Modelling Time Series of Counts

Richard A. Davis
Colorado State University

William Dunsmuir
University of New South Wales

Ying Wang
Colorado State University

Two Types of Models for Poisson Counts

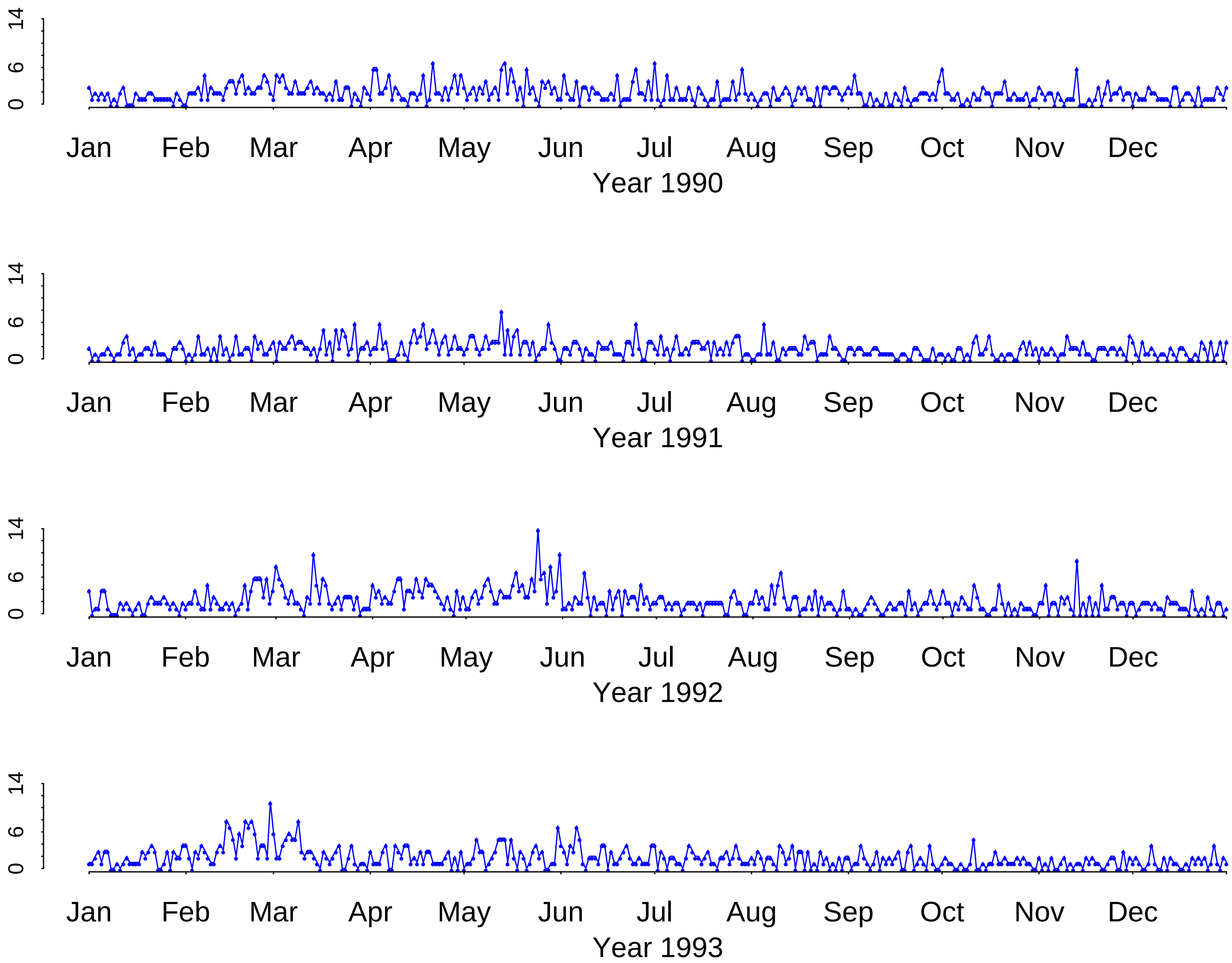
☞ Parameter-driven models

- Poisson regression when serial dependence
- Testing for a latent process
- Estimating serial dependence
- Fitting latent processes

☞ Observation-driven models

- Fitting, distribution, and standard errors
- Application to asthma data

Example: Daily Asthma Presentations (1990:1993)

