Observation Driven Models for Poisson Counts

by R. Davis, W. Dunsmuir and S. Street

Evangelos Evangelou

SAMSI Extreme Value Methodology Workgroup

January 14, 2008

Poisson time-series

Examples related to extreme events that can be modeled as Poisson:

- The number of occurrences of a meteorological phenomenon e.g. hurricanes that strike every year
- Epidemiology data e.g. death counts

Observation driven v.s. parameter driven models

GLM idea
Mean process $\mu_t := \mathbb{E}(Y_t|Z_t) = \exp(x_t^\intercal \beta + Z_t)$

Cox (1981):

- *Parameter-driven models*: The dist'n of Z_{t+1} depends only on Z_t and does not depend on $(Z_{t-1}, Z_{t-2}, \ldots), (Y_t, Y_{t-1}, \ldots)$. The dependence structure of $\{Y_t\}$ is inherited from that of $\{Z_t\}$.

- *Observation-driven models*: The dist'n of Z_{t+1} depends on $(Y_t, Y_{t-1}, \ldots)$.

Observation driven models are easier to forecast but theoretical properties are more difficult to establish.

Desirable properties

Zeger and Qaqish (1988):

1. Approximate relationship for the mean
   $\mathbb{E}(Y_t) = \mathbb{E}(\mu_t) \approx \exp(x_t^\intercal \beta)$
   so that β is interpreted as the proportional change in the expectation of Y_t given a unit change in x_t.

2. The model should allow for both positive and negative serial dependence. For example, $\log(\mu_t) = \beta + \gamma Y_{t-1}$ is not stationary unless $\gamma \leq 0$.

3. The estimators for β and γ should be approximately orthogonal for easier computation.

In addition Davis, Dunsmuir and Wang (1999):

4. The model should be easy to forecast.
5. The computational procedure should be easy to implement; earlier obs-d models are computationally intensive.
6. Diagnostic tools should be available.

Poisson GLARMA(p, q) model

1. $Y_t | Y_{t-1}, Y_{t-2}, \ldots \sim \mathrm{Poisson}(\mu_t)$
2. Define the *state* process $W_t := \log(\mu_t)$
3. $e_t := (Y_t - e^{W_t}) e^{-\lambda W_t}$, $0 < \lambda \leq 1$ fixed, is a martingale difference sequence
4. $Z_t := \sum_{i=1}^{\infty} \gamma_i e_{t-i}$, $\sum |\gamma_i| < \infty$
5. $W_t = x_t^\intercal \beta + Z_t$

Reparameterization

Suppose $\{U_t\} \sim \mathrm{ARMA}(p,q)$
$U_t = \phi_1 U_{t-1} + \ldots + \phi_p U_{t-p} + e_t + \theta_1 e_{t-1} + \ldots + \theta_q e_{t-q}$

where $\displaystyle\sum_{i=1}^{\infty} \gamma_i z^i \equiv \left(1 - \sum_{i=1}^{p} \phi_i z^i\right)^{-1} \left(1 + \sum_{i=1}^{q} \theta_i z^i\right) - 1$

Then Z_t satisfies

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

Set $e_s = 0$ and $Z_s = 0$ for $s \leq 0$

A few properties

1. $\mathbb{E}(e_t | e_{t-1}, e_{t-2}, \ldots) = 0$ and $\mathbb{E}(e_t) = 0$
2. $\mathbb{E}(e_t^2) = \mathbb{E}(\mu_t^{1-2\lambda})$
3. $\text{cov}(e_t, e_s) = 0$ for $t \neq s$
4. $\mathbb{E}(W_t) = x_t^\intercal \beta$
5. $\text{cov}(W_t, W_{t+h}) = \sum_{i=1}^{\infty} \gamma_i \gamma_{i+h} \mathbb{E}(\mu_{t-i}^{1-2\lambda})$
6. for $\lambda = 0.5$, $\mathbb{E}(\mu_t) \approx \exp(x_t^\intercal \beta + \frac{1}{2} \sum_{i=1}^{\infty} \gamma_i^2)$

Remarks

- The accuracy of the approximation at 6 depends on how close $\{e_t\}$ is to a process of independent normal random variables.
- $\hat{\mu}_t := \exp\left(\hat{W}_t - \frac{1}{2}\sum_{i=1}^{\infty}\hat{\gamma}_i^2\right)$ is an unbiased prediction of Y_t.

