Observation Driven Models for Time Series of Counts

Richard A. Davis
Colorado State University
(http://www.stat.colostate.edu/~rdavis/lectures)

Joint work with:
William Dunsmuir, University of New South Wales
Sarah Streett, National Center for Atmospheric Research
Outline

- Introduction
  - Examples
- Parameter-driven models
- Observation-driven models
- Generalized Linear ARMA (GLARMA) models for Poisson counts
  - Properties
  - Existence and uniqueness of stationary distributions
  - Estimation and asymptotic theory for MLE
  - Application to asthma data
- GLARMA extensions
  - Bernoulli
- Other (BIN)
Example: Monthly Polio Counts in USA (Zeger 1988)
Count data: \( Y_1, \ldots, Y_n \)

Regression (explanatory) variable: \( x_t \)

Model: Distribution of the \( Y_t \) given \( x_t \) and a stochastic process \( \nu_t \) are indep Poisson distributed with mean

\[
\mu_t = \exp(x_t^T \beta + \nu_t).
\]

The distribution of the stochastic process \( \nu_t \) may depend on a vector of parameters \( \gamma \).

Note: \( \nu_t = 0 \) corresponds to standard Poisson regression model.

Primary objective: Inference about \( \beta \).
Regression function:

$$x_t^T = (1, \ t´/1000, \ \cos(2\pi t´/12),\ \sin(2\pi t´/12),\ \cos(2\pi t´/6),\ \sin(2\pi t´/6))$$

where \(t´=(t-73)\).

Summary of various models fits to Polio data:

<table>
<thead>
<tr>
<th>Study</th>
<th>Trend(β)</th>
<th>SE(β)</th>
<th>t-ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLM Estimate</td>
<td>-4.80</td>
<td>1.40</td>
<td>-3.43</td>
</tr>
<tr>
<td>Zeger (1988)</td>
<td>-4.35</td>
<td>2.68</td>
<td>-1.62</td>
</tr>
<tr>
<td>Kuk &amp; Chen (1996) MCNR</td>
<td>-3.79</td>
<td>2.95</td>
<td>-1.28</td>
</tr>
<tr>
<td>Fahrmeir and Tutz (1994)</td>
<td>-3.33</td>
<td>2.00</td>
<td>-1.67</td>
</tr>
<tr>
<td>Durbin and Koopman</td>
<td>-3.78</td>
<td>2.86</td>
<td>-1.32</td>
</tr>
</tbody>
</table>
**Parameter-Driven Model for the Mean Function $\mu_t$**

**Parameter-driven specification:** (Assume $Y_t | \mu_t$ is Poisson($\mu_t$))

$$\log \mu_t = x_t^T \beta + \nu_t,$$

where $\{\nu_t\}$ is a stationary Gaussian process.

**e.g.** (AR(1) process)

$$(\nu_t + \sigma^2/2) = \phi(\nu_{t-1} + \sigma^2/2) + \epsilon_t, \quad \{\epsilon_t\} \sim \text{IID } N(0, \sigma^2(1-\phi^2)).$$

**Advantages:**

- properties of model (ergodicity and mixing) easy to derive.
- interpretability of regression parameters

$$E(Y_t) = \exp(x_t^T \beta) E \exp(\nu_t) = \exp(x_t^T \beta), \quad \text{if } E \exp(\nu_t) = 1.$$

**Disadvantages:**

- estimation is difficult-likelihood function not easily calculated (MCEM, importance sampling, estimating eqns).
- model building can be laborious
- prediction is more difficult.
Observation Driven Model for the Mean Function $\mu_t$

Observation-driven specification: (Assume $Y_t \mid \mu_t$ is Poisson($\mu_t$))

$$\log\mu_t = x_t^T \beta + \nu_t,$$

where $\nu_t$ is a function of past observations $Y_s, s < t$.

E.g. $\nu_t = \gamma_1 Y_{t-1} + \ldots + \gamma_p Y_{t-p}$

Advantages:

- likelihood easy to calculate
- prediction is straightforward (at least one lead-time ahead).