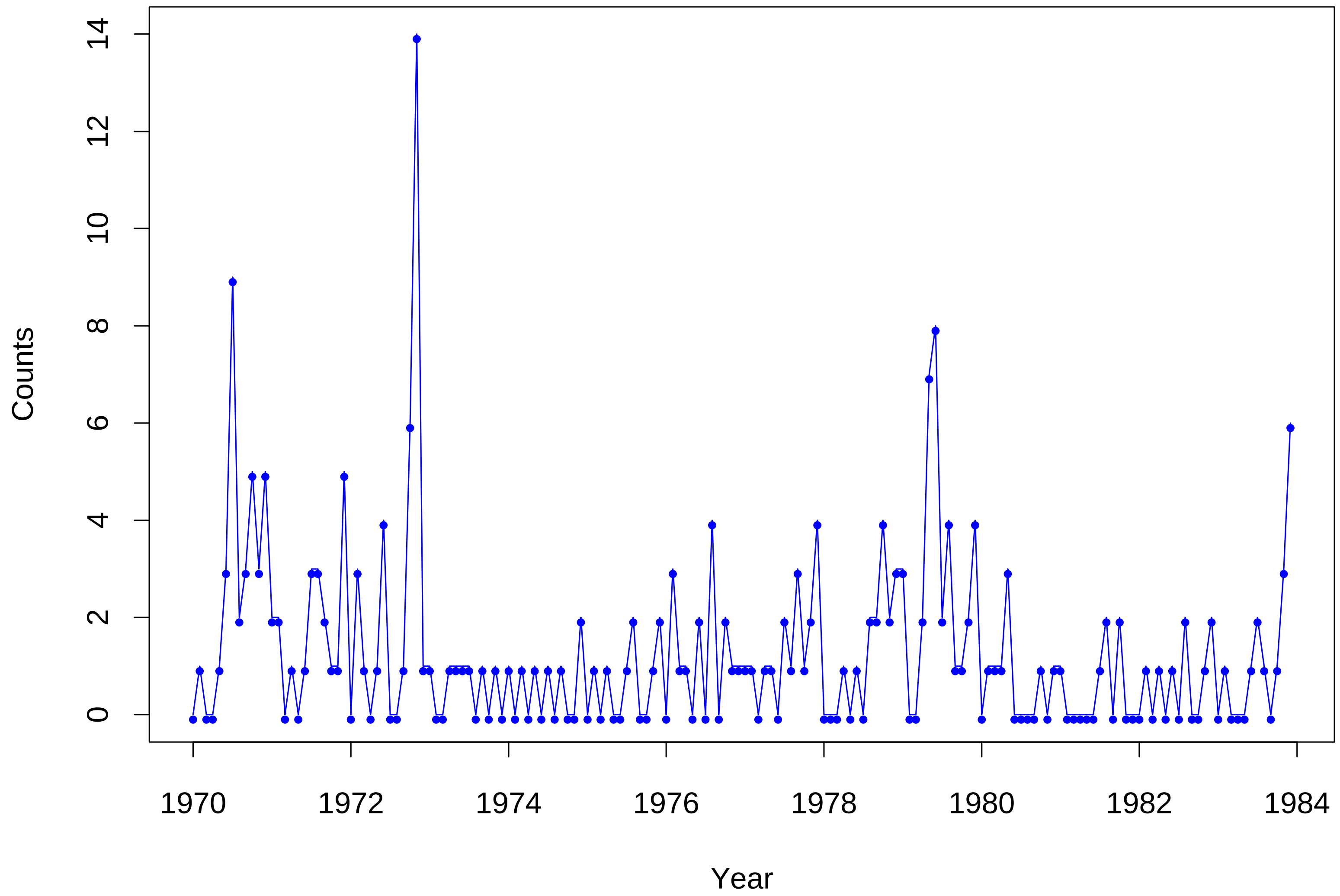


Polio Incidence in the U.S.

The rate of polio infection dropped dramatically following the inactivated polio vaccine (IPV) introduction in 1955. The decline continued following the introduction of live oral polio vaccine (OPV) in 1961. In 1960, there were 2,525 cases of paralytic polio reported in the United States, and in 1965 there were only 61. Between 1980 and 1990 an average of 8 cases were reported per year, most of which were vaccine associated. Since 1979 there has not been a single case of polio caused by wild virus in the United States and only an average of one imported case per year.

CENTER FOR DISEASE CONTROL AND PREVENTION
Date Last Revised: March 9, 1995

Example: Polio Counts (Zeger 1988)



Notation and Setup

Count Data: $Y_1, \ldots, Y_n$

Regression 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

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

The distribution of the stochastic process ν_t may depend on a vector of parameters γ.

Note: If $\nu_t = 0$, then in standard Poisson regression model.

Objective: Inference about β.

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.

Example: Polio (cont)

Regression function:

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

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

Desiderata for models - Zeger and Qaqish

Zeger & Qaqish (1988) offer 3 desiderata that should be met.

1. Ease of interpretation. Marginal mean of Y_t should be approximately

$$E(Y_t) = \mu_t = \exp(\mathbf{x}_t^T\beta)$$

(regression coefficient β can be interpreted as the proportional change in the marginal expectation of Y_t given a unit change in $\mathbf{x}_t$)

2. Flexibility. Both positive and negative serial correlation should be possible in the model.

3. Orthogonality of the estimates of β and γ. (Enables implementation of a 2-stage estimation procedure?)

Desiderata for models - continued

Condition 3 is met for linear regression models with time series errors. For count data, this condition may be overly restrictive since the mean and variance of Y_t are linked.

4. Ease of producing forecasts. Often this is primary goal of time series modelling.

5. Procedures for model fitting and inference.

6. Diagnostic tools. Required for assessing model adequacy.

Latent Process or Parameter Driven Model

Count Data: $Y_1, \ldots, Y_n$

Conditional distribution of Y_t given $\mathbf{x_t}$ and a non-negative stochastic process ε_t , is Poisson distributed with mean $\varepsilon_t \exp(\mathbf{x}_t^T \boldsymbol{\beta})$, i.e.,

$$Y_t \mid \varepsilon_t , \mathbf{x}_t \sim P(\varepsilon_t \exp(\mathbf{x}_t^T \boldsymbol{\beta})).$$

Note: $\mathrm{E}\, Y_t = \exp(\mathbf{x}_t^T \boldsymbol{\beta})\mathrm{E}\, \varepsilon_t$. We assume $\mathrm{E}\, \varepsilon_t = 1$ for identification purposes.

