Modeling Time Series of Counts

Richard A. Davis
Colorado State University

William Dunsmuir
University of New South Wales

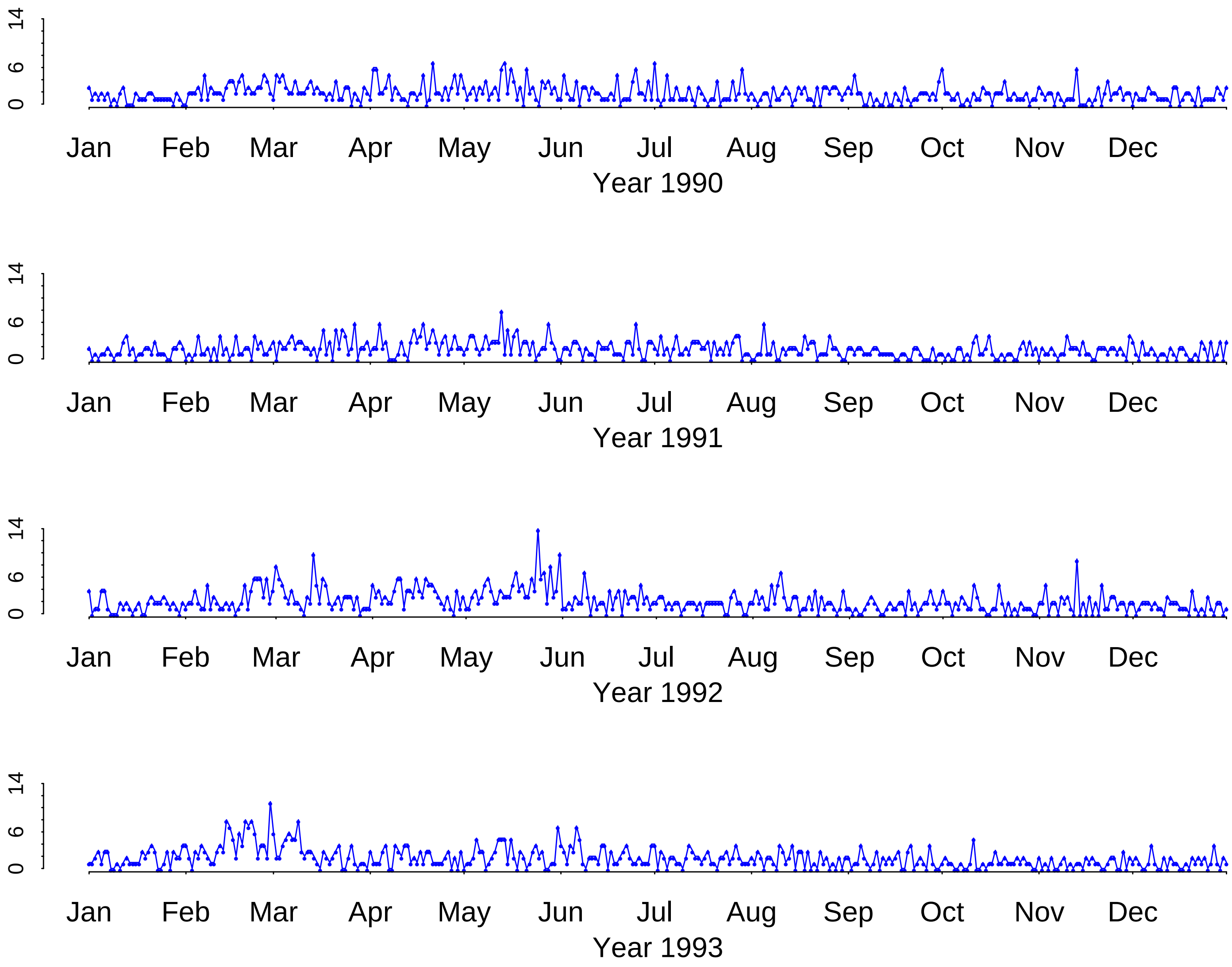
Sarah Streett
National Center for Atmospheric Research

(Other collaborators: Richard Tweedie, Ying Wang)

