mid \mathcal{F}_{t-1}\right)\right] \\
&= E\left[A_{t-1}^2\cdots A_{t-i+1}^2 U_{t-i}^2 E\left(e^{W_t}A_t^2 \mid \mathcal{F}_{t-1}\right)\right] \\
&\le b_1^2 E\left[E\left(A_{t-1}^2\cdots A_{t-i+1}^2 U_{t-i}^2 \mid \mathcal{F}_{t-2}\right)\right] \\
&= b_1^2 E\left[A_{t-2}^2\cdots A_{t-i+1}^2 U_{t-i}^2 E\left(A_{t-1}^2 \mid \mathcal{F}_{t-2}\right)\right] \\
&\le b_1^2\gamma^2(1+e^{-(\beta-\gamma)})E\left[A_{t-2}^2\cdots A_{t-i+1}^2 U_{t-i}^2\right] \\
&\vdots \\
&\le b_1^2\left(\gamma^2(1+e^{-(\beta-\gamma)})\right)^{i-1} E\left[E\left(U_{t-i} \mid \mathcal{F}_{t-i-1}\right)\right] \\
&\le b_1^2 b_2^2\left(\gamma^2(1+e^{-(\beta-\gamma)})\right)^{i-1}.
\end{aligned}$$

Therefore,

$$\|e^{W_t/2}\dot{W}_{t,1}\|_2 \le c_1 + c_2\sum_{i=1}^{\infty}\gamma^{i-1}(1+e^{\gamma-\beta})^{(i-1)/2},$$

where $c_2 = b_1 b_2$. Likewise

$$\begin{aligned}
\|e^{W_t/2}\dot{W}_{t,2}\|_2 &\le \|e^{W_t/2}\|_2 + \sum_{i=0}^{\infty}\|e^{W_t/2}A_t\cdots A_{t-i}\|_2 \\
&\le c_3 + c_4\sum_{i=1}^{\infty}\gamma^{i-1}(1+e^{\gamma-\beta})^{(i-1)/2},
\end{aligned}$$

where

$$c_3 = \left[e^{\beta-\gamma} e^{e^{\beta-\gamma}(e^{\gamma e^{-(\beta-\gamma)}}-1)}\right]^{1/2},$$

and

$$c_4 = \left[\gamma^2 e^{\beta-\gamma} e^{e^{\beta-\gamma}(e^{\gamma e^{-(\beta-\gamma)}}-1)} e^{\gamma e^{-(\beta-\gamma)}} \left(e^{\gamma e^{-(\beta-\gamma)}} + e^{-(\beta-\gamma)}\right)\right]^{1/2}.$$

It follows that $E|e^{W_t}\dot{W}_t\dot{W}_t^T|$ will be finite for $\gamma(1+e^{\gamma-\beta})^{1/2} < 1$.

The convergence required in condition (9) is easily established using condition (8) and the stationarity of $\{W_t\}$. Now,

$$\begin{aligned}
&\sum_{t=1}^n E\left(\eta_{nt}\eta_{nt}^T I[|\eta_{nt}| > \epsilon] \mid \mathcal{F}_{t-1}\right) \\
&\qquad = \frac{1}{n}\sum_{t=1}^n E\left[(Y_{t-1} - e^{W_{t-1}})^2 \dot{W}_t\dot{W}_t^T I[|(Y_{t-1} - e^{W_{t-1}})\dot{W}_t| > \epsilon\sqrt{n}] \mid \mathcal{F}_{t-1}\right] \\
&\qquad \le \frac{1}{n}\sum_{t=1}^n E\left[(Y_{t-1} - e^{W_{t-1}})^2 \dot{W}_t\dot{W}_t^T I[|(Y_{t-1} - e^{W_{t-1}})\dot{W}_t| > M] \mid \mathcal{F}_{t-1}\right] \\
&\qquad \stackrel{n\to\infty}{\longrightarrow} E\left[(Y_1 - e^{W_1})^2 \dot{W}_1\dot{W}_1^T I[|(Y_1 - e^{W_1})\dot{W}_1| > M]\right] \\
&\qquad \longrightarrow 0 \text{ as } M \to \infty.
\end{aligned}$$

