Observation Driven Models for Time Series of Counts

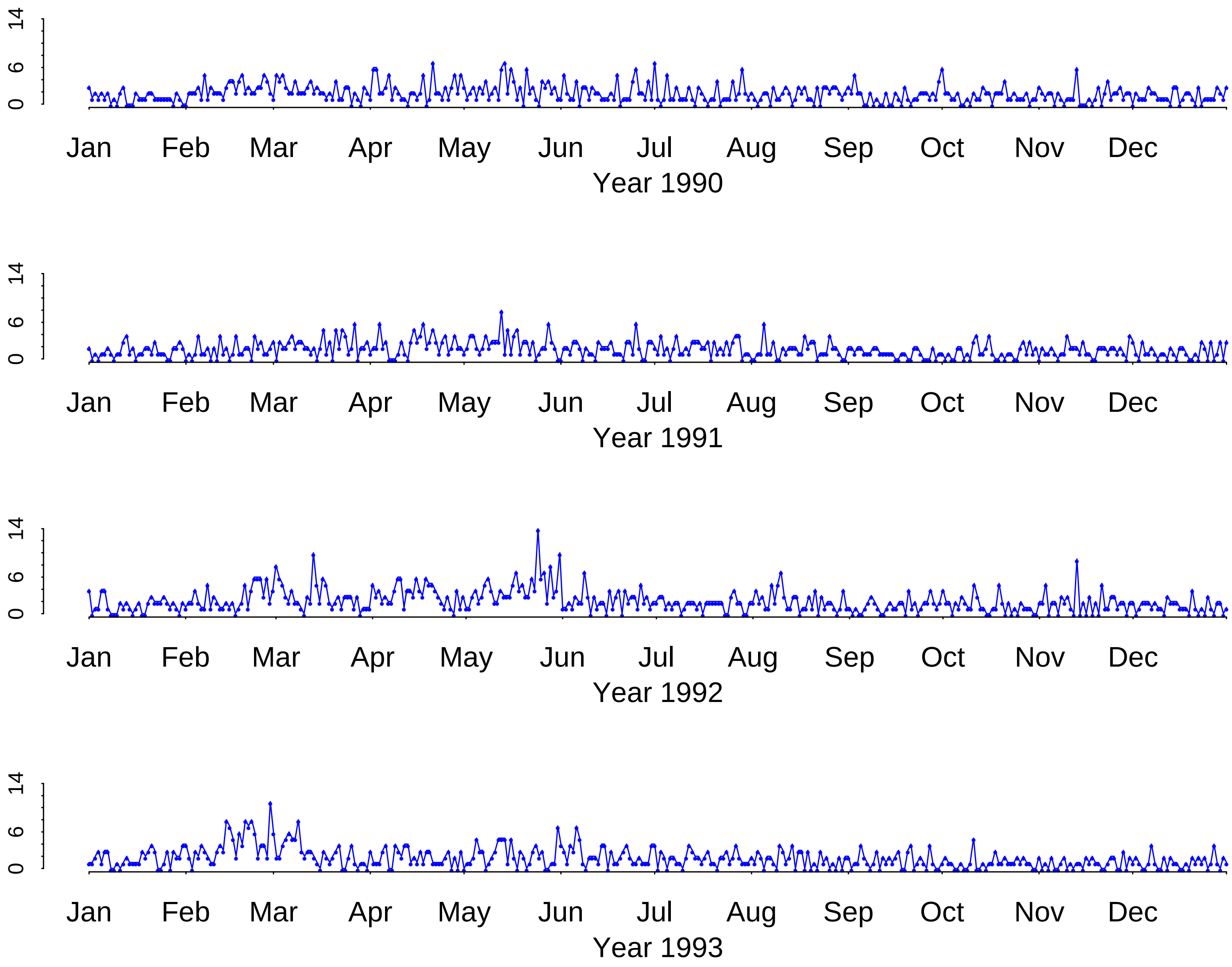
Richard A. Davis
Colorado State University
(http://www.stat.colostate.edu/~rdavis/lectures)

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

Outline

- Introduction
  - Example: asthma data
- Parameter-driven models
- Observation-driven models
- Generalized Linear ARMA (GLARMA) models for time series of counts
  - Properties
  - Existence and uniqueness of stationary distributions
  - Maximum likelihood estimation and asymptotic theory
  - Application to asthma data
- GLARMA extensions
  - Bernoulli
- Other (BIN)

Example: Daily Asthma Presentations (1990:1993)



Notation and Setup

Count data: $Y_1, \ldots, Y_n$

Regression (explanatory) variable: $\mathbf{x}_t$

Model: Distribution of the Y_t given $\mathbf{x}_t$ and a stochastic process ν_t are indep Poisson distributed with mean

$$\mu_t = \exp(\mathbf{x}_t^T \boldsymbol{\beta} + \nu_t).$$

The distribution of the stochastic process ν_t may depend on a vector of parameters $\boldsymbol{\gamma}$.