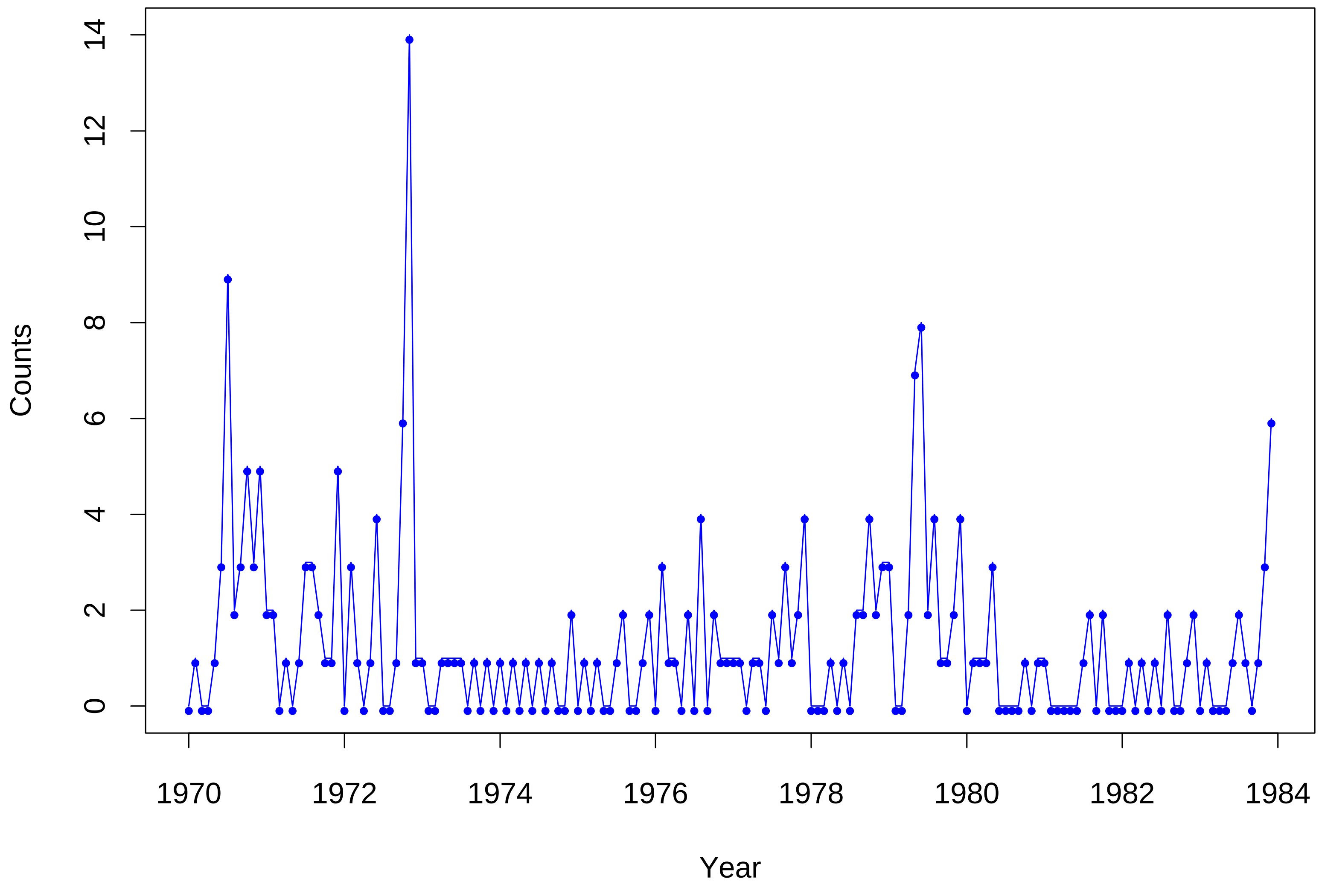
Outline

- Introduction
  - Examples
- Linear regression model
- Parameter-driven models
  - Poisson regression with serial dependence
  - Theory for GLM estimates
- Observation-driven models
  - Properties
  - Existence and uniqueness of stationary distributions
  - Model for stock prices (number of trades and price activity)
  - Estimation and asymptotic theory for MLE
  - Application to asthma data

Example: Daily Asthma Presentations (1990:1993)



Example: Monthly Polio Counts in USA (Zeger 1988)



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 $\boldsymbol{\beta}$.