Disadvantages:

- stability behavior, such as stationarity and ergodicity, is difficult to derive.
- $x_t^T \beta$ is not easily interpretable. In the special case above,

$$E(Y_t) = \exp(x_t^T \beta) \cdot \exp(\gamma_1 Y_{t-1} + \ldots + \gamma_p Y_{t-p})$$
Generalized Linear ARMA (GLARMA) Model for Poisson Counts

Two components in the specification of $\nu_t$ (see also Shephard (1994)).

1. Uncorrelated (martingale difference sequence)

For $\lambda > 0$, define

$$e_t = (Y_t - \mu_t) / \mu_t^\lambda$$

(Specification of $\lambda$ will be described later.)

2. Form a linear process driven by the MGD sequence $\{e_t\}$

$$\log \mu_t = x_t^T \beta + \nu_t,$$

where

$$\nu_t = \sum_{i=1}^{\infty} \psi_i e_{t-i}.$$  

Since the conditional mean $\mu_t$ is based on the whole past, the model is no longer Markov. Nevertheless, this specification could lead to stationary solutions, although the stability theory appears difficult.
Properties of the New Model

\[ e_t = (Y_t - \mu_t) / \mu_t^\lambda, \quad \log \mu_t = x_t^T \beta + \nu_t, \quad \nu_t = \sum_{i=1}^{\infty} \psi_i e_{t-i}. \]

1. \{e_t\} is a MG difference sequence \( E(e_t \mid F_{t-1}) = 0 \)
2. \{e_t\} is an uncorrelated sequence (follows from 1)
3. \( E(e_t^2) = E(\mu_t^{1-2\lambda}) \)
   \[ = 1 \text{ if } \lambda = .5 \]
4. Set, \( W_t = \log \mu_t = x_t^T \beta + \nu_t, \)

so that

\[ E(W_t) = x_t^T \beta \quad \text{and} \quad \text{Var}(W_t) = \sum_{i=1}^{\infty} \psi_i^2 E(\mu_{t-i}^{1-2\lambda}) \]

\[ = \sum_{i=1}^{\infty} \psi_i^2 \quad (\text{if } \lambda = .5) \]
5. \( \text{Cov}(W_t, W_{t+h}) = \sum_{i=1}^{\infty} \psi_i \psi_{i+h} E(\mu_{t-i}^{1-2\lambda}) \)

It follows that \( \{W_t\} \) has properties similar to the latent process specification:

\[
W_t = x_t^T \beta + \sum_{i=1}^{\infty} \psi_i e_{t-i}
\]

which, by using the results for the latent process case and assuming the linear process part is nearly Gaussian, we obtain

\[
E(e^{W_t}) = E(e^{x_t^T \beta + \sum_i \psi_i e_{t-i}})
\approx e^{x_t^T \beta + \text{Var}(\nu_t)/2}
\approx e^{x_t^T \beta + \sum_i \psi_i^2 / 2}
= e^{x_t^T \beta + \sum_i \psi_i^2 / 2},
\]

By adjusting the intercept term, \( E(\mu_t) \) can be interpreted as \( \exp(x_t^T \beta) \).
6. (GLARMA model). Let \( \{U_t\} \) be an ARMA process with driven by the MGD sequence \( \{e_t\} \), i.e.,

\[
U_t = \phi_1 U_{t-1} + \ldots + \phi_p U_{t-p} + e_t + \theta_1 e_{t-1} + \ldots + \theta_q e_{t-q}
\]

Then the best predictor of \( U_t \) based on the infinite past is

\[
\hat{U}_t = \sum_{i=1}^{\infty} \psi_i e_{t-i}
\]

where

\[
\sum_{i=1}^{\infty} \psi_i z^i = \phi(z)^{-1} \theta(z) - 1.
\]

The model for \( \log \mu_t \) is then

\[
W_t = x_t^T \beta + Z_t,
\]

where

\[
Z_t = \hat{U}_t = \phi_1 (Z_{t-1} + e_{t-1}) + \ldots + \phi_p (Z_{t-p} + e_{t-p}) + \theta_1 e_{t-1} + \ldots + \theta_q e_{t-q}.
\]
Existence and uniquess of a stationary distr in the simple case.