The basic model

Set

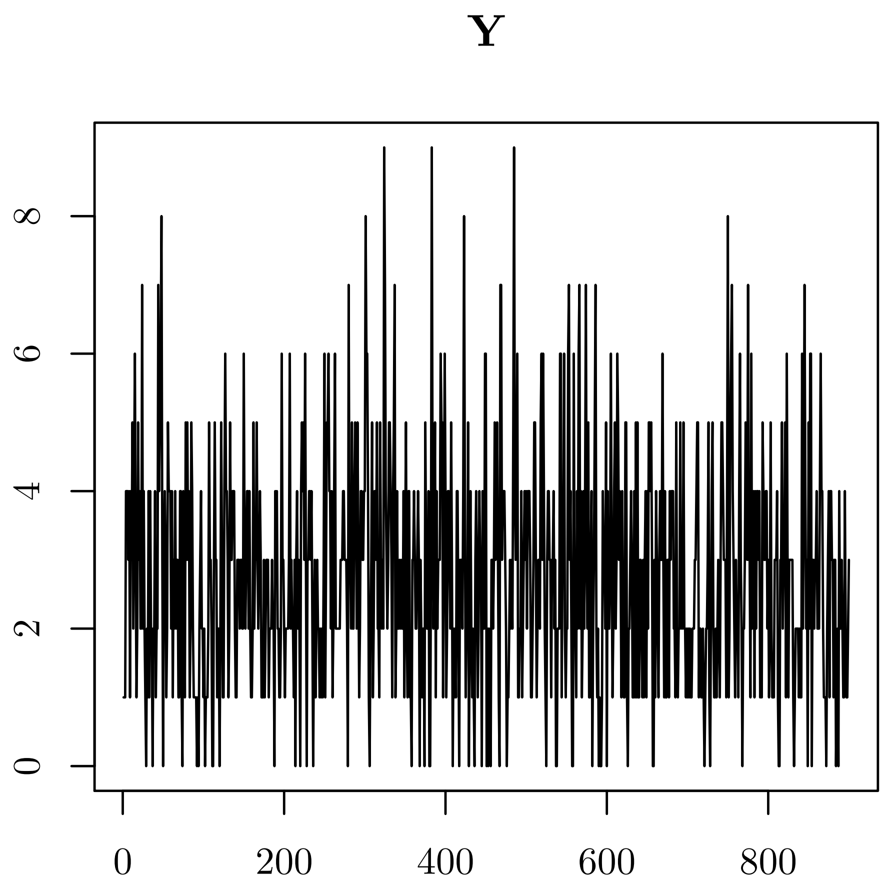
- $x_t^\mathsf{T}\beta = \beta$
- $p = 0$
- $q = 1$