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

Primary objective: Inference about $\boldsymbol{\beta}$.

Parameter-Driven Model for the Mean Function μ_t

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

$$\log \mu_t = \mathbf{x}_t^T \boldsymbol{\beta} + \nu_t ,$$

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

e.g. (AR(1) process)

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

Advantages:

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

$$E(Y_t) = \exp(\mathbf{x}_t^T \boldsymbol{\beta})E\exp(\nu_t) = \exp(\mathbf{x}_t^T \boldsymbol{\beta}), \text{ if } E\exp(\nu_t) = 1.$$

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 μ_t

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

$$\log \mu_t = \mathbf{x}_t^T \boldsymbol{\beta} + \nu_t \; ,$$

where ν_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 ergodicty, is difficult to derive.
- $\mathbf{x}_t^T \boldsymbol{\beta}$ is not easily interpretable. In the special case above,

$$E(Y_t) = \exp(\mathbf{x}_t^T \boldsymbol{\beta})E\exp(\gamma_1 Y_{t-1} + \ldots + \gamma_p Y_{t-p})$$

Generalized Linear ARMA (GLARMA) Model for Poisson Counts

Two components in the specification of ν_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 λ will be described later.)

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

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

where

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

Since the conditional mean μ_t is based on the whole past, the model is no longer Markov.

Properties of the New Model

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

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. $E(e_t^2) = E(\mu_t^{1-2\lambda})$

   $= 1$ if $\lambda = .5$

4. Set, $W_t = \log \mu_t = x_t^T \boldsymbol{\beta} + \nu_t,$

so that

$$E(W_t) = x_t^T \boldsymbol{\beta} \qquad \text{and} \qquad \mathrm{Var}(W_t) = \sum_{i=1}^{\infty} \psi_i^2 E(\mu_{t-i}^{1-2\lambda})$$

$$= \sum_{i=1}^{\infty} \psi_i^2 \quad (\text{if } \lambda = .5)$$

Properties continued

5. $$\mathrm{Cov}(W_t, W_{t+h}) = \sum_{i=1}^{\infty} \psi_i \psi_{i+h} \mathrm{E}(\mu_{t-i}^{1-2\lambda})$$

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

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

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

$$E(e^{W_t}) = E(e^{x_t^T \boldsymbol{\beta} + \sum_i \psi_i e_{t-i}})$$

$$\approx e^{x_t^T \boldsymbol{\beta} + Var(\nu_t)/2}$$

$$= e^{x_t^T \boldsymbol{\beta} + \sum_{i=1}^{\infty} \psi_i^2 / 2},$$

By adjusting the intercept term, $\mathrm{E}(\mu_t)$ can be interpreted as $\exp(\mathbf{x}_t^T \beta)$.

Properties continued

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

$$Z_t = \hat{U}_t = \phi_1(Z_{t-1} + e_{t-1}) + \cdots + \phi_p(Z_{t-p} + e_{t-p}) + \theta_1 e_{t-1} + \cdots + \theta_q e_{t-q}.$$

Existence and uniqueness 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 ν such for some $m > 1$, $\varepsilon > 0$, and $\delta > 0$,

  $\nu(A) > \varepsilon$ implies $P^m(x,A) \geq \delta$ 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 $.5 \le \lambda < 1$.

Consider the process

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

Propostion: 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 $\{P(x, \bullet),\ x \in X\}$ 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 $\{P(x, \bullet),\ x \in X\}$ using a Markov style inequality.

Uniqueness of Stationary Distr in Case $.5 \leq \lambda < 1$?

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^*,

$$\sum_{n=1}^{\infty} P^n(x, O) > 0 \quad \text{for all } x.$$

e-chain: For every continuous f with compact support, the sequence of functions $\{P^n f, \ n = 1, \ldots\}$ is equicontinuous, on compact sets.

Estimation for Poisson GLARMA

Let $\boldsymbol{\delta} = (\boldsymbol{\beta}^T, \boldsymbol{\gamma}^T)^T$ be the parameter vector for the model ($\boldsymbol{\gamma}$ corresponds to the parameters in the linear process part).