Consider the simplest form of the model with \( \lambda = 1 \), given by

\[
W_t = \beta + \gamma (Y_{t-1} - e^{W_{t-1}}) e^{-W_{t-1}}.
\]

**Theorem:** The Markov process \( \{W_t\} \) has a unique stationary distribution.

**Idea of proof:**

- State space is \([\beta - \gamma, \infty)\) (if \( \gamma > 0 \)) and \((-\infty, \beta - \gamma]\) (if \( \gamma < 0 \)).
- Satisfies Doeblin’s condition:

  There exists a prob measure \( \nu \) such for some \( m > 1, \epsilon > 0, \) and \( \delta > 0, \)

  \[
  \nu(A) > \epsilon \quad \text{implies} \quad P^m(x,A) \geq \delta \quad \text{for all } x.
  \]
- Chain is strongly aperiodic.
- It follows that the chain \( \{W_t\} \) is *uniformly ergodic* (Thm 16.0.2 (iv) in Meyn and Tweedie (1993))
Existence of Stationary Distr in Case \(0.5 \leq \lambda < 1\).

Consider the process
\[
W_t = \beta + \gamma(Y_{t-1} - e^{W_{t-1}})e^{-\lambda W_{t-1}}.
\]

Proposition: The Markov process \(\{W_t\}\) has at least one stationary distribution.

Idea of proof:

- \(\{W_t\}\) is weak Feller.
- \(\{W_t\}\) is bounded in probability on average, i.e., for each \(x\), the sequence \(k^{-1}\sum_{i=1}^{k} P^i(x, \cdot), \ k = 1, 2, \ldots\) is tight.
- There exists at least one stationary distribution (Thm 12.0.1 in M&T)

Lemma: If a MC \(\{X_t\}\) is weak Feller and \(\{P(x, \cdot), \ x \in X\}\) is tight, then \(\{X_t\}\) is bounded in probability on average and hence has a stationary distribution.

Note: For our case, we can show tightness of \(\{P(x, \cdot), \ x \in X\}\) using a Markov style inequality.
Theorem (M&T `93): If the Markov process \( \{X_t\} \) is an e-chain which is bounded in probability on average, then there exists a unique stationary distribution if and only if there exists a reachable point \( x^* \).

For the process \( W_t = \beta + \gamma(Y_{t-1} - e^{W_{t-1}})e^{-\lambda W_{t-1}} \), we have

- \( \{W_t\} \) is bounded in probability uniformly over the state space.
- \( \{W_t\} \) has a reachable point \( x^* \) that is a zero of the equation
  \[
  0 = x^* + \gamma \exp\{(1-\lambda) x^*\}
  \]
- e-chain?

Reachable point: \( x^* \) is a reachable point if for every open set \( O \) containing \( x^* \),

\[
\sum_{n=1}^{\infty} P^n(x, O) > 0 \text{ for all } x.
\]
e-chain: For every continuous \( f \) with compact support, the sequence of functions \( \{P^n f, n = 1, \ldots\} \) is equicontinuous, on compact sets.
Let $\delta = (\beta^T, \gamma^T)^T$ be the parameter vector for the model ($\gamma$ corresponds to the parameters in the linear process part).

Log-likelihood:

$$L(\delta) = \sum_{t=1}^{n} (Y_t W_t(\delta) - e^{W_t(\delta)}),$$

where

$$W_t(\delta) = x_t \beta + \sum_{i=1}^{\infty} \psi_i(\delta)e_{t-i}.$$ 

First and second derivatives of the likelihood can easily be computed recursively and Newton-Raphson methods are then implementable. For example,

$$\frac{\partial L(\delta)}{\partial \delta} = \sum_{t=1}^{n} (Y_t - e^{W_t(\delta)}) \frac{\partial W_t(\delta)}{\partial \delta}$$

and the term $\partial W_t(\delta) / \partial \delta$ can be computed recursively.

