Regression Models for Time Series Analysis

Benjamin Kedem[1]
and
Konstantinos Fokianos[2]
[1]University of Maryland, College Park, MD
[2]University of Cyprus, Nicosia, Cyprus

Wiley, New York, 2002

Cox (1975). Partial likelihood, *Biometrika*, 62, 69–76.

Fahrmeir and Tutz (2001). *Multivariate Statistical Modelling Based on GLM*. 2nd ed., Springer, NY.

Fokianos (1996). Categorical Time Series: Prediction and Control. Ph.D. Thesis, University of Maryland.

Fokianos and Kedem (1998), Prediction and classification of non-stationary categorical time series. *J. Multivariate Analysis*,67, 277-296.

Kedem (1980). *Binary Time Series*, Marcel Dekker, NY

(•) Kedem and Fokianos (2002). *Regression Models for Time Series Analysis*, Wiley, NY.

McCullagh and Nelder (1989). *Generalized Linear Models*. 2nd edi., Chapman & Hall, London.

Nelder and Wedderburn (1972). Generalized linear models. *JRSS*, A, 135, 370–384.

Slud (1982). Consistency and efficiency of inference with the partial likelihood. *Biometrika*, 69, 547–552.

Slud and Kedem (1994). Partial likelihood analysis of logistic regression and autoregression. *Statistica Sinica*, 4, 89–106.

Wong (1986). Theory of partial likelihood. *Annals of Statistics*, 14, 88–123.

Part I: GLM and Time Series

Overview

Extension of Nelder and Wedderburn (1972), McCullagh and Nelder(1989) GLM to time series is possible due to:

- Increasing sequence of histories relative to an observer.

- Partial likelihood.

- The partial score is a martingale.

- Well behaved covariates.

Partial Likelihood

Suppose we observe a pair of jointly distributed time series, (X_t, Y_t), $t = 1, \ldots, N$, where $\{Y_t\}$ is a *response* series and $\{X_t\}$ is a *time dependent random covariate*. Employing the rules of conditional probability, the joint density of all the X, Y observations can be expressed as,

$$f_{\boldsymbol{\theta}}(x_1, y_1, \ldots, x_N, y_N) = f_{\boldsymbol{\theta}}(x_1) \left[\prod_{t=2}^{N} f_{\boldsymbol{\theta}}(x_t \mid d_t) \right] \left[\prod_{t=1}^{N} f_{\boldsymbol{\theta}}(y_t \mid c_t) \right] \quad (1)$$

$d_t = (y_1, x_1, \ldots, y_{t-1}, x_{t-1})$
$c_t = (y_1, x_1, \ldots, y_{t-1}, x_{t-1}, x_t)$.

The second product on the right hand side of (1) constitutes a partial likelihood according to Cox(1975).

An increasing sequence of σ-fields
$\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \ldots$.

$Y_1, Y_2, \ldots$ a sequence of random variables on some common probability space such that Y_t is $\mathcal{F}_t$ measurable.

$Y_t \mid \mathcal{F}_{t-1} \sim f_t(y_t; \boldsymbol{\theta})$.

$\boldsymbol{\theta} \in R^p$ is a fixed parameter.

The partial likelihood (PL) function relative to $\boldsymbol{\theta}$, $\mathcal{F}_t$, and the data $Y_1, Y_2, \ldots, Y_N$, is given by the product

$$\text{PL}(\boldsymbol{\theta}; y_1, \ldots, y_N) = \prod_{t=1}^{N} f_t(y_t; \boldsymbol{\theta}) \tag{2}$$

The General Regression Problem

$\{Y_t\}$ is a *response* time series with the corresponding p–dimensional *covariate process*,

$$\mathbf{Z}_{t-1} = (Z_{(t-1)1}, \ldots, Z_{(t-1)p})'.$$

Define,

$$\mathcal{F}_{t-1} = \sigma\{Y_{t-1}, Y_{t-2}, \ldots, \mathbf{Z}_{t-1}, \mathbf{Z}_{t-2}, \ldots\}.$$

Note: $\mathbf{Z}_{t-1}$ may already include past Y_t's.

The conditional expectation of the response given the past:

$$\mu_t = \mathsf{E}[Y_t \mid \mathcal{F}_{t-1}],$$

($\bullet$) The problem is to relate μ_t to the covariates.

Time Series Following GLM

1. **Random Component.** The conditional distribution of the response given the past belongs to the exponential family of distributions in *natural* or *canonical* form,

$$f(y_t;\theta_t,\phi \mid \mathcal{F}_{t-1}) = \exp\left\{\frac{y_t\theta_t - b(\theta_t)}{\alpha_t(\phi)} + c(y_t;\phi)\right\}. \tag{3}$$

$\alpha_t(\phi) = \phi/\omega_t$, *dispersion* ϕ, *prior weight* ω_t.

2. **Systematic Component.** There is a monotone function $g(\cdot)$ such that,

$$g(\mu_t) = \eta_t = \sum_{j=1}^{p} \beta_j Z_{(t-1)j} = \mathbf{Z}'_{t-1}\boldsymbol{\beta}. \tag{4}$$

$g(\cdot)$: the *link function*

η_t: the linear predictor of the model.

Typical choices for $\eta_t = \mathbf{Z}'_{t-1}\boldsymbol{\beta}$, could be

$$\beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \beta_3 X_t \cos(\omega_0 t)$$

or

$$\beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \beta_3 Y_{t-1} X_t + \beta_4 X_{t-1}$$

or

$$\beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2}^7 + \beta_3 Y_{t-1} \log(X_{t-12})$$

GLM Equations:

$$\int f(y; \theta_t; \phi \mid \mathcal{F}_{t-1}) dy = 1.$$

This implies the relationships:

$$\mu_t = \mathsf{E}[Y_t \mid \mathcal{F}_{t-1}] = b'(\theta_t). \quad (5)$$

$$\mathsf{Var}[Y_t \mid \mathcal{F}_{t-1}] = \alpha_t(\phi) b''(\theta_t) \equiv \alpha_t(\phi) V(\mu_t). \quad (6)$$

Since $\mathsf{Var}[Y_t \mid \mathcal{F}_{t-1}] > 0$, it follows that b' is monotone. Therefore, equation (5) implies that

$$\theta_t = (b')^{-1}(\mu_t). \quad (7)$$

We see that θ_t itself is a monotone function of μ_t and hence it can be used to define a link function. The link function

$$g(\mu_t) = \theta_t(\mu_t) = \eta_t = \mathbf{Z}_{t-1}^{'} \boldsymbol{\beta} \quad (8)$$

is called the *canonical link* function.