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

$$\log \mu_t = \mathrm{x}_t^T \boldsymbol{\beta} + \nu_t,$$

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

Log-likelihood:

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

where

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

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

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

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

Asymptotic Results for MLE

Define the array of random variables by

$$\eta_{nt} = n^{-1/2}(\mathrm{Y}_t - e^{\mathrm{W}_t(\boldsymbol{\delta})})\frac{\partial \mathrm{W}_t(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}}.$$

Properties of $\{\eta_{nt}\}$:

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

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

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

where $$V = \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} e^{W_t(\boldsymbol{\delta})} \partial W_t(\boldsymbol{\delta}) \partial W_t^T(\boldsymbol{\delta}).$$

Simulation Results

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

Parameter	Mean	SD	SD(from like)
$\beta_0 = 1.50$	1.499	0.0263	**0.0265**
$\gamma = 0.25$	0.249	0.0403	**0.0408**
$\beta_0 = 1.50$	1.499	0.0366	**0.0364**
$\gamma = 0.75$	0.750	0.0218	**0.0218**
$\beta_0 = 3.00$	3.000	0.0125	**0.0125**
$\gamma = 0.25$	0.249	0.0431	**0.0430**
$\beta_0 = 3.00$	3.000	0.0175	**0.0174**
$\gamma = 0.75$	0.750	0.0270	**0.0271**

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

$\beta_0 = 1.00$	1.000	0.0286	**0.0284**
$\beta_1 = 0.50$	0.500	0.0035	**0.0034**
$\gamma = 0.25$	0.248	0.0420	**0.0426**
$\beta_0 = 1.50$	0.998	0.0795	**0.0805**
$\beta_1 = -.15$	-.150	0.0171	**0.0173**
$\gamma = 0.25$	0.247	0.0337	**0.0339**

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
- annual cycle modeled by cos(2πt/365), sin(2πt/365)
- seasonal effect modeled by

$$P_{ij}(t) = \frac{1}{B(2.5,5)}\left(\frac{t-T_{ij}}{100}\right)^{2.5}\left(1-\frac{t-T_{ij}}{100}\right)^{5}$$

  where *B(2.5,5)* is the beta function and T_{ij} is the start of the j^{th} school term in year *i*.

- day of the week effect modeled by separate indicator variables for Sunday 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 and NO_2(max) appear to have an association.

Model for Asthma Data

Trend function.

$\mathbf{x}_t^T = (1, S_t, M_t, \cos(2\pi t/365), \sin(2\pi t/365), P_{11}(t), P_{12}(t), P_{21}(t), P_{22}(t), P_{31}(t), P_{32}(t), P_{41}(t), P_{42}(t), H_t, N_t)$

($H_t = \frac{1}{7}\sum_{i=0}^{6} h_{t-12-i}$ and h_t is the residual from an annual cycle fitted to the daily average humidity at 0900 and 1500.)