Example: Polio (cont)

Regression function:

$$\mathbf{x}_t^T = (1,\ t'/1000,\ \cos(2\pi t'/12),\ \sin(2\pi t'/12),\ \cos(2\pi t'/6),\ \sin(2\pi t'/6))$$

where $t' = (t-73)$.

Summary of various models fits to Polio data:

Study	Trend(β)	SE(β)	t-ratio
GLM Estimate	-4.80	1.40	-3.43
Zeger (1988)	-4.35	2.68	-1.62
Chan and Ledolter (1995)	-4.62	1.38	-3.35
Kuk&Chen (1996) MCNR	-3.79	2.95	-1.28
Jorgensen et al (1995)	-1.64	.018	-91.1
Fahrmeir and Tutz (1994)	-3.33	2.00	-1.67

Linear Regression Model-A Review

Suppose $\{Y_t\}$ follows the linear model with time series errors given by

$$Y_t = \mathbf{x}_t^T \boldsymbol{\beta} + W_t ,$$

where $\{W_t\}$ is a stationary (ARMA) time series.

- Estimate $\boldsymbol{\beta}$ by ordinary least squares (OLS).
- OLS estimate has same asymptotic efficiency as MLE.
- Asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}_{OLS}$ depends on ARMA parameters.
- Identify and estimate ARMA parameters using the estimated residuals,

$$W_t = Y_t - \mathbf{x}_t^T \hat{\boldsymbol{\beta}}_{OLS}$$

- Re-estimate $\boldsymbol{\beta}$ and ARMA parameters using full MLE.

GLM Estimation

Model: $Y_t \mid \nu_t, \mathbf{x}_t \sim P(\exp(\mathbf{x}_t^T \beta + \nu_t))$.

GLM log-likelihood:

$$l(\beta) = -\sum_{t=1}^{n} e^{\mathbf{x}_t^T \beta} + \sum_{t=1}^{n} Y_t \mathbf{x}_t^T \beta - \log\left[\prod_{t=1}^{n} Y_t!\right]$$

(Likelihood ignores presence of the latent process.)

Assumptions on regressors:

$$\Omega_{I,n} = n^{-1} \sum_{t=1}^{n} \mathbf{x}_t \mathbf{x}_t^T \mu_t \to \Omega_I(\beta),$$

$$\Omega_{II,n} = n^{-1} \sum_{t=1}^{n} \sum_{s=1}^{n} \mathbf{x}_t \mathbf{x}_s^T \mu_t \mu_s \gamma_\varepsilon(s-t) \to \Omega_{II}(\beta),$$

Theorem for GLM Estimates

Theorem (Davis, Dunsmuir, Wang `00). Let $\hat{\beta}$ be the GLM estimate of β obtained by maximizing $l(\beta)$ for the Poisson regression model with a stationary lognormal latent process. Then

$$n^{1/2}(\hat{\beta}-\beta) \xrightarrow{d} \mathrm{N}(0, \Omega_I^{-1} + \Omega_I^{-1}\Omega_{\mathrm{II}}\,\Omega_I^{-1}).$$

Notes:

1. $n^{-1}\Omega_I^{-1}$ is the asymptotic cov matrix from a std GLM analysis.

2. $n^{-1}\Omega_I^{-1}\,\Omega_{II}\,\Omega_I^{-1}$ is the additional contribution due to the presence of the latent process.

3. Result also valid for more general latent processes (mixing, etc),

4. Can have $\mathbf{x}_t$ depend on the sample size n.

When does CLT Apply?

Conditions on the regressors hold for:

1. Trend functions.

$$\mathbf{x}_{\mathbf{n}t} = \mathbf{f}(t/n)$$

where $\mathbf{f}$ is a continuous function on [0,1]. In this case,

$$n^{-1}\sum_{t=1}^{n}\mathbf{x}_{\mathbf{t}}\mathbf{x}_{\mathbf{t}}^{\mathbf{T}}\mu_t \to \int_0^1 \mathbf{f}(t)\mathbf{f}^{\mathbf{T}}(t)e^{\mathbf{f}^{\mathbf{T}}(t)\beta}dt,$$

$$n^{-1}\sum_{t=1}^{n}\sum_{s=1}^{n}\mathbf{x}_{\mathbf{t}}\mathbf{x}_{\mathbf{s}}^{\mathbf{T}}\mu_t\mu_s\gamma_\varepsilon(s-t) \to \int_0^1 \mathbf{f}(t)\mathbf{f}^{\mathbf{T}}(t)e^{2\mathbf{f}^{\mathbf{T}}(t)\beta}dt\sum_h \gamma_\varepsilon(h)\,.$$

Remark. $\mathbf{x}_{\mathbf{nt}}$ = (1, t/n) corresponds to linear regression and works. However $\mathbf{x}_{\mathbf{t}}$ = (1, t) does **not** produce consistent estimates say if the true slope is negative.