Example: *Poisson Time Series.*

$f(y_t; \theta_t, \phi \mid \mathcal{F}_{t-1}) = \exp\{(y_t \log \mu_t - \mu_t) - \log y_t!\}.$

$\mathrm{E}[Y_t|\mathcal{F}_{t-1}] = \mu_t$, $b(\theta_t) = \mu_t = \exp(\theta_t)$, $V(\mu_t) = \mu_t$, $\phi = 1$, and $\omega_t = 1$. The canonical link is

$$g(\mu_t) = \theta_t(\mu_t) = \log \mu_t = \eta_t = \mathbf{Z}'_{t-1}\boldsymbol{\beta}.$$

As an example , if $\mathbf{Z}_{t-1} = (1, X_t, Y_{t-1})'$, then

$$\log \mu_t = \beta_0 + \beta_1 X_t + \beta_2 Y_{t-1}$$

with $\{X_t\}$ standing for some covariate process, or a possible trend, or a possible seasonal component.

Example: *Binary Time Series*

$\{Y_t\}$ takes the values 0,1. Let

$$\pi_t = P(Y_t = 1 \mid \mathcal{F}_{t-1}).$$

Then

$$f(y_t; \theta_t, \phi \mid \mathcal{F}_{t-1}) = \exp\left\{ y_t \log\left(\frac{\pi_t}{1-\pi_t}\right) + \log(1-\pi_t) \right\}$$

The canonical link gives the *logistic regression* model

$$g(\pi_t) = \theta_t(\pi_t) = \log\frac{\pi_t}{1-\pi_t} = \eta_t = \mathbf{Z}'_{t-1}\boldsymbol{\beta}. \quad (9)$$

Note:

$$\pi_t = F_l(\eta_t) \quad (10)$$

Partial Likelihood Inference

Given a time series $\{Y_t\}$, $t = 1, \ldots, N$, conditionally distributed as (3).

The partial likelihood of the observed series is

$$\text{PL}(\boldsymbol{\beta}) = \prod_{t=1}^{N} f(y_t; \theta_t, \phi \mid \mathcal{F}_{t-1}). \tag{11}$$

Then from (3), the log–partial likelihood, $l(\boldsymbol{\beta})$, is given by

$$\begin{aligned} l(\boldsymbol{\beta}) &= \sum_{t=1}^{N} \log f(y_t; \theta_t, \phi \mid \mathcal{F}_{t-1}) \equiv \sum_{t=1}^{N} l_t \\ &= \sum_{t=1}^{N} \left\{ \frac{y_t\theta_t - b(\theta_t)}{\alpha_t(\phi)} + c(y_t, \phi) \right\} \\ &= \sum_{t=1}^{N} \left\{ \frac{y_t u(\mathbf{z}'_{t-1}\boldsymbol{\beta}) - b(u(\mathbf{z}'_{t-1}\boldsymbol{\beta}))}{\alpha_t(\phi)} + c(y_t, \phi) \right\} \end{aligned} \tag{12}$$

$$\nabla \equiv \left(\frac{\partial}{\partial \beta_1}, \frac{\partial}{\partial \beta_2}, \cdots, \frac{\partial}{\partial \beta_p}\right)'.$$

The *partial score* is a p–dimensional vector,

$$\mathbf{S}_N(\boldsymbol{\beta}) \equiv \nabla l(\boldsymbol{\beta}) = \sum_{t=1}^{N} \mathbf{Z}_{t-1} \frac{\partial \mu_t}{\partial \eta_t} \frac{(Y_t - \mu_t(\boldsymbol{\beta}))}{\sigma_t^2(\boldsymbol{\beta})} \quad (13)$$

with $\sigma_t^2(\boldsymbol{\beta}) = \text{Var}[Y_t \mid \mathcal{F}_{t-1}]$.

The *partial score vector process* $\{\mathbf{S}_t(\boldsymbol{\beta})\}$, $t = 1, \ldots, N$, is defined from the partial sums

$$\mathbf{S}_t(\boldsymbol{\beta}) = \sum_{s=1}^{t} \mathbf{Z}_{s-1} \frac{\partial \mu_s}{\partial \eta_s} \frac{(Y_s - \mu_s(\boldsymbol{\beta}))}{\sigma_s^2(\boldsymbol{\beta})}. \quad (14)$$

The solution of the score equation,

$$\mathbf{S}_N(\boldsymbol{\beta}) = \nabla \log \mathsf{PL}(\boldsymbol{\beta}) = \mathbf{0} \qquad (15)$$

is denoted by $\widehat{\boldsymbol{\beta}}$, and is referred to as the maximum partial likelihood estimator (MPLE) of $\boldsymbol{\beta}$. The system of equations (15) is non–linear and is customarily solved by the Fisher scoring method, an iterative algorithm resembling the Newton–Raphson procedure. Before turning to the Fisher scoring algorithm in our context of conditional inference, it is necessary to introduce several important matrices.

An important role in partial likelihood inference is played by the *cumulative conditional information matrix*, $\mathbf{G}_N(\boldsymbol{\beta})$, defined by a sum of conditional covariance matrices,

$$\begin{aligned}\mathbf{G}_N(\boldsymbol{\beta}) &= \sum_{t=1}^{N} \mathrm{Cov}\left[\mathbf{Z}_{t-1}\frac{\partial \mu_t}{\partial \eta_t}\frac{(Y_t - \mu_t(\boldsymbol{\beta}))}{\sigma_t^2(\boldsymbol{\beta})} \mid \mathcal{F}_{t-1}\right] \\ &= \sum_{t=1}^{N} \mathbf{Z}_{t-1}\left(\frac{\partial \mu_t}{\partial \eta_t}\right)^2 \frac{1}{\sigma_t^2(\boldsymbol{\beta})}\mathbf{Z}'_{t-1} \\ &= \mathbf{Z}'\mathbf{W}(\boldsymbol{\beta})\mathbf{Z}.\end{aligned}$$

We also need:

$$\mathbf{H}_N(\boldsymbol{\beta}) \equiv -\nabla\nabla\,' l(\boldsymbol{\beta}).$$

Define $\mathbf{R}_N(\boldsymbol{\beta})$ from the difference

$$\mathbf{H}_N(\boldsymbol{\beta}) = \mathbf{G}_N(\boldsymbol{\beta}) - \mathbf{R}_N(\boldsymbol{\beta}).$$

Fact: For canonical links $\mathbf{R}_N(\boldsymbol{\beta}) = \mathbf{0}$.