Model: $Y_t \mid \mu_t$ is Poisson($\mu_t$)

$$\log \mu_t = x_t^T \beta + \nu_t,$$

$$\nu_t = \sum_{i=1}^{\infty} \psi_i e_{t-i}.$$
Asymptotic Results for MLE

Define the array of random variables by

\[ \eta_{nt} = n^{-1/2} (Y_t - e^{W_t(\delta)}) \frac{\partial W_t(\delta)}{\partial \delta}. \]

Properties of \( \{\eta_{nt}\} \):

- \( \{\eta_{nt}\} \) is a martingale difference sequence.
- \( \sum_{t=1}^{n} E(\eta_{nt}^T \eta_{nt} | F_{t-1}) \xrightarrow{P} V(\delta). \)
- \( \sum_{t=1}^{n} E(\eta_{nt}^T I(|\eta_{nt}| > \varepsilon) | F_{t-1}) \xrightarrow{P} 0. \)

Using a MG central limit theorem, it “follows” that

\[ n^{1/2} (\hat{\delta} - \delta) \xrightarrow{D} N(0, V^{-1}), \]

where

\[ V = \lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} e^{W_t(\delta)} \partial W_t(\delta) \partial W_t^T(\delta). \]
Simulation Results

Model 1: \( W_t = \beta_0 + \gamma(Y_{t-1} - e^{W_{t-1}})e^{-W_{t-1}}, \) \( n = 500, \text{nreps} = 5000 \)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>SD</th>
<th>SD(from like)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta_0 = 1.50 )</td>
<td>1.499</td>
<td>0.0263</td>
<td>0.0265</td>
</tr>
<tr>
<td>( \gamma = 0.25 )</td>
<td>0.249</td>
<td>0.0403</td>
<td>0.0408</td>
</tr>
<tr>
<td>( \beta_0 = 1.50 )</td>
<td>1.499</td>
<td>0.0366</td>
<td>0.0364</td>
</tr>
<tr>
<td>( \gamma = 0.75 )</td>
<td>0.750</td>
<td>0.0218</td>
<td>0.0218</td>
</tr>
<tr>
<td>( \beta_0 = 3.00 )</td>
<td>3.000</td>
<td>0.0125</td>
<td>0.0125</td>
</tr>
<tr>
<td>( \gamma = 0.25 )</td>
<td>0.249</td>
<td>0.0431</td>
<td>0.0430</td>
</tr>
<tr>
<td>( \beta_0 = 3.00 )</td>
<td>3.000</td>
<td>0.0175</td>
<td>0.0174</td>
</tr>
<tr>
<td>( \gamma = 0.75 )</td>
<td>0.750</td>
<td>0.0270</td>
<td>0.0271</td>
</tr>
</tbody>
</table>

Model 2: \( W_t = \beta_0 + \beta_1 \frac{t}{500} + \gamma(Y_{t-1} - e^{W_{t-1}})e^{-W_{t-1}}, \) \( n = 500, \text{nreps} = 5000 \)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>SD</th>
<th>SD(from like)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta_0 = 1.00 )</td>
<td>1.000</td>
<td>0.0286</td>
<td>0.0284</td>
</tr>
<tr>
<td>( \beta_1 = 0.50 )</td>
<td>0.500</td>
<td>0.0035</td>
<td>0.0034</td>
</tr>
<tr>
<td>( \gamma = 0.25 )</td>
<td>0.248</td>
<td>0.0420</td>
<td>0.0426</td>
</tr>
<tr>
<td>( \beta_0 = 1.50 )</td>
<td>0.998</td>
<td>0.0795</td>
<td>0.0805</td>
</tr>
<tr>
<td>( \beta_1 = -.15 )</td>
<td>0.150</td>
<td>0.0171</td>
<td>0.0173</td>
</tr>
<tr>
<td>( \gamma = 0.25 )</td>
<td>0.247</td>
<td>0.0337</td>
<td>0.0339</td>
</tr>
</tbody>
</table>
Application to Sydney Asthma Count Data

Data: \( Y_1, \ldots, Y_{1461} \) daily asthma presentations in a Campbelltown hospital.

Preliminary analysis identified.

- no upward or downward trend

- a triple peaked annual cycle modelled by pairs of the form \( \cos(2\pi kt/365), \sin(2\pi kt/365), k=1,2,3,4 \).