When does CLT apply? (cont)

2. Harmonic functions to specify annual or weekly effects, e.g.,

$$x_t = \cos(2\pi t/7)$$

3. Stationary process. (e.g. seasonally adjusted temperature series.)

Application to Model for Polio Data

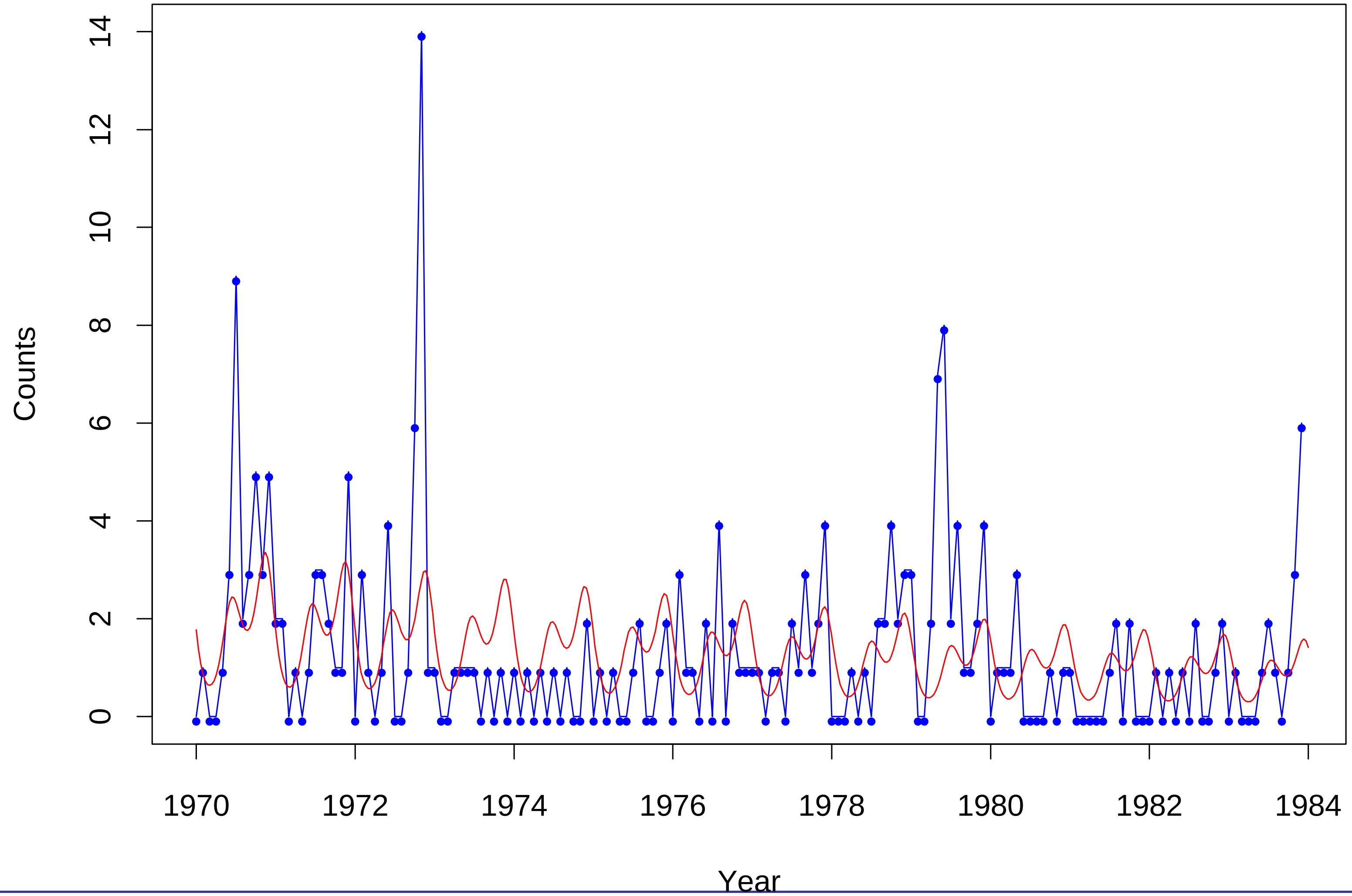
Use the same regression function as before. Assume the $\{\nu_t\}$ follows a log-normal AR(1), where

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

with $\phi = .82$, $\sigma^2 = .57$.