Fisher Scoring: In Newton-Raphson replace $\mathbf{H}_N(\boldsymbol{\beta})$ by its conditional expectation:

$$\widehat{\boldsymbol{\beta}}^{(k+1)} = \widehat{\boldsymbol{\beta}}^{(k)} + \mathbf{G}_N^{-1}(\widehat{\boldsymbol{\beta}}^{(k)})\mathbf{S}_N(\widehat{\boldsymbol{\beta}}^{(k)}).$$

Fisher scoring becomes Newton-Raphson for canonical links.

Fisher scoring simplifies to Iterative Reweighted Least Squares:

$$\widehat{\boldsymbol{\beta}}^{(k+1)} = \left(\mathbf{Z}'\mathbf{W}(\widehat{\boldsymbol{\beta}}^{(k)})\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{W}(\widehat{\boldsymbol{\beta}}^{(k)})\mathbf{q}^{(k)}.$$

Asymptotic Theory

Assumption A

A1. The true parameter $\boldsymbol{\beta}$ belongs to an open set $B \subseteq R^p$.

A2. The covariate vector $\mathbf{Z}_{t-1}$ almost surely lies in a nonrandom compact subset Γ of R^p, such that $P[\sum_{t=1}^{N} \mathbf{Z}_{t-1}\mathbf{Z}'_{t-1} > \mathbf{0}] = 1$. In addition, $\mathbf{Z}'_{t-1}\boldsymbol{\beta}$ lies almost surely in the domain H of the inverse link function $h = g^{-1}$ for all $\mathbf{Z}_{t-1} \in \Gamma$ and $\boldsymbol{\beta} \in B$.

A3. The inverse link function h–defined in (A2)–is twice continuously differentiable and $|\partial h(\gamma)/\partial\gamma| \neq 0$.

A4. There is a probability measure ν on R^p such that $\int_{R^p} \mathbf{z}\mathbf{z}'\nu(d\mathbf{z})$ is positive definite, and such that under (3) and (4) for Borel sets $A \subset R^p$,

$$\frac{1}{N}\sum_{t=1}^{N} I_{[\mathbf{Z}_{t-1} \in A]} \rightarrow \nu(A)$$

in probability as $N \to \infty$, at the true value of $\boldsymbol{\beta}$.

A4 calls for asymptotically "well behaved" covariates:

$$\frac{\sum_{t=1}^{N} f(\mathbf{Z}_{t-1})}{N} \rightarrow \int_{R^p} f(\mathbf{z})\nu(d\mathbf{z})$$

in probability as $N \to \infty$. Thus, there exists a $p \times p$ limiting *information matrix per observation*, $\mathbf{G}(\boldsymbol{\beta})$, such that

$$\frac{\mathbf{G}_N(\boldsymbol{\beta})}{N} \rightarrow \mathbf{G}(\boldsymbol{\beta}) \qquad (16)$$

in probability, as $N \to \infty$.

Slud and K (1994), Fokianos and K (1998):

1. $\{\mathbf{S}_t(\boldsymbol{\beta})\}$ relative to $\{\mathcal{F}_t\}$, $t = 1, \ldots, N$, is a martingale.

2. $\frac{\mathbf{R}_N(\boldsymbol{\beta})}{N} \rightarrow \mathbf{0}$.

3. $\frac{\mathbf{S}_N(\boldsymbol{\beta})}{\sqrt{N}} \rightarrow \mathcal{N}_p(\mathbf{0}, \mathbf{G}(\boldsymbol{\beta}))$.

4. $\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \rightarrow \mathcal{N}_p(\mathbf{0}, \mathbf{G}^{-1}(\boldsymbol{\beta}))$.

100(1 − α)% prediction interval ($h = g^{-1}$)

$$\mu_t(\boldsymbol{\beta}) \doteq \mu_t(\widehat{\boldsymbol{\beta}}) \pm z_{\alpha/2} \frac{|h'(\mathbf{Z}'_{t-1}\boldsymbol{\beta})|}{\sqrt{N}} \sqrt{\mathbf{Z}'_{t-1}\mathbf{G}^{-1}(\boldsymbol{\beta})\mathbf{Z}_{t-1}}$$

Hypothesis Testing

Let $\tilde{\boldsymbol{\beta}}$ be the MPLE of $\boldsymbol{\beta}$ obtained under H_0 with $r < p$ restrictions,

$\mathrm{H}_0 : \beta_1 = \cdots = \beta_{\mathrm{r}} = 0.$

Let $\widehat{\boldsymbol{\beta}}$ be the unrestricted MPLE. The log–partial likelihood ratio statistic

$$\lambda_N = 2\left\{ l(\widehat{\boldsymbol{\beta}}) - l(\tilde{\boldsymbol{\beta}}) \right\} \qquad (17)$$

converges to χ^2_r.

More generally:

Assume $\mathbf{C}$ is a known matrix with full rank r, $r < p$.

Under the general linear hypothesis,

$$\mathsf{H}_0: \ \mathbf{C}\boldsymbol{\beta} = \boldsymbol{\beta}_0 \text{ against } \mathsf{H}_1: \ \mathbf{C}\boldsymbol{\beta} \neq \boldsymbol{\beta}_0, \quad (18)$$

$$\{\mathbf{C}\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\}'\{\mathbf{C}\mathbf{G}^{-1}(\widehat{\boldsymbol{\beta}})\mathbf{C}'\}^{-1}\{\mathbf{C}\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\} \rightarrow \chi^2_r$$

Diagnostics

$l(\mathbf{y};\mathbf{y})$: maximum log partial likelihood corresponding to the saturated model.
$l(\hat{\boldsymbol{\mu}};\mathbf{y})$: maximum log partial likelihood from the reduced model.