Assumptions on latent process: $\{\varepsilon_t\}$ is a non-negative stationary time series with mean 1 and ACVF

$$\gamma_\varepsilon(h) = \mathrm{E}(\varepsilon_{t+h}-1)(\varepsilon_t-1).$$

Often assume $\varepsilon_t = \exp(\alpha_t)$, where $\{\alpha_t\}$ is a stationary Gaussian T.S.

($\alpha_t \sim \mathrm{N}(-\sigma_\alpha^2/2, \sigma_\alpha^2)$)

Moment Properties of the Poisson Count Process

Mean of Y_t:

$$\mu_t = E(Y_t) = \exp(\mathbf{x}_t^T \beta)$$

Variance of Y_t:

$$\mathrm{Var}(Y_t) = \mu_t + \mu_t^2 \sigma_\varepsilon^2$$

Autocovariance function of Y_t:

$$\mathrm{Cov}(Y_{t+h}, Y_t) = \mu_t \, \mu_{t+h} \, \gamma_\varepsilon (h) \, .$$

Autocorrelation function of Y_t:

$$\mathrm{Cor}(Y_{t+h}, Y_t) = \rho_\varepsilon (h) / ((1+ \mu_t^{-1} \sigma_\varepsilon^{-2})(1+ \mu_{t+h}^{-1} \sigma_\varepsilon^{-2}))^{1/2}$$

Special case $\mathbf{x}_t = 1$ and $\varepsilon_t = \exp(\alpha_t)$:

$$0 \le \mathrm{Cor}(Y_{t+h}, Y_t) \le \rho_\alpha (h),$$

Implication: difficult to detect correlation in latent process from Y_t

GLM Estimates

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

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:

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

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

Theorem for GLM Estimates

Theorem. Let $\hat{\beta}$ be the GLM estimate of $\boldsymbol{\beta}$ 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 **f** is a continuous function on [0,1]. In this case,

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

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

Remark. $\mathbf{x_{nt}}$ = (1, t/n) corresponds to linear regression and works. However $\mathbf{x_t}$ = (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 Polio Data

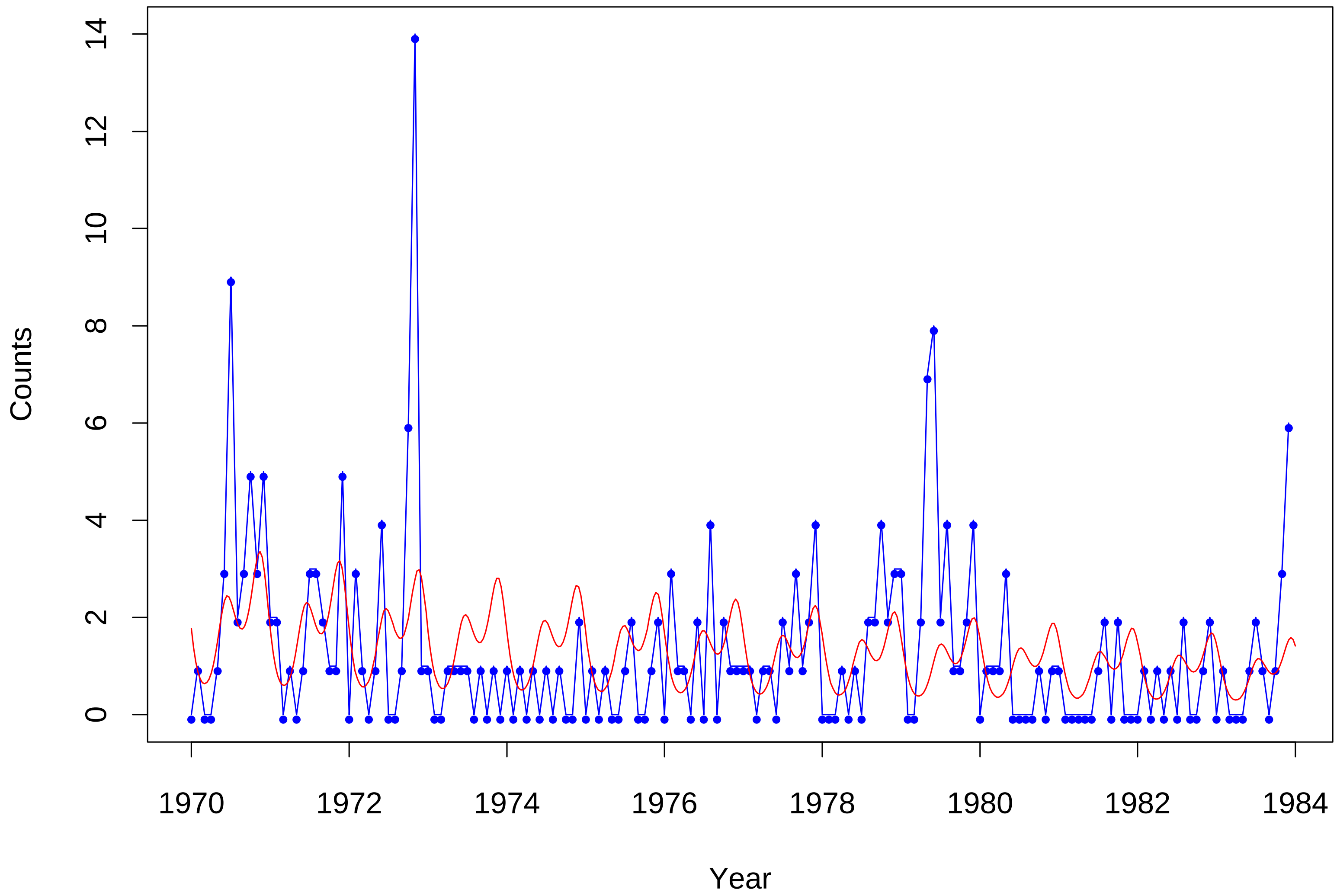
Use the same regression function as before. Assume latent process is a log-normal AR(1), i.e., $\ln \varepsilon_t = \alpha_t$, where