- day of the week effect modelled by separate indicator variables for Sundays and Monday (increase in admittance on these days compared to Tues-Sat).

- Of the meteorological variables (max/min temp, humidity) and pollution variables (ozone, NO, NO\(_2\)), only humidity at lags of 12-20 days appears to have an association.
Model for Asthma Data

Trend function.

\[ x_t^T = (1, S_t, M_t, \cos(2\pi t/365), \sin(2\pi t/365), \cos(4\pi t/365), \sin(4\pi t/365), \cos(6\pi t/365), \sin(6\pi t/365), \cos(8\pi t/365), \sin(8\pi t/365)) \]

(No humidity used in this model.)

Model for \( \{v_t\} \).

\[ v_t = \left( \frac{1}{\phi(B)} - 1 \right) e_t , \text{ where } \phi(B) \text{ is the AR}(10) \text{ with autoregressive polynomial} \]

\[ \phi(B) = 1 - \phi_1 B - \phi_3 B^3 - \phi_7 B^7 - \phi_{10} B^{10}. \]

Note: the \( v_t \) can be computed recursively.
## Results for Asthma Data

<table>
<thead>
<tr>
<th>Term</th>
<th>Est</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.533</td>
<td>0.029</td>
</tr>
<tr>
<td>Sunday effect</td>
<td>0.240</td>
<td>0.054</td>
</tr>
<tr>
<td>Monday effect</td>
<td>0.249</td>
<td>0.054</td>
</tr>
<tr>
<td>$\cos(2\pi t/365)$</td>
<td>-0.162</td>
<td>0.036</td>
</tr>
<tr>
<td>$\sin(2\pi t/365)$</td>
<td>0.362</td>
<td>0.035</td>
</tr>
<tr>
<td>$\cos(4\pi t/365)$</td>
<td>-0.067</td>
<td>0.036</td>
</tr>
<tr>
<td>$\sin(4\pi t/365)$</td>
<td>0.023</td>
<td>0.034</td>
</tr>
<tr>
<td>$\cos(6\pi t/365)$</td>
<td>-0.083</td>
<td>0.035</td>
</tr>
<tr>
<td>$\sin(6\pi t/365)$</td>
<td>0.009</td>
<td>0.035</td>
</tr>
<tr>
<td>$\cos(8\pi t/365)$</td>
<td>-0.157</td>
<td>0.034</td>
</tr>
<tr>
<td>$\sin(8\pi t/365)$</td>
<td>-0.062</td>
<td>0.034</td>
</tr>
<tr>
<td>$\phi_1$</td>
<td>0.053</td>
<td>0.024</td>
</tr>
<tr>
<td>$\phi_3$</td>
<td>0.061</td>
<td>0.024</td>
</tr>
<tr>
<td>$\phi_7$</td>
<td>0.078</td>
<td>0.024</td>
</tr>
<tr>
<td>$\phi_{10}$</td>
<td>0.053</td>
<td>0.024</td>
</tr>
</tbody>
</table>
Asthma Data w/ Deterministic Part of Mean Fcn
Asthma Data: Deterministic Part + AR in Pearson Resid
GLARMA Extensions (Binary data)

Binary data: \( Y_1, \ldots, Y_n \)

Regression (explanatory) variable: \( x_t \)

Model: Distribution of the \( Y_t \) given \( x_t \) and the past is Bernoulli(\( p_t \)), i.e.,

\[
P(Y_t = 1) = p_t \quad \text{and} \quad P(Y_t = 0) = 1 - p_t.
\]

As before construct a MGD sequence

\[
e_t = (Y_t - p_t) / (p_t (1 - p_t))^{1/2}
\]

and using the logistic link function, the GLARMA model becomes

\[
W_t = \log \frac{p_t}{1 - p_t} \quad \text{with} \quad W_t = x_t^T \beta + Z_t,
\]

and

\[
Z_t = \hat{U}_t = \phi_1 (Z_{t-1} + e_{t-1}) + \cdots + \phi_p (Z_{t-p} + e_{t-p}) + \theta_1 e_{t-1} + \cdots + \theta_q e_{t-q}.
\]
A Simple GLARMA Model for Price Activity (R&S)

Model for price change: The price change $C_i$ of the $i^{th}$ transaction has the following components:

- $Y_t$ activity \{0,1\}
- $D_t$ direction \{-1,1\}
- $S_t$ size \{1, 2, 3, \ldots\}

Rydberg and Shephard consider a model for these components. An autologistic model is used for $Y_t$.