	Zeger		GLM Fit		Asym	Simulation	
	$\hat\beta_Z$	s.e.	$\hat\beta_{GLM}$	s.e.	s.e.	$\hat\beta_{GLM}$	s.d.
Intercept	0.17	0.13	.207	.075	.205	.150	.213
Trend($\times 10^{-3}$)	-4.35	2.68	-4.80	1.40	4.12	-4.89	3.94
$\cos(2\pi t/12)$	-0.11	0.16	-0.15	.097	.157	-.145	.144
$\sin(2\pi t/12)$	-.048	0.17	-0.53	.109	.168	-.531	.168
$\cos(2\pi t/6)$	0.20	0.14	.169	.098	.122	.167	.123
$\sin(2\pi t/6)$	-0.41	0.14	-.432	.101	.125	-.440	.125

Polio Data With Estimated Regression Function



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 hard.

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:

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

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,

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

New Observation Driven Model

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 $\{e_t\}$

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

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. Nevertheless, this specification could lead to stationary solutions, although the stability theory appears difficult.

Properties of the New Model

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

1. $E(e_t \mid F_{t-1}) = 0$
2. $E(e_t^2) = E(\mu_t^{1-2\lambda})$

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

3. Set,

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

so that

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

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

Properties continued

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

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

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

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

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

It follows that the intercept term can be adjusted in order for $E(\mu_t)$ to be interpretable as $\exp(\mathbf{x}_t^T \beta)$.

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

  $\nu(A) > \varepsilon$ implies $P^m(x,A) \geq \delta$.

- 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 $.5 \le \lambda < 1$?

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^*,

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

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

Modeling Framework for Stock Prices (Rydberg & Shephard)

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 $[t\,\Delta, (t+1)\,\Delta)$.

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

A Simple GLARMA Model for Price Activity (R&S)

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

- A_t activity {0,1}
- D_t direction {-1,1}
- S_t size {1, 2, 3, . . .}

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

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

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

where

$$p_t = \frac{e^{\sigma U_t}}{(1+e^{\sigma U_t})} \text{ and } U_t = \frac{A_{t-1} - p_{t-1}}{\sqrt{p_{t-1}(1-p_{t-1})}}.$$

Existence of Stationary for the Simple GLARMA Model .

Consider the process

$$U_t = \frac{A_{t-1} - p_{t-1}}{\sqrt{p_{t-1}(1-p_{t-1})}},$$

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

Propostion: The Markov process $\{U_t\}$ has a unique stationary distribution.

Idea of proof:

- $\{U_t\}$ is an e-chain.
- $\{U_t\}$ is bounded in probability on uniformly over the state space
- Possesses a reachable point (x^* is soln to $x+e^{\sigma x/2}=0$)

Estimation for Poisson Observation Driven Model

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

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

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

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

Log-likelihood:

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

where

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

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.

Asymptotic Results for MLE

Define the array of random variables by

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

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

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

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

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

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, 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
- 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.

$\mathbf{x_t^T}$=(1, S_t, M_t, cos(2πt/365), sin(2πt/365), cos(4πt/365), sin(4πt/365),

cos(6πt/365), sin(6πt/365), cos(8πt/365), sin(8πt/365))

(No humidity used in this model.)

Model for $\{\nu_t\}$.