- Scaled Deviance:

$$D \equiv 2\{l(\mathbf{y};\mathbf{y}) - l(\hat{\boldsymbol{\mu}};\mathbf{y})\} \sim \chi^2_{N-p}$$

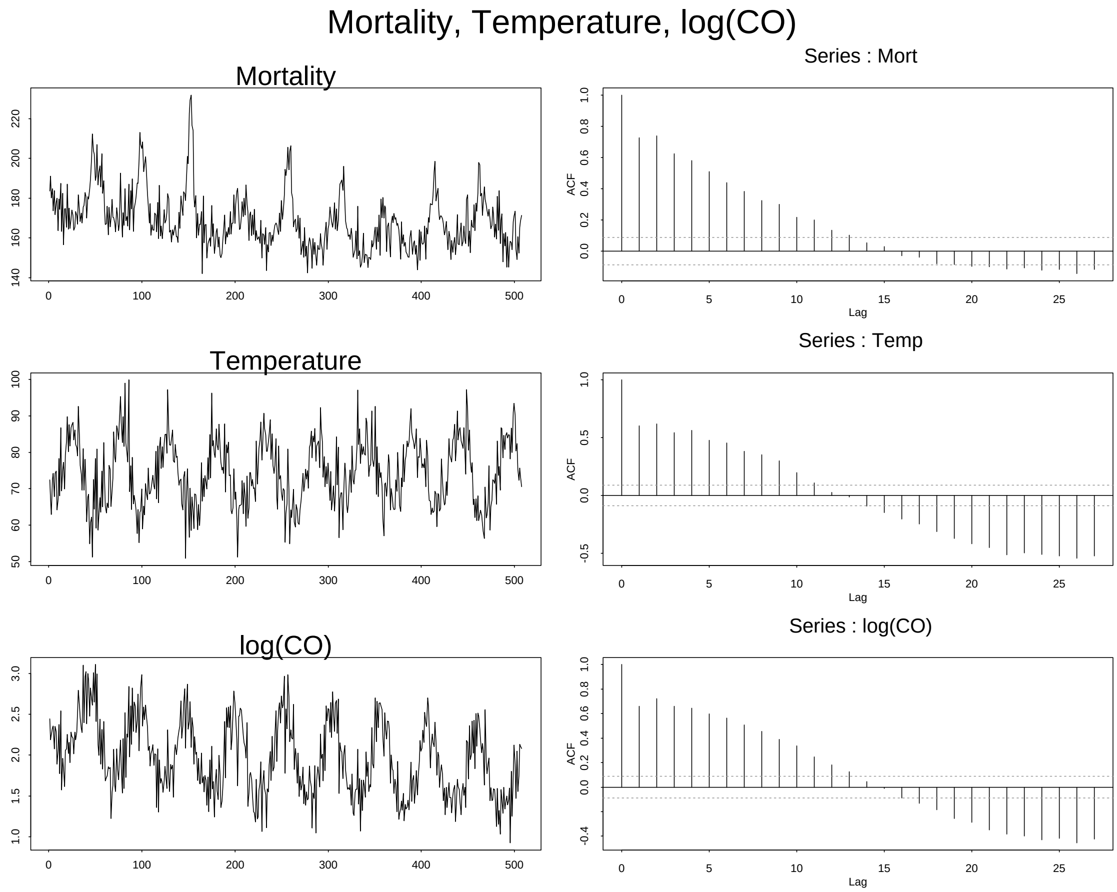
- $\text{AIC}(p) = -2\log \text{PL}(\hat{\boldsymbol{\beta}}) + 2p,$

- $\text{BIC}(p) = -2\log \text{PL}(\hat{\boldsymbol{\beta}}) + p\log N$

Analysis of Mortality Count Data in LA

Weekly data from Los Angeles County during a period of 10 years from January 1, 1970, to December 31, 1979: Weekly sampled filtered time series. $N = 508$.

Response	
Y	Total Mortality (filtered)
Weather	
T	Temperature
RH	Relative humidity
Pollution	
CO	Carbon monoxide
SO_2	Sulfur dioxide
NO_2	Nitrogen dioxide
HC	Hydrocarbons
OZ	Ozone
KM	Particulates



Weekly data of filtered total mortality and temperature, and log-filtered CO, and the corresponding estimated autocorrelation functions. $N = 508$.

Covariates and η_t used in Poisson regression. $S = \text{SO}_2$, $N = \text{NO}_2$. To recover η_t, insert the β's. For Model 2, $\eta_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2}$, etc.

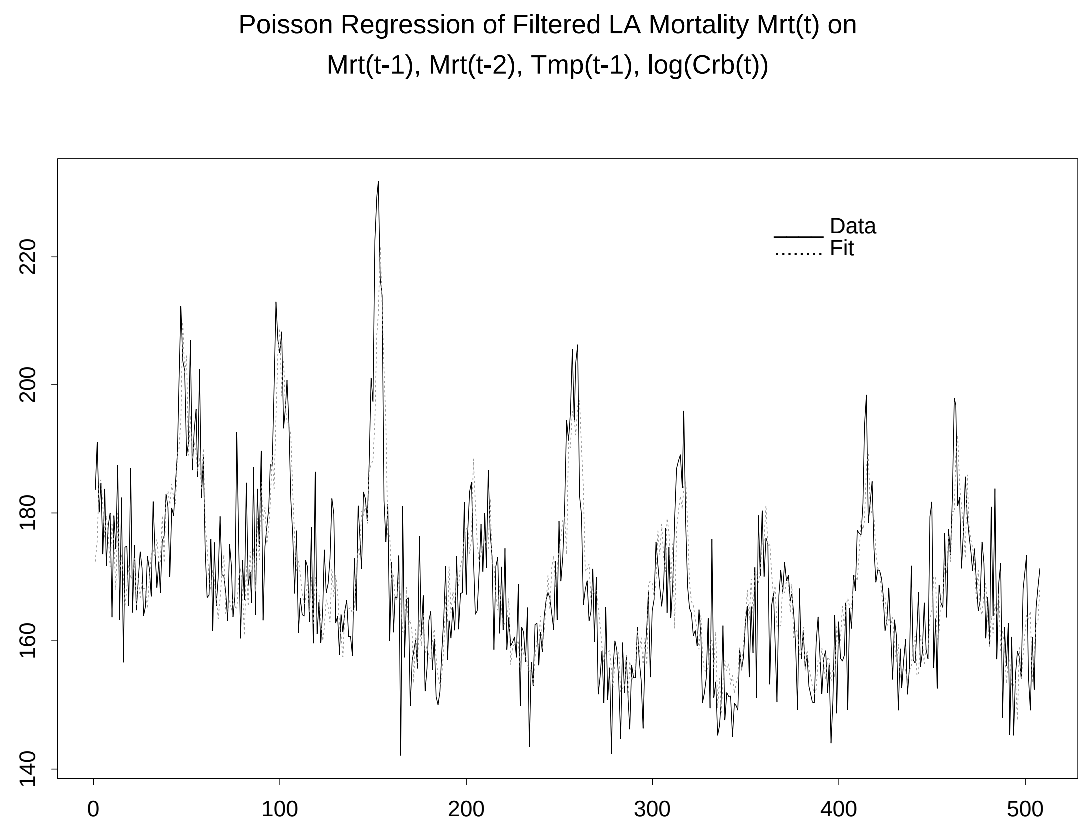
Model 0	$T_t + RH_t + CO_t + S_t + N_t$ $+ HC_t + OZ_t + KM_t$
Model 1	Y_{t-1}
Model 2	$Y_{t-1} + Y_{t-2}$
Model 3	$Y_{t-1} + Y_{t-2} + T_{t-1}$
Model 4	$Y_{t-1} + Y_{t-2} + T_{t-1} + \log(CO_t)$
Model 5	$Y_{t-1} + Y_{t-2} + T_{t-1} + T_{t-2} + \log(CO_t)$
Model 6	$Y_{t-1} + Y_{t-2} + T_t + T_{t-1} + \log(CO_t)$