Simple GLARMA(0,1) model for price activity: $Y_t$ is a Bernoulli rv representing a price change at the $i^{th}$ transaction. Assume $Y_t$ given $F_{t-1}$ is Bernoulli($p_t$), i.e.,

$$P(Y_t = 1 \mid F_{t-1}) = p_t = 1 - P(Y_t = 0 \mid F_{t-1}),$$

where

$$p_t = \frac{e^{\sigma U_t}}{1 + e^{\sigma U_t}}$$

and

$$Z_t = \frac{Y_{t-1} - p_{t-1}}{\sqrt{p_{t-1}(1 - p_{t-1})}} = e_{t-1}.$$
Existence of Stationary for the Simple GLARMA Model

Consider the process

\[ Z_t = \frac{Y_{t-1} - p_{t-1}}{\sqrt{p_{t-1}(1 - p_{t-1})}} , \]

where \( Y_{t-1} \) is Bernoulli with parameter \( p_t = e^{\sigma Z_t} (1 + e^{\sigma Z_t})^{-1} \).

Propostion: The Markov process \( \{Z_t\} \) has a unique stationary distribution.

Idea of proof:

- \( \{Z_t\} \) is an e-chain.
- \( \{Z_t\} \) is bounded in probability on uniformly over the state space
- Possesses a reachable point ( \( x^* \) is soln to \( x + e^{\sigma x/2} = 0 \) )
Consider the model of a price of an asset at time $t$ given by

$$p(t) = p(0) + \sum_{i=1}^{N(t)} Z_i,$$

where

- $N(t)$ is the number of trades up to time $t$
- $Z_i$ is the price change of the $i^{th}$ transaction.

Then for a fixed time period $\Delta$,

$$p_t := p((t + 1)\Delta -) - p(t\Delta) = \sum_{i=N(t\Delta)+1}^{N((t+1)\Delta -)} Z_i,$$

denotes the rate of return on the investment during the $t^{th}$ time interval and

$$N_t := N((t + 1)\Delta -) - N(t\Delta)$$

denotes the number of trades in $[t \Delta, (t+1) \Delta)$. 
The Bin Model for the Number of Trades

Bin(p,q) model: The distribution of the number of trades $N_t$ in $[t \Delta, (t+1) \Delta)$, conditional on information up to time $t \Delta$ is Poisson with mean

$$\lambda_t = \alpha + \sum_{j=1}^{p} \gamma_j N_{t-j} + \sum_{j=1}^{q} \delta_j \lambda_{t-j}, \alpha \geq 0, 0 \leq \gamma_j, \delta_j < 1.$$ 

Proposition: For the Bin(1,1) model,

$$\lambda_t = \alpha + \gamma N_{t-1} + \delta \lambda_{t-1},$$

there exists a unique stationary solution.

Idea of proof:

• $\{\lambda_t\}$ is an e-chain.

• $\{\lambda_t\}$ is bounded in probability on average.

• Possesses a reachable point ($x^* = \alpha/(1-\gamma)$)
The observation model for the Poisson counts proposed here is

1. Easily interpretable on the linear predictor scale and on the scale of the mean $\mu_t$ with the regression parameters directly interpretable as the amount by which the mean of the count process at time $t$ will change for a unit change in the regressor variable.

2. An approximately unbiased plot of the $\mu_t$ can be generated by

$$\hat{\mu}_t = \exp(\hat{W}_t - .5 \sum_{i=1}^{\infty} \psi_i^2).$$

3. Is easy to predict with.

4. Provides a mechanism for adjusting the inference about the regression parameter $\beta$ for a form of serial dependence.

5. Generalizes to ARMA type lag structure.

6. Estimation (approx MLE) is easy to carry out.