$\nu_t = (1/\phi(B) - 1)\, e_t$, where $\phi(B)$ is the AR(10) with autoregressive polynomial

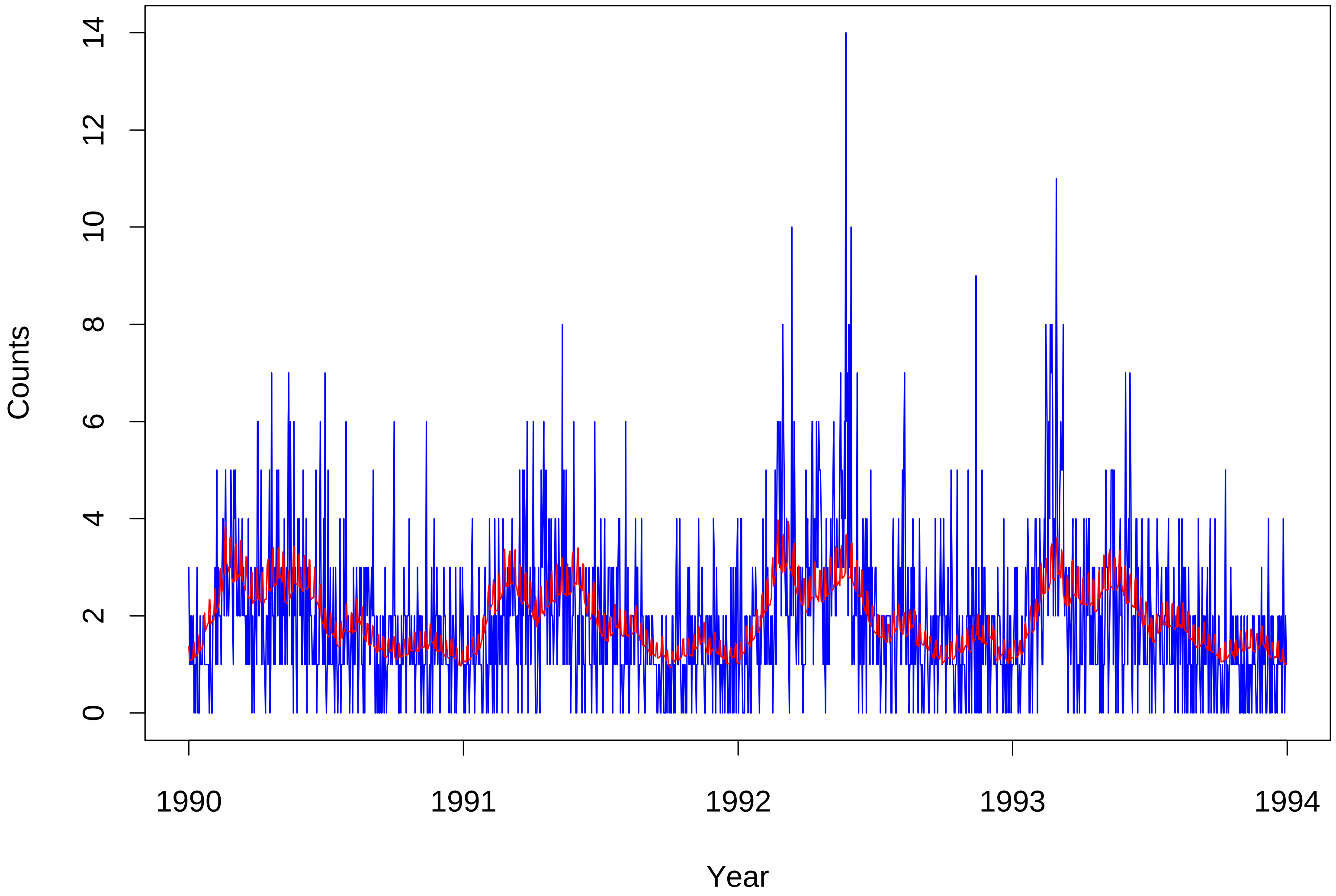
$$\phi(B) = 1 - \phi_1 B - \phi_3 B^3 - \phi_7 B^7 - \phi_{10} B^{10}.$$

Note: the ν_t can be computed recursively.