Then

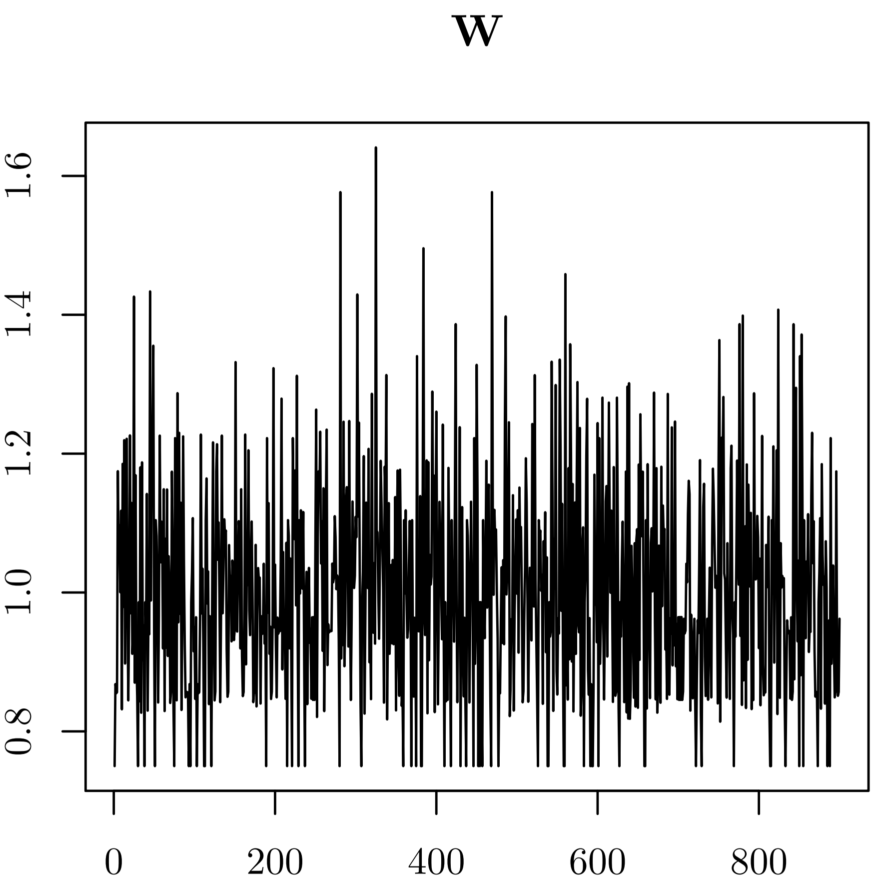
$$W_t = \beta + \gamma \left(Y_{t-1} - e^{W_{t-1}} \right) e^{-\lambda W_{t-1}} \tag{2}$$

Proposition: For $\frac{1}{2} \leq \lambda \leq 1$ and $\gamma \neq 0$, $\{W_t\}$ has a stationary dist'n. If $\lambda = 1$, the dist'n is unique.

$$\lambda = 1$$



Y
0
2
4
6
8
0
200
400
600
800
Time



W
0.8
1.0
1.2
1.4
1.6
0
200
400
600
800
Time

Maximum likelihood estimation

Details in Davis, Dunsmuir and Street (2005) section 2.1

$\delta := (\beta, \gamma^\intercal)^\intercal$

$\ell(\delta) \approx \sum_{t=1}^{n} \left(Y_t \, W_t(\delta) - e^{W_t(\delta)} \right)$

Iteratively:

`Set` $e_0 = \ldots = e_{1-q} = 0$ `and` $Z_0 = \ldots = Z_{1-p} = 0$
`for` $t = 1, \ldots, n$ $\{$
$\quad Z_t = \sum_{i=1}^{p} \phi_i (Z_{t-i} + e_{t-i}) + \sum_{i=1}^{q} \theta_i e_{t-i}$
$\quad W_t = x_t^\intercal \beta + Z_t$
$\quad e_t = (Y_t - e^{W_t}) e^{-\lambda W_t}$
$\}$
$\ell = \sum_{t=1}^{n} (Y_t \, W_t - e^{W_t})$

1st and 2nd derivatives are calculated similarly.

The Newton-Raphson algorithm is used to obtain $\hat{\delta}$. Convergence is achieved within 10 iterations.

When $\lambda = 1, q = 1$, $\hat{\delta} \overset{\text{aprx}}{\sim} \mathrm{N}(0, V^{-1})$ where

$$V = \lim_n \frac{1}{n} \sum_{t=1}^{n} e^{W_t} \left(\frac{\partial W_t}{\partial \delta} \right) \left(\frac{\partial W_t}{\partial \delta} \right)^{\mathsf{T}} \tag{3}$$

Connection with extreme value theory

Two approaches of analyzing extremes in time-series (from the book statistics of extremes):

1. Time-series model for the complete process
2. Model only the extreme values

Idea:

1. Different form of $\{Z_t\}$
   e.g. moving maxima process $\rightarrow$ parameter-driven model
2. Use the model proposed by Davis et al but focus only on the extreme values