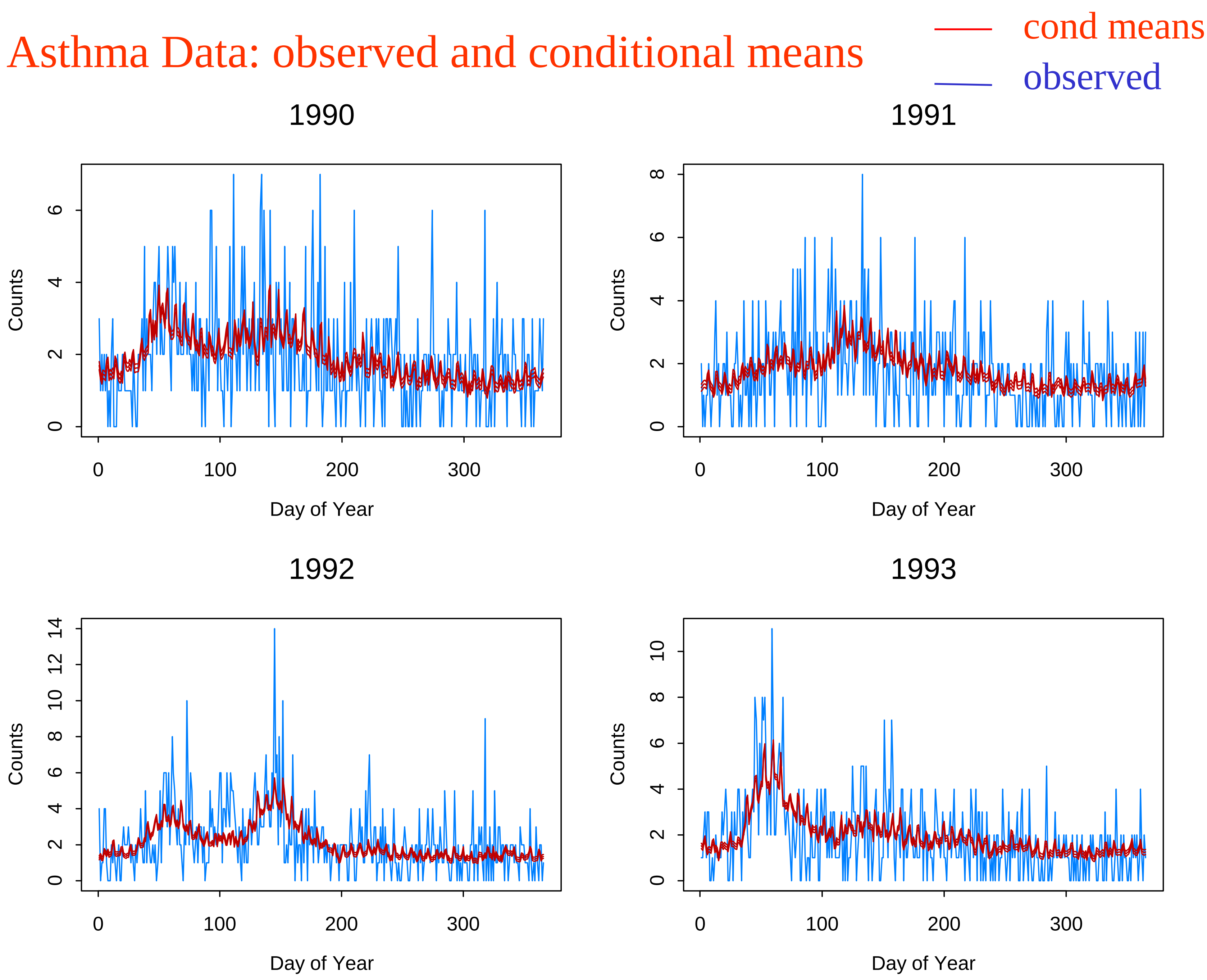
Model for $\{\nu_t\}$.

MA(7): $\nu_t = \theta_7 e_{t-7}$

Results for Asthma Data

Term	Est	SE	T-ratio
Intercept	0.583	0.062	9.46
Sunday effect	0.197	0.056	3.53
Monday effect	0.230	0.055	4.20
$\cos(2\pi t/365)$	-0.214	0.039	-5.54
$\sin(2\pi t/365)$	0.176	0.040	4.35
Term 1, 1990	0.200	0.056	3.54
Term 2, 1990	0.132	0.057	2.31
Term 1, 1991	0.087	0.066	1.32
Term 2, 1991	0.172	0.057	2.99
Term 1, 1992	0.254	0.055	4.66
Term 2, 1992	0.308	0.049	6.31
Term 1, 1993	0.439	0.050	8.77
Term 2, 1993	0.116	0.061	1.91
Humidity $H_t/20$	0.169	0.055	3.09
NO_2 max	-0.104	0.033	-3.16
MA, lag 7 θ_7	0.042	0.018	2.32

Asthma Data: observed and conditional means



GLARMA Extensions (Binary data)

Binary data: $Y_1, \ldots, Y_n$

Regression (explanatory) variable: $\mathbf{x_t}$

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

$$P(Y_t = 1 | F_{t-1}) = p_t \text{ and } P(Y_t = 0 | F_{t-1}) = 1 - p_t.$$

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

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

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 t^{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

$$p_t = \frac{e^{\sigma Z_t}}{(1+e^{\sigma Z_t})} \quad \text{and} \quad Z_t = \frac{Y_{t-1} - p_{t-1}}{\sqrt{p_{t-1}(1-p_{t-1})}} = e_{t-1}.$$

Existence of Stationary Solns 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 $p_{t-1} = e^{\sigma Z_{t-1}}(1+e^{\sigma Z_{t-1}})^{-1}$.

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 $x+e^{\sigma x/2}=0$)

BIN Models: A Modeling Framework for Stock Prices
(Davis, Ryderberg, Shephard, Streett)

Consider the model of a price of an asset at time *t* given by

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

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 Δ,

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

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 $[t\,\Delta, (\mathrm{t}+1)\,\Delta)$.

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 ($x^* = \alpha/(1-\gamma)$)

Summary Remarks

The observation model for the Poisson counts proposed here has the following properties.

1. Easily interpretable on the linear predictor scale and on the scale of the mean μ_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 μ_t can be generated by

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

3. Is easy to predict with.

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

5. Generalizes to ARMA type lag structure.

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