Comparison of 7 Poisson regression models. $N = 508$.

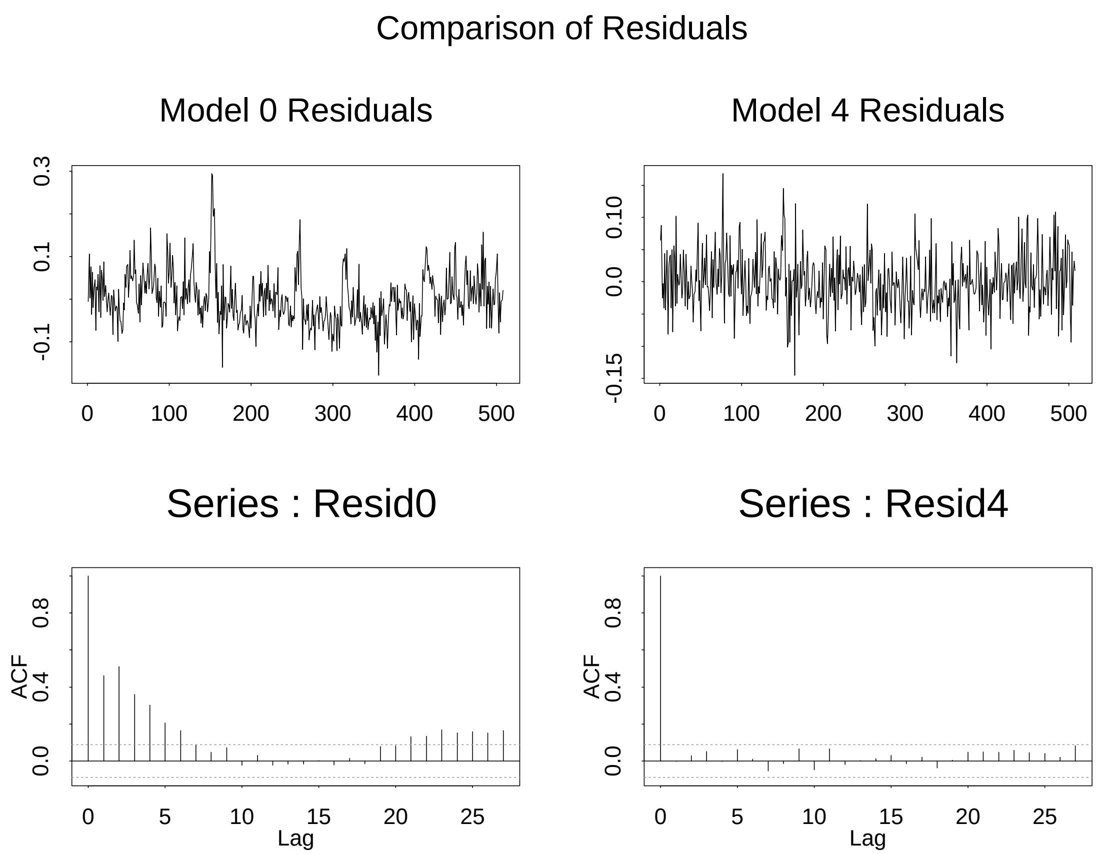
Model	p	D	df	AIC	BIC
0	9	315.69	499	333.69	371.76
1	2	276.07	506	280.07	288.53
2	3	222.23	505	228.23	240.92
3	4	203.52	504	211.52	228.44
4	5	174.55	503	184.55	205.71
5	6	174.53	502	186.53	211.91
6	6	171.41	502	183.41	208.79

Choose Model 4:

$$\log(\widehat{\mu}_t) = \widehat{\beta}_0 + \widehat{\beta}_1 Y_{t-1} + \widehat{\beta}_2 Y_{t-2} + \widehat{\beta}_3 T_{t-1} + \widehat{\beta}_4 \log(CO_t)$$



Observed and predicted weekly filtered total mortality from Model 4.



Working residuals from Models 0 and 4, and their respective estimated autocorrelation.

Part II: Binary Time Series

Example of bts: Two categories obtained by clipping.

$$Y_t \equiv I_{[X_t \in C]} = \begin{cases} 1, & \text{if} \quad X_t \in C \\ 0, & \text{if} \quad X_t \in \overline{C} \end{cases} \tag{19}$$

$$Y_t \equiv I_{[X_t \geq r]} = \begin{cases} 1, & \text{if} \quad X_t \geq r \\ 0, & \text{if} \quad X_t < r \end{cases} \tag{20}$$

Other examples: at time $t = 1, 2, ...,$

(Rain, No Rain), (S&P Up, S&P Down), etc.

$\{Y_t\}$ taking the values 0 or 1, $t = 1, 2, 3, \cdots$.

$\{\mathbf{Z}_{t-1}\}$ p-dim covariate stochastic data.

Against the backdrop of the general framework presented above we wish to relate

$$\mu_t(\boldsymbol{\beta}) = \pi_t(\boldsymbol{\beta}) = P_{\boldsymbol{\beta}}(Y_t = 1|\mathcal{F}_{t-1}) \qquad (21)$$

to the covariates. For this we need good links!

Standard logistic distribution,

$$F_l(x) = \frac{e^x}{1+e^x} = \frac{1}{1+e^{-x}}, \quad -\infty < x < \infty$$

Then,

$$F_l^{-1}(x) = \log(x/(1-x))$$

is the natural link under some conditions.