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

with ϕ =.82, σ^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



Testing for the Existence of a Latent Process

Under H_0: no latent process (i.e., $\varepsilon_t \equiv 1$), the Pearson residuals

$$e_t = \frac{Y_t - \hat{\mu}_t}{\sqrt{\hat{\mu}_t}}$$

are approx IID N(0,1). Test statistic

$$Q = \left(n^{-1} \sum_{t=1}^{n} e_t^2 - 1 \right) / \hat{\sigma}_Q, \quad \hat{\sigma}_Q^2 = n^{-1} \left(n^{-1} \sum_{t=1}^{n} \hat{\mu}_t^{-1} + 2 \right),$$

has an approx N(0,1) distribution. Test does not perform well.

α	.100	.050	.025
$P(Q>z_{1-\alpha})$	.036	.010	.004

Adjustments to Test Statistic

Standardized Pearson residuals: $\tilde{e}_t = \dfrac{Y_t - \hat{\mu}_t}{\sqrt{\hat{\mu}_t (1 - h_t)}}$,

where h_t is the t^{th} diagonal value of the "hat" matrix.

Brannas and Johansson (1994) test statistic: based on a local alternative hypothesis against a neg binomial alternative.

$$S_a = \frac{\sum_{t=1}^{n} \left[(Y_t - \hat{\mu}_t)^2 - Y_t + \hat{h}_t \hat{\mu}_t \right]}{\left[2 \sum_{t=1}^{n} \hat{\mu}_t^2 \right]^{1/2}}$$

(S_a is the version adapted by Dean and Lawless (1989) and generally worked best.)

Zeger's estimates of autocovariances

Zeger (1988) proposed the following estimates of the ACVF of the latent process

$$\hat{\sigma}^2_{\varepsilon,Z} = \sum_{t=1}^{n}\left[(Y_t - \hat{\mu}_t)^2 - \hat{\mu}_t\right] / \sum_{t=1}^{n}\hat{\mu}_t^2$$

$$\hat{\gamma}_{\varepsilon,Z}(h) = \sum_{t=1}^{n-h}(Y_t - \hat{\mu}_t)(Y_{t+h} - \hat{\mu}_{t+h}) / \sum_{t=1}^{n-h}\hat{\mu}_t\hat{\mu}_{t+h}$$

$$\hat{\rho}_{\varepsilon,Z}(h) = \hat{\gamma}_{\varepsilon,Z}(h) / \hat{\sigma}^2_{\varepsilon,Z}$$

Bias Adjustments to Zeger's estimates

Letting β_0 denote the true parameter value, write

$$\hat{\mu}_t = \mu_t \exp(\mathrm{x}_t^{\mathrm{T}}(\hat{\beta} - \beta_0))$$

Using the theorem, $\hat{\beta} - \beta_0$ is approximately distributed as $\mathrm{N}(0,\mathrm{G}_n)$, where $\mathrm{G}_n = \Omega_{\mathrm{I},n}^{-1} + \Omega_{\mathrm{I},n}^{-1}\,\Omega_{\mathrm{II},n}\,\Omega_{\mathrm{I},n}^{-1}$, $\hat{\mu}_t$ has an approximate log-normal distribution with mean and second moment,

$$E(\hat{\mu}_t) = \mu_t E\left(\exp(\mathrm{x}_t^{\mathrm{T}}(\hat{\beta} - \beta_0))\right) = \mu_t \exp(\mathrm{x}_t^{\mathrm{T}} G_n \mathrm{x}_t / 2)$$

$$E(\hat{\mu}_t^2) = \mu_t^2 E\left(\exp(2\mathrm{x}_t^{\mathrm{T}}(\hat{\beta} - \beta_0))\right) = \mu_t^2 \exp(2\mathrm{x}_t^{\mathrm{T}} G_n \mathrm{x}_t)$$

Thus both first and second moments have **positive** bias. A nearly unbiased estimate of μ_t is then

$$\hat{\mu}_t \exp(-\mathrm{x}_t^{\mathrm{T}} G_n \mathrm{x}_t / 2)$$

Bias Adjustments to Zeger's estimates (cont)