□

To argue that the MLE is asymptotically normal, consider the the linearized version of the log-likelihood obtained by linearizing $W_t(\boldsymbol{\delta})$ in a neighborhood of the true value $\boldsymbol{\delta}_0$ in (4). Specifically, let

$$W_t^\dagger(\boldsymbol{\delta}) = W_t(\boldsymbol{\delta}_0) + (\boldsymbol{\delta} - \boldsymbol{\delta}_0)^T \dot{W}_t,$$

so that the linearized log-likelihood takes the form

$$L^\dagger(\boldsymbol{\delta}) = \sum_{t=1}^n \left(Y_t W_t^\dagger(\boldsymbol{\delta}) - e^{W_t^\dagger(\boldsymbol{\delta})}\right).$$

After re-parameterizing with the transformation $u = n^{1/2}(\boldsymbol{\delta} - \boldsymbol{\delta}_0)$, we have

$$\begin{aligned}
R_n^\dagger(u) &:= L^\dagger(\boldsymbol{\delta}_0) - L^\dagger(\boldsymbol{\delta}_0 + un^{-1/2}) \\
&= -u^T n^{-1/2} \sum_{t=1}^n Y_t \dot{W}_t + \sum_{t=1}^n e^{W_t}\left(e^{u^T n^{-1/2}\dot{W}_t} - 1\right) \\
&= -u^T n^{-1/2} \sum_{t=1}^n \left(Y_t - e^{W_t}\right)\dot{W}_t + \sum_{t=1}^n e^{W_t}\left(e^{u^T n^{-1/2}\dot{W}_t} - 1 - u^T n^{-1/2}\dot{W}_t\right). \qquad (12)
\end{aligned}$$

Note that $R_n^\dagger(u)$ is a convex function of u. The first term in (12) is $-u^T H_n$ which, by Proposition 2.2, is asymptotically normal with mean 0 and variance $-u^T V(\boldsymbol{\delta}_0)u$. The second term in (12) is

$$u^T\left[(2n)^{-1}\sum_{t=1}^n e^{W_t}\dot{W}_t\dot{W}_t^T\right]u + O_p\left(n^{-3/2}\sum_{t=1}^n e^{W_t}(u^T\dot{W}_t)^3\right) = 2^{-1}u^T V(\boldsymbol{\delta}_0)u + o_P(1),$$

so that $R_n^\dagger(u) \xrightarrow{d} R(u)$ where

$$R(u) = -u^T N(0, V) + u^T V u/2.$$

This convergence extends to finite dimensional distributions and since $R_n^\dagger$ is convex, the convergence is actually on $C(\mathbb{R}^2)$, the space of continuous functions on $\mathbb{R}^2$ (see Remark 1 of Davis et al. (1992)). Moreover, the convexity implies that the minimizer, $\hat{u}_n = n^{1/2}(\hat{\boldsymbol{\delta}}_n^\dagger - \boldsymbol{\delta}_0)$, of $R_n^\dagger$ converges in distribution to the unique minimizer of $R(u)$. It is easy to see that this minimizer is $V^{-1}N(0, V)$ which has a $N(0, V^{-1})$ distribution.

Under suitable smoothness conditions, the convergence of $R_n^\dagger(u)$ can be transferred to $R_n(u) := L(\boldsymbol{\delta_0}) - L(un^{-1/2} + \boldsymbol{\delta_0})$. That is, $R_n^\dagger(u) - R_n(u) \xrightarrow{P} 0$ uniformly for $|u| \leq K$. In this case, $\hat{\boldsymbol{\delta}}_n^\dagger$ and the maximum likelihood estimator $\hat{\boldsymbol{\delta}}_n$ have the same limiting distribution, namely, $N(0, V^{-1})$.

Simulation results from these models (see Davis et al. (2003)) show close agreement between the theoretical values and the model estimates, thereby supporting the derived theory.

Acknowledgment

The research of the first author was supported in part by NSF grant DMS-0308109 and EPA STAR CR-829095.

References

[1] Davis, R.A., Knight, K., and Liu, J. (1992). M-estimation for Autoregressions with Infinite Variance. *Stoch. Proc. Appl.* **40**, 145–180.

[2] Davis, R.A., Dunsmuir, W.T.M., and Wang, Y. (1999). "Modelling Time Series of Count Data." *Asymptotics, Nonparametrics, and Time Series* (Subir Ghosh, editor) Marcel-Dekker, New York, 63–114.

[3] Davis, R.A., Dunsmuir, W.T.M., and Streett, S.B. (2003). "Observation-Driven Models for Poisson Counts. " (To appear in *Biometrika*.)

[4] Hall, P. and Heyde, C.C. (1980) "Martingale limit theory and its application", Academic Press, New York.