- Fact: For any bts, there are θ_j such that

$$\log\left\{\frac{P(Y_t = 1|Y_{t-1} = y_{t-1}, ..., Y_1 = y_1)}{P(Y_t = 0|Y_{t-1} = y_{t-1}, ..., Y_1 = y_1)}\right\} = \theta_0 + \theta_1 y_{t-1} + \cdots + \theta_p y_{t-p}$$

or

$$\pi_t(\boldsymbol{\beta}) = \frac{1}{1 + \exp[-(\theta_0 + \theta_1 y_{t-1} + \cdots + \theta_p y_{t-p})]}$$

- Fact: Consider an AR(p) time series

$$X_t = \gamma_0 + \gamma_1 X_{t-1} + \ldots + \gamma_p X_{t-p} + \lambda\epsilon_t$$

where ϵ_t are i.i.d. logistically distributed. Define: $Y_t = I_{[X_t \geq r]}$. Then

$$\pi_t(\boldsymbol{\beta}) = \frac{1}{1 + \exp[-(\gamma_0 - r + \gamma_1 X_{t-1} + \cdots + \gamma_p X_{t-p})/\lambda]}$$

This motivates *logistic regression:*

$$\pi_t(\boldsymbol{\beta}) \equiv P_{\boldsymbol{\beta}}(Y_t = 1|\mathcal{F}_{t-1}) = F_l(\boldsymbol{\beta}'\mathbf{Z}_{t-1})$$
$$= \frac{1}{1 + \exp[-\boldsymbol{\beta}'\mathbf{Z}_{t-1}]}$$

or equivalently, the link function is

$$\text{logit}(\pi_t(\boldsymbol{\beta})) \equiv \log\left\{\frac{\pi_t(\boldsymbol{\beta})}{1 - \pi_t(\boldsymbol{\beta})}\right\} = \boldsymbol{\beta}'\mathbf{Z}_{t-1}$$

This is the *canonical link*.

Link functions for binary time series.

logit	$\boldsymbol{\beta}'\mathbf{Z}_{t-1} = \log\{\pi_t(\boldsymbol{\beta})/(1-\pi_t(\boldsymbol{\beta}))\}$
probit	$\boldsymbol{\beta}'\mathbf{Z}_{t-1} = \Phi^{-1}\{\pi_t(\boldsymbol{\beta})\}$
log-log	$\boldsymbol{\beta}'\mathbf{Z}_{t-1} = -\log\{-\log(\pi_t(\boldsymbol{\beta}))\}$
C-log-log	$\boldsymbol{\beta}'\mathbf{Z}_{t-1} = \log\{-\log(1-\pi_t(\boldsymbol{\beta}))\}$

Note: Here all the inverse links are cdf's. In what follows we always assume the inverse link is a differentiable cdf $F(x)$.

The partial likelihood of $\boldsymbol{\beta}$ takes on the simple product form,

$$\begin{aligned} \mathrm{PL}(\boldsymbol{\beta}) &= \prod_{t=1}^{N} [\pi_t(\boldsymbol{\beta})]^{y_t} [1 - \pi_t(\boldsymbol{\beta})]^{1-y_t} \\ &= \prod_{t=1}^{N} [F(\boldsymbol{\beta}' \mathbf{Z}_{t-1})]^{y_t} [1 - F(\boldsymbol{\beta}' \mathbf{Z}_{t-1})]^{1-y_t} \end{aligned}$$

We have under Assumption A:

$$\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \rightarrow \mathcal{N}_p(\mathbf{0}, \mathbf{G}^{-1}(\boldsymbol{\beta})).$$

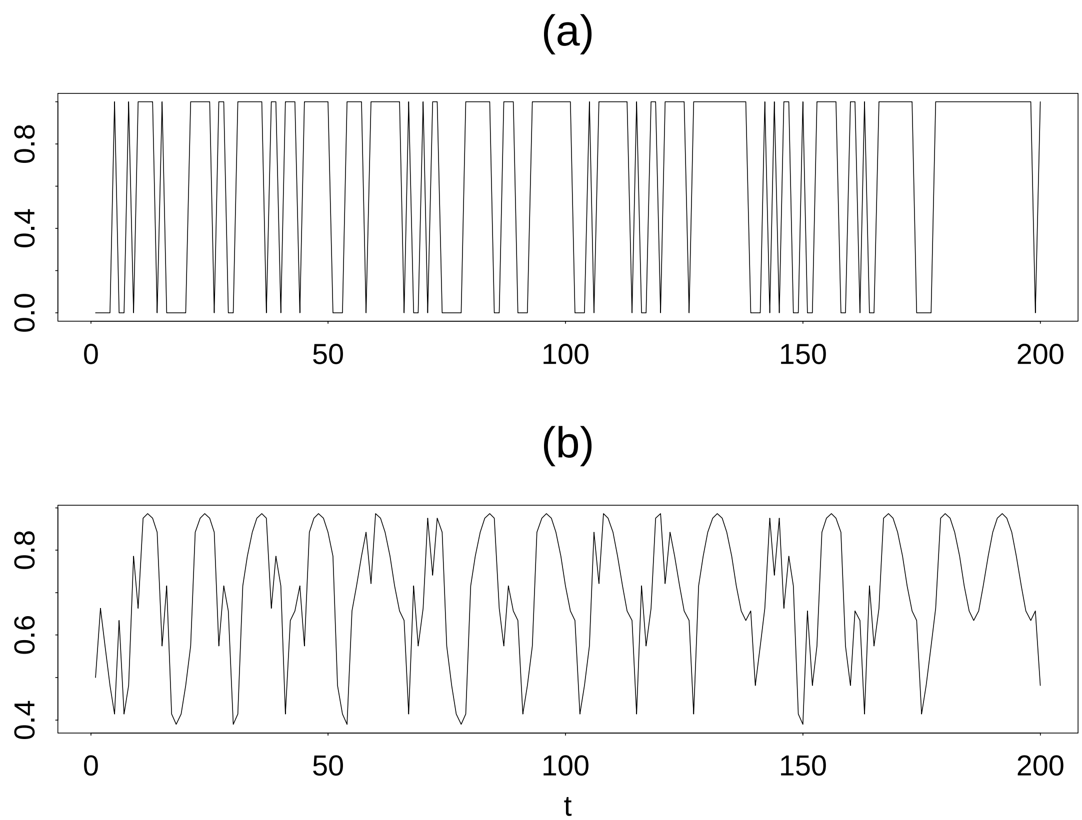
For the canonical link (logistic regression):

$$\frac{\mathbf{G}_N(\boldsymbol{\beta})}{N} \rightarrow \mathbf{G}(\boldsymbol{\beta}) = \int_{\mathcal{R}^p} \frac{e^{\boldsymbol{\beta}' \mathbf{z}}}{(1 + e^{\boldsymbol{\beta}' \mathbf{z}})^2} \mathbf{z} \mathbf{z}' \nu(d\mathbf{z})$$