Using these results, a biased adjustment of the variance of the latent process is

$$\hat{\sigma}^2_{\varepsilon,UB} = \frac{\sum_{t=1}^{n}\left[(Y_t - \hat{\mu}_t)^2 + \hat{\mu}_t^2 e^{-2\mathrm{x}_t^{\mathrm{T}}\hat{G}_n\mathrm{x}_t}\left(e^{2\mathrm{x}_t^{\mathrm{T}}\hat{G}_n\mathrm{x}_t} - 2e^{\mathrm{x}_t^{\mathrm{T}}\hat{G}_n\mathrm{x}_t/2} + 1\right) - \hat{\mu}_t\right]}{\sum_{t=1}^{n}\hat{\mu}_t^2 e^{-2\mathrm{x}_t^{\mathrm{T}}\hat{G}_n\mathrm{x}_t}},$$

where the limiting covariance matrix is estimated by

$$\hat{G}_n = \hat{\Omega}_{I,n}^{-1} + \hat{\Omega}_{I,n}^{-1}\hat{\Omega}_{II,n}\hat{\Omega}_{I,n}^{-1},$$

$$\hat{\Omega}_{I,n} = \sum_{t=1}^{n}\mathrm{x}_t\mathrm{x}_t^{\mathrm{T}}\hat{\mu}_t,$$

$$\hat{\Omega}_{II,n} = \sum_{h=-L}^{L}\ \sum_{t=\max(1-h,1)}^{\min(n-h,n)}\mathrm{x}_t\mathrm{x}_{t+h}^{\mathrm{T}}\hat{\mu}_t\hat{\mu}_{t+h}\hat{\gamma}_{\varepsilon,Z}(h)$$

Simulation Results (linear regression function)

Autocovariance estimates of a log-normal AR(1) latent process with $\phi = .9$, variance .6931 and reg function $1+t/n$ (n=100).

	Means			SD	
Lag	True	Zeg	Z.UB	Zeg	Z.UB
0	1.00	.50	.70	.30	.63
1	.87	.40	.58	.27	.56
2	.75	.31	.48	.24	.51
3	.66	.24	.39	.21	.46
4	.58	.19	.32	.19	.41
5	.51	.14	.26	.17	.36
6	.45	.10	.21	.15	.33

Simulation Results (cont)

Autocorrelation estimates of a log-normal AR(1) latent process with $\phi = .9$, variance .6931 and regression function 1+t/n (n=100).

	Means			SD	
Lag	True	Zeg	Z.UB	Zeg	Z.UB
1	.87	.79	.81	.17	.16
2	.75	.60	.64	.20	.19
3	.66	.45	.50	.23	.22
4	.58	.38	.40	.24	.23
5	.51	.33	.30	.25	.24
6	.45	.21	.23	.25	.25

Simulation Results (cosine regression function)

Autocovariance estimates of a log-normal AR(1) latent process with $\phi = .9$, variance .6931 and reg function $1+\cos(2\pi t/12)$ (n=100).

	Means			SD	
Lag	True	Zeg	Z.UB	Zeg	Z.UB
0	1.00	.73	1.06	.44	.87
1	.87	.61	.90	.39	.79
2	.75	.52	.78	.36	.72
3	.66	.45	.68	.33	.66
4	.58	.38	.59	.30	.61
5	.51	.33	.52	.28	.56
6	.45	.21	.46	.26	.53

Simulation Results (cont)

Autocorrelation estimates of a log-normal AR(1) latent process with $\phi = .9$, variance .6931 and reg function $1+\cos(2\pi t/12)$ (n=100).

	Means			SD	
Lag	True	Zeg	Z.UB	Zeg	Z.UB
1	.87	.82	.84	.15	.14
2	.75	.69	.71	.17	.16
3	.66	.58	.60	.19	.18
4	.58	.49	.51	.20	.19
5	.51	.42	.44	.21	.21
6	.45	.35	.38	.22	.22

Optimaly Weighted Estimates

Consider weighted estimates of the variance of the latent process of the form

$$\hat{\sigma}^2_{\varepsilon,W} = \sum_{t=1}^{n} W_t^2 E_t \Big/ \sum_{t=1}^{n} W_t^2, \quad E_t = \hat{\mu}_t^{-1}\left[(Y_t - \hat{\mu}_t)^2 / \hat{\mu}_t - 1\right]$$

This estimate is approximately unbiased for any latent process. Choose weights to *minimize* variance of the estimate when latent process is IID.

Optimal weights: $W_t^{*2} = 1/\mathrm{Var}(E_t)$ given by complicated formula!

Zeger estimates: $W_{t,Z}^2 = \mu_t^2$

Variance Formulas for Optimal Estimates

Suppose

$$\mu_t = g(t/n) = \exp(\mathbf{x}_t^T \boldsymbol{\beta}).$$

Then under an IID latent process assumption

$$n\mathrm{Var}(\hat{\gamma}_{\varepsilon,W^Z}(h)) \approx I_Z := \int_0^1 g^2(x)(\sigma_\varepsilon^2 g(x)+1)^2 dx \Big/ \left(\int_0^1 g^2(x)dx\right)^2$$

$$n\mathrm{Var}(\hat{\gamma}_{\varepsilon,W^*}(h)) \approx I_{Opt} := 1 \Big/ \int_0^1 g^2(x)(\sigma_\varepsilon^2 g(x)+1)^{-2} dx$$

Clearly, $I_Z \geq I_{opt}$ and for the polio data regression function $\mathbf{f}^{\mathbf{T}}(t)\boldsymbol{\beta}$,

Scenario	sqrt(I_Z)	sqrt(I_{Opt})
1. $\mu(x) = e^{\mathbf{f}(x)\beta}, \sigma_\varepsilon^2 = .77$	.131	.120
2. $\mu(x) = e^{\mathbf{f}(x)\beta}, \sigma_\varepsilon^2 = 1.54$	.212	.184
3. $\mu(x) = e^{2\mathbf{f}(x)\beta}, \sigma_\varepsilon^2 = .77$	.149	.105
4. $\mu(x) = e^{2\mathbf{f}(x)\beta}, \sigma_\varepsilon^2 = 1.54$	.275	.173

Tests for Zero Autocorrelation in Latent Process

Use Box-Pierce or Ljung-Box portmanteau tests applied to correlation estimates of residuals.

Pearson residuals: $e_t = (Y_t - \hat{\mu}_t)/\sqrt{\hat{\mu}_t}$ nearly IID if latent process is IID.

ACF of Pearson residuals: $\hat{\rho}_P(h) = \sum_{t=1}^{n-h} e_t e_{t+h} / \sum_{t=1}^{n} e_t^2$

Ljung-Box statistic: $H_P = \sum_{h=1}^{L} \hat{\rho}_P^2(h) / \mathrm{Var}(\hat{\rho}_P(h))$ has a chi-square distribution with L degrees of freedom under H_0: no spatial correlation .

Tests for Zero Autocorrelation in Latent Process (cont)

Lack of power of H_P for some alternatives: To see this note,

$$E\hat{\rho}_P(h) \approx \frac{\int_0^1 e^{f(x)\beta} dx}{\sigma_\varepsilon^{-2} + \int_0^1 e^{f(x)\beta} dx} \rho_\varepsilon(h) \to 0, \quad \text{as} \quad \sigma_\varepsilon^2 \to 0$$

while $\mathrm{Var}(\hat{\rho}_P(h)) \approx n^{-1}$, for σ_ε^2 small. This problem arises in the analysis of the asthma data (see later).

Alternative LB estimate: $H_{Z,UB} = \sum_{h=1}^{L} \hat{\rho}_{Z,UB}^2(h) / \mathrm{Var}(\hat{\rho}_{Z,UB}(h))$

Relative performance of test statistics depend on regression fcn.

A Simulation Illustration

Model: $Y_t \mid \varepsilon_t, \mathbf{x}_t \sim P(\varepsilon_t \exp(\mathbf{x}_t^T \boldsymbol{\beta}))$, where

- $\mathbf{x}_t^T \boldsymbol{\beta}$ is the estimated regression function from polio data
- $\ln \varepsilon_t = \alpha_t$, where $(\alpha_t + \sigma^2/2) = \phi(\alpha_{t-1} + \sigma^2/2) + \eta_t$, $\{\eta_t\} \sim \text{IID } N(0, \sigma^2(1-\phi^2))$, with $\phi = .82$, $\sigma^2 = .57$.
- Sample size is n=168
- 1000 reps.

Results: H_0 was rejected 97.7% using test based on S_a ($\alpha = .05$). 88% of these cases rejected 0 correlation in latent process using $H_{Z,UB}$ (78% using H_P)

	True	Mean	SD	Min	Max	%<1
$\hat{\rho}_{\varepsilon,UB}(1)$	.78	.79	.24	.05	2.19	84%
$\hat{\rho}_{\alpha,UB}(1)$	.82	.82	.21	.06	2.0	84%

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 $\cos(2\pi kt/365)$, $\sin(2\pi kt/365)$, $k=1,2,3,4,5,8$.
- 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.

Application to Sydney Asthma Count Data (cont)

Humidity variable: $$H_t = \frac{1}{7}\sum_{i=0}^{6} h_{t\text{-}12\text{-}i}$$

where h_t is the residual from an annual cycle harmonic model fit to the daily average of humidity at 0900 and 1500 hours.

GLM analysis:

Effect	$\hat{\beta}$	GLM s.e.	Theorem s.e.
Sunday	.230	.051	.055
Monday	.236	.051	.055
H_t	.210	.048	.066

t-ratios for humidity are 4.41 and 3.19

Application to Sydney Asthma Count Data (cont)

Test for presence of latent process: S_a was 3.30 (highly significant)

Tests of correlation in latent process:

	Degrees of freedom		
Test statistic	5	10	15
$H_{Z,UB}$	44.63(2e-08)	74.86(5e-12)	81.32(4e-11)
H_P	10.78(.056)	25.60(.004)	26.83(.030)

Application to Sydney Asthma Count Data (cont)

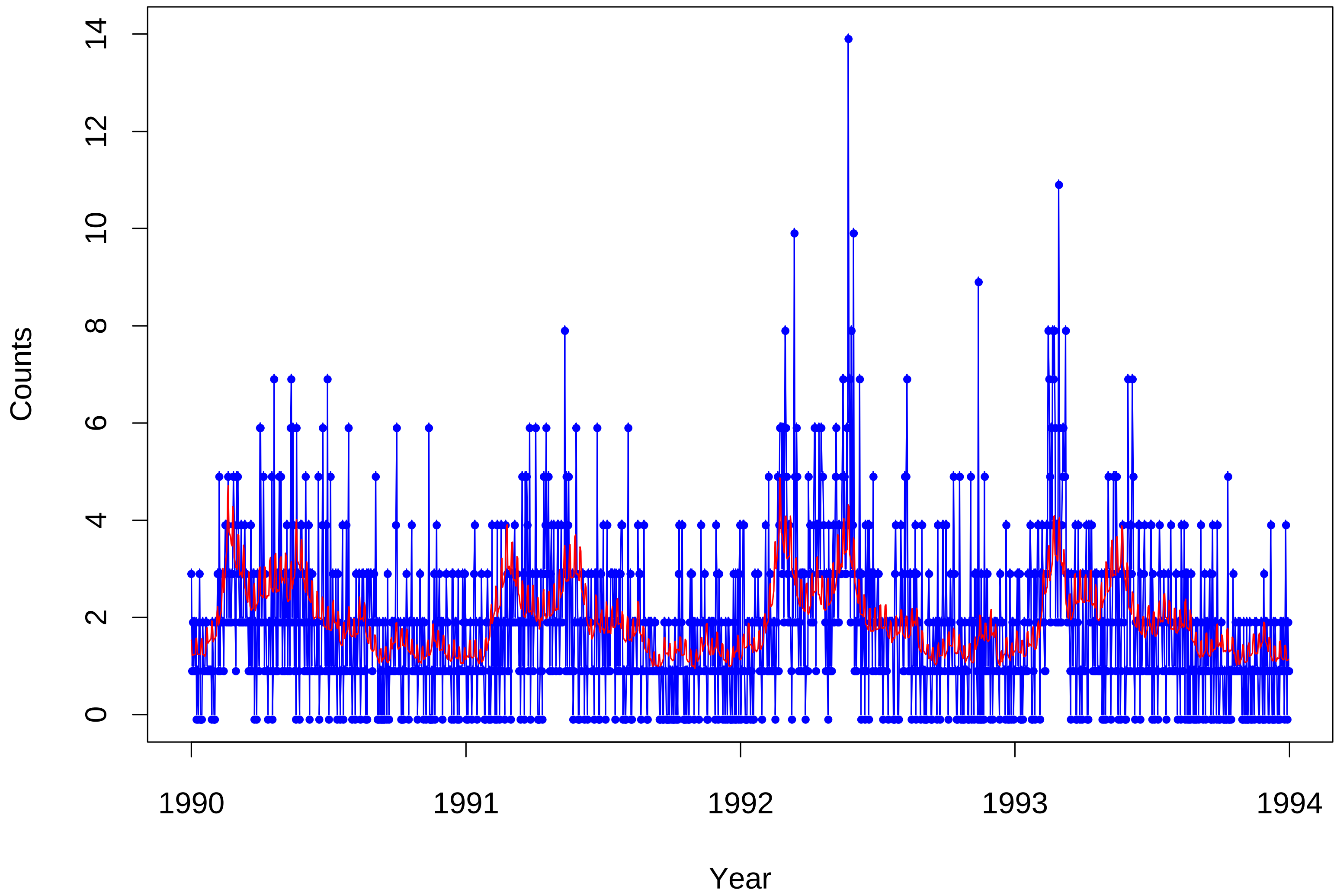
ACVF and ACV estimates.

lag h	$\hat{\gamma}_Z$	$\hat{\gamma}_{Z,UB}$	s.e.	$\hat{\rho}_{Z,UB}$	$\hat{\rho}_P$
0	.054	.067		1.00	1.000
1	.041	.053	.209	.79	.047
2	.030	.041	.224	.62	.021
3	.038	.050	.224	.74	.055
4	.023	.033	.224	.50	.033
5	.025	.036	.224	.54	.026
6	.020	.030	.224	.45	.025

Note: $(1461)^{-.5} = .026$ implies ACF for Pearson residuals are barely significant at lags 1 and 3? The small values of ACF can be partially explained by

$$E(\hat{\rho}_P(1)) \approx \frac{1.934}{.054^{-1} + 1.934}(.76) = .0718$$

Asthma Counts With Estimated Trend Function



Observation Driven Models

Count Data: $Y_1, \ldots, Y_n$

Let $H_t = (\mathbf{Y}^{(t-1)}, \mathbf{X}^{(t)})$ be information contained in the past of the observed count process and the past and present of the regressor variables.

Zeger & Qaqish (1988) models: Assume $Y_t \mid H_t$ is Poisson with mean μ_t where

Model 1: $$\mu_t = \exp(\mathbf{x}_t^{\mathbf{T}}\beta)\prod_{i=1}^{p}\left[\frac{\max(Y_{t-i}, c)}{\exp(\mathbf{x}_{t-i}^{\mathbf{T}}\beta)}\right]^{\gamma_i}, \; c > 0,$$

Model 2: $$\mu_t = \exp(\mathbf{x}_t^{\mathbf{T}}\beta)\prod_{i=1}^{p}\left[\frac{Y_{t-i} + c}{\exp(\mathbf{x}_{t-i}^{\mathbf{T}}\beta) + c}\right]^{\gamma_i}, \; c > 0,$$

Model 3: $$\mu_t = \exp(\mathbf{x}_t^{\mathbf{T}}\beta + \sum_{i=1}^{p}\gamma_i Y_{t-i}).$$

Observation Driven Models (cont)

Remarks:

- Z&Q argue that model 1 is preferred on their three desiderata.
- Model 3 cannot be stationary (if $p=1$ and $\gamma_1<0$).
- In Model 2 in the case $p=1$, c is interpreted as an immigration rate adding to counts at every time point.
- Estimation of c in both Models 1 & 2 is problematic.

New Observation Driven Model

For $\lambda > 0$, define

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

and assume that

$$\log \mu_t = W_t = \mathrm{x}_\mathrm{t}^{\mathrm{T}}\beta + \sum_{i=1}^{p} \theta_i e_{t-i}.$$

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

Assuming that $\lambda = .5$, we have

$$Var(W_t) = Var(\sum_{i=1}^{p} \theta_i e_{t-i}) = \sum_{i=1}^{p} \theta_i^2,$$

so that

$$\begin{aligned} E(\mu_t) &= E(e^{W_t}) \\ &\approx e^{x_t^T \beta + Var(W_t)/2} \\ &= e^{x_t^T \beta + \sum_{i=1}^{p} \theta_i^2 / 2}, \end{aligned}$$

which holds approximately if W_t is nearly Gaussian.

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

Properties of the New Model (cont)

The model 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

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

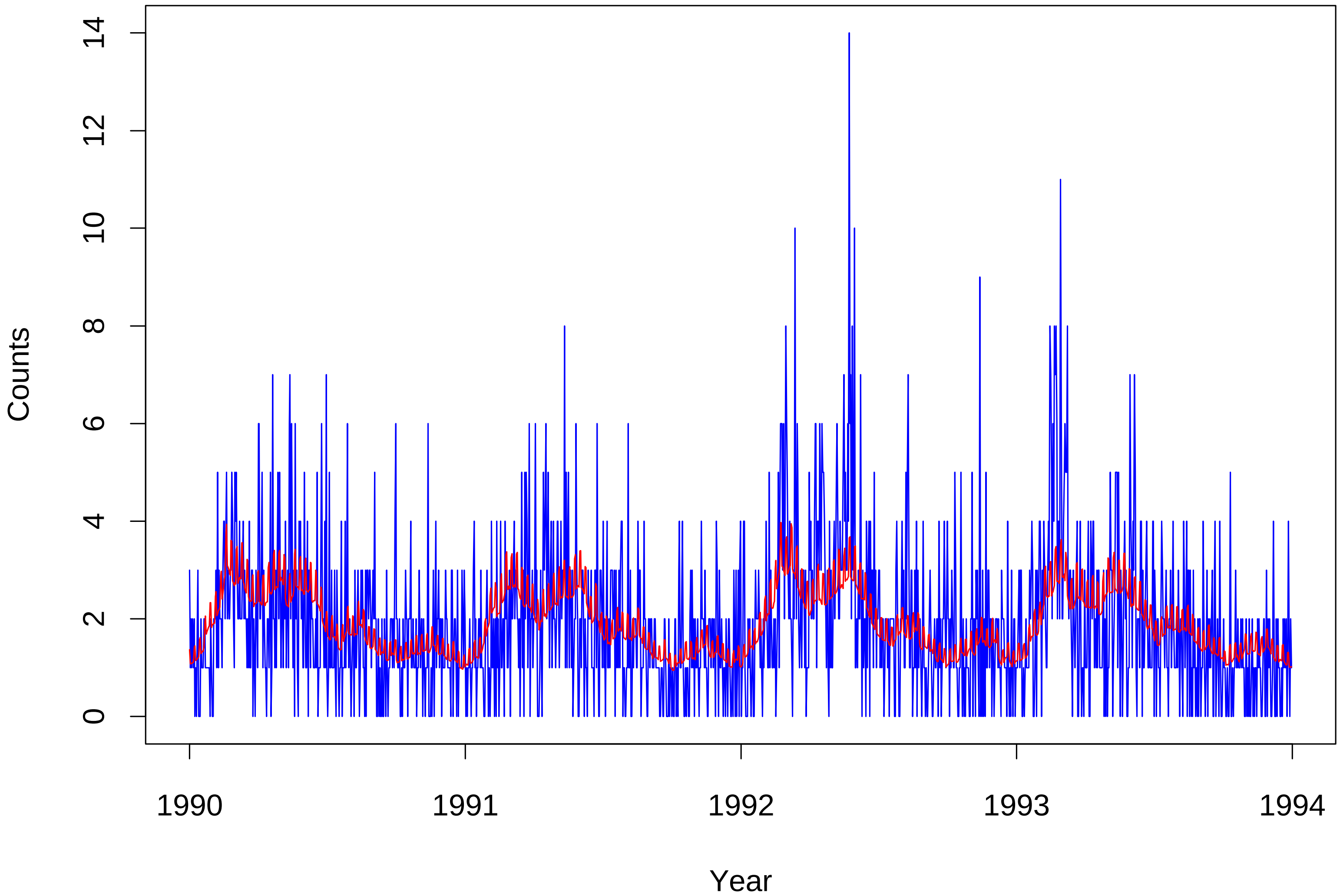
3. Is easy to predict with.

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

5. Generalizable to ARMA type lag structure.

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

Asthma Data w/ Deterministic Part of Mean Fcn



Asthma Data: Deterministic Part + AR in Pearson Resid