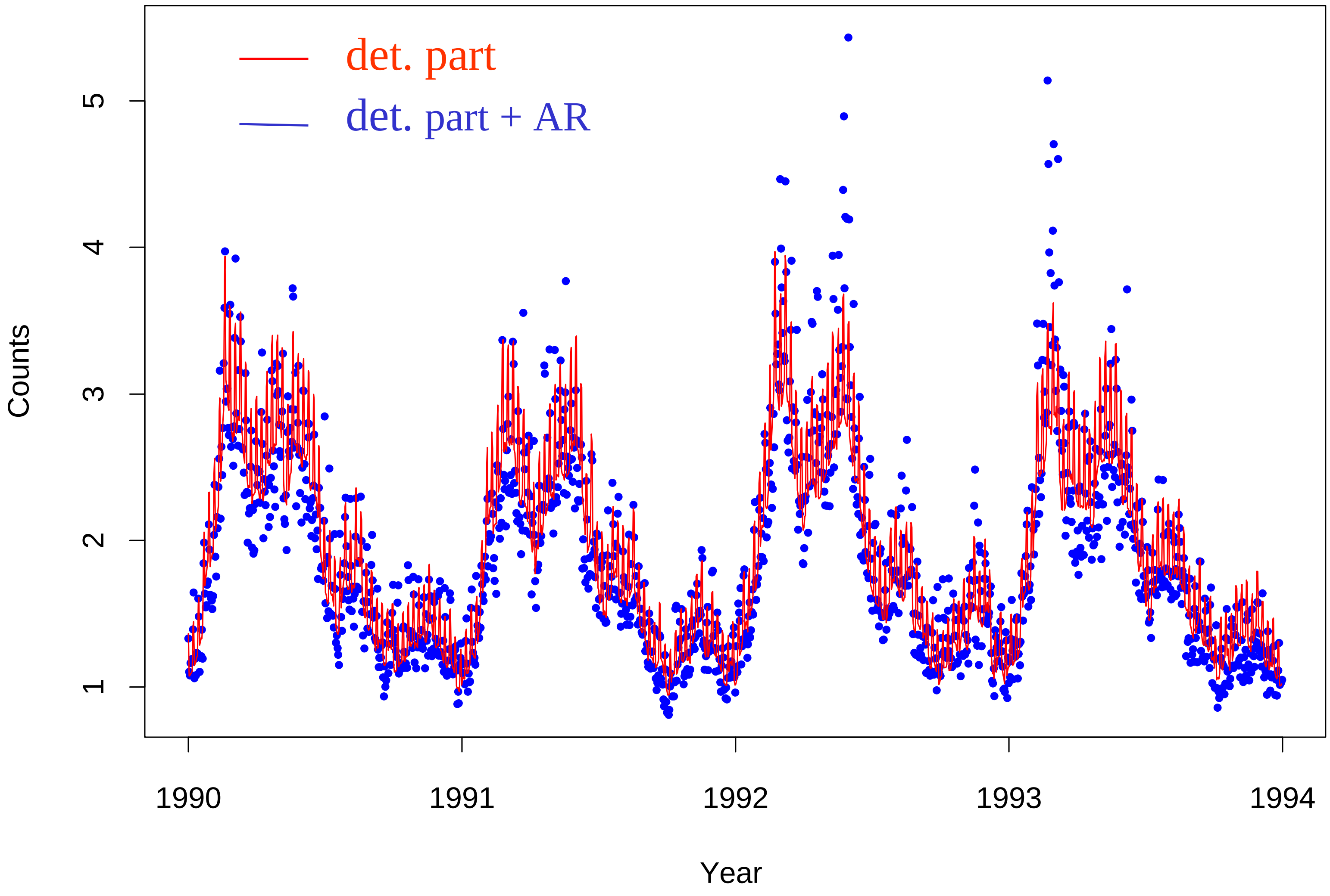
Results for Asthma Data

Term	Est	SE
Intercept	0.533	0.029
Sunday effect	0.240	0.054
Monday effect	0.249	0.054
$\cos(2\pi t/365)$	-0.162	0.036
$\sin(2\pi t/365)$	0.362	0.035
$\cos(4\pi t/365)$	-0.067	0.036
$\sin(4\pi t/365)$	0.023	0.034
$\cos(6\pi t/365)$	-0.083	0.035
$\sin(6\pi t/365)$	0.009	0.035
$\cos(8\pi t/365)$	-0.157	0.034
$\sin(8\pi t/365)$	-0.062	0.034
ϕ_1	0.053	0.024
ϕ_3	0.061	0.024
ϕ_7	0.078	0.024
ϕ_{10}	0.053	0.024

Asthma Data w/ Deterministic Part of Mean Fcn



Asthma Data: Deterministic Part + AR in Pearson Resid



Summary Remarks

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 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.