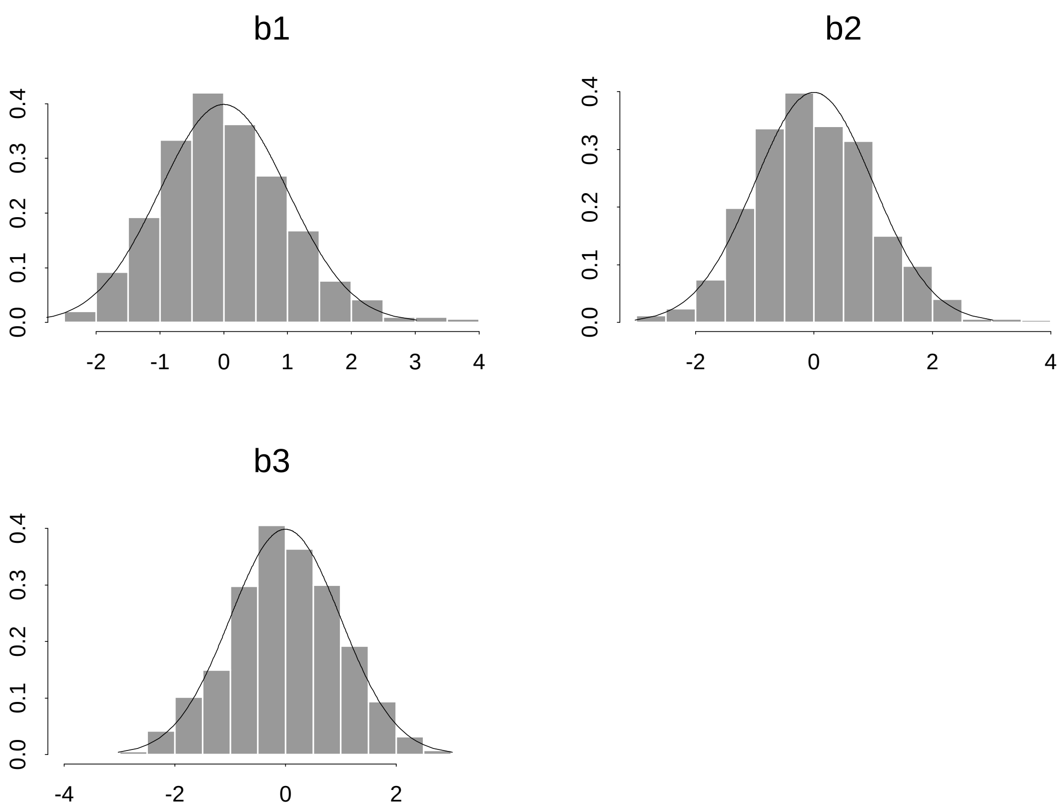
Illustration of asymptotic normality.

$$\text{logit}(\pi_t(\boldsymbol{\beta})) = \beta_1 + \beta_2 \cos\left(\frac{2\pi t}{12}\right) + \beta_3 Y_{t-1}$$

so that $\mathbf{Z}_{t-1} = (1, \cos(2\pi t/12), Y_{t-1})'$.



Logistic autoregression with a sinusoidal component. a. Y_t. b. $\pi_t(\boldsymbol{\beta})$ where $\text{logit}(\pi_t(\boldsymbol{\beta})) = 0.3 + 0.75\cos(2\pi t/12) + y_{t-1}$.



Histograms of normalized MPLE's where $\boldsymbol{\beta} = (0.3, 0.75, 1)'$, $N = 200$. Each histogram consists of 1000 estimates.

Goodness of Fit

$C_1, \cdots, C_k$, a partition of $\mathcal{R}^p$. For $j = 1, \cdots, k$, define,

$$M_j \equiv \sum_{t=1}^{N} I_{[\mathbf{Z}_{t-1} \in C_j]} Y_t$$

and

$$E_j(\boldsymbol{\beta}) \equiv \sum_{t=1}^{N} I_{[\mathbf{Z}_{t-1} \in C_j]} \pi_t(\boldsymbol{\beta})$$

Put:

$\mathbf{M} \equiv (M_1, \cdots, M_k)'$,
$\mathbf{E}(\beta) \equiv (E_1(\boldsymbol{\beta}), \cdots, E_k(\boldsymbol{\beta}))'$.

Slud and K (1994), K and Fokianos (2002):

With

$$\sigma_j^2 \equiv \int_{C_j} F(\boldsymbol{\beta}'\mathbf{z})(1 - F(\boldsymbol{\beta}'\mathbf{z}))\nu(d\mathbf{z})$$

$$\chi^2(\boldsymbol{\beta}) \equiv \frac{1}{N} \sum_{j=1}^{k} (M_j - E_j(\boldsymbol{\beta}))^2 / \sigma_j^2 \rightarrow \chi_k^2.$$

In practice need to adjust the df when replacing $\boldsymbol{\beta}$ by $\widehat{\boldsymbol{\beta}}$.

When $\widehat{\boldsymbol{\beta}}$ and $(\mathbf{M}-\mathbf{E}(\boldsymbol{\beta}))$ are obtained from the *same data set*,

$$\mathsf{E}(\chi^2(\widehat{\boldsymbol{\beta}})) \approx k - \sum_{j=1}^{k} (\mathbf{B}'\mathbf{G}(\boldsymbol{\beta})\mathbf{B})_{jj}/\sigma_j^2$$

When $\widehat{\boldsymbol{\beta}}$ and $(\mathbf{M}-\mathbf{E}(\boldsymbol{\beta}))$ are obtained from *independent data sets*,

$$\mathsf{E}(\chi^2(\widehat{\boldsymbol{\beta}})) \approx k + \sum_{j=1}^{k} (\mathbf{B}'\mathbf{G}(\boldsymbol{\beta})\mathbf{B})_{jj}/\sigma_j^2$$

Illustration of the distribution of $\chi^2(\boldsymbol{\beta})$ using Q-Q plots.

Consider the previous logistic regression model with a periodic component. Use the partition

$$C_1 = \{\mathbf{Z} : Z_1 = 1, -1 \leq Z_2 < 0, Z_3 = 0\}$$

$$C_2 = \{\mathbf{Z} : Z_1 = 1, -1 \leq Z_2 < 0, Z_3 = 1\}$$

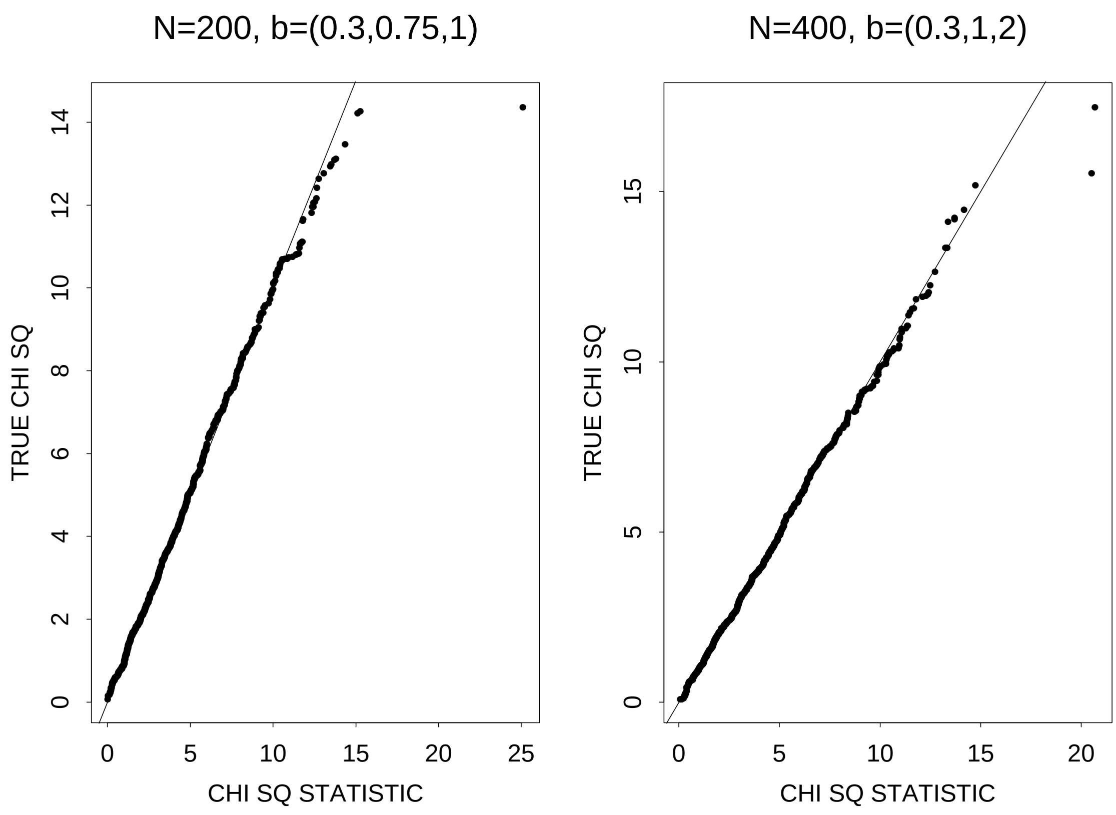
$$C_3 = \{\mathbf{Z} : Z_1 = 1, 0 \leq Z_2 \leq 1, Z_3 = 0\}$$

$$C_4 = \{\mathbf{Z} : Z_1 = 1, 0 \leq Z_2 \leq 1, Z_3 = 1\}$$

Then, k=4, M_j is the sum of those Y_t's for which $\mathbf{Z}_{t-1}$ is in C_j, $j = 1, 2, 3, 4$, and the $E_j(\boldsymbol{\beta})$ are obtained similarly. Estimate σ_j^2 by,

$$\tilde{\sigma}_j^2 = \frac{1}{N} \sum_{t=1}^{N} I_{[\mathbf{Z}_{t-1} \in C_j]} \pi_t(\boldsymbol{\beta})(1 - \pi_t(\boldsymbol{\beta}))$$

The χ^2_4 approximation is quite good.



Example: Modeling successive eruptions of the Old Faithful geyser in Yellowstone National Park, Wyoming.

1 if duration is greater than 3 minutes
0 if duration is less than 3 minutes

101110110101011010110101010
111110101010101010101010101
011111010101011010111011111
011101010101010101010101010
101101010101011101111111011
111011111110101010101011111
101010101110101011010111101
010101110101011011011101010
101101111111010101111011011
101101011101011111011101010
110101111111101010101010101
10

$N = 299$

Candidate models $\eta_t = \boldsymbol{\beta}'\mathbf{Z}_{t-1}$ for Old Faithful.

1	$\beta_0 + \beta_1 Y_{t-1}$
2	$\beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2}$
3	$\beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \beta_3 Y_{t-3}$
4	$\beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \beta_3 Y_{t-3} + \beta_4 Y_{t-4}$

Comparison of models using the logistic link: Model 2 is "best". Probit reg. gives similar results.

Model	p	$\mathcal{X}^2$	D	AIC	BIC
1	2	165.00	227.38	231.38	238.46
2	3	165.00	215.53	221.53	232.15
3	4	165.00	215.08	223.08	237.24
4	5	164.97	213.99	223.99	241.69

$$\widehat{\pi}_t = \pi_t(\widehat{\boldsymbol{\beta}}) = \frac{1}{1 + \exp\left\{-(\widehat{\beta}_0 + \widehat{\beta}_1 Y_{t-1} + \widehat{\beta}_2 Y_{t-2})\right\}}.$$

Part III: Categorical Time Series

EEG sleep state classified or quantized in four categories as follows.

$$\begin{array}{ll} 1: & \text{quiet sleep,} \\ 2: & \text{indeterminate sleep,} \\ 3: & \text{active sleep,} \\ 4: & \text{awake.} \end{array}$$

Here the sleep state categories or levels are assigned integer values. This is an example of a *categorical time series* $\{Y_t\}$, $t = 1, ..., N$, taking the values $1, ..., 4$.

This is an arbitrary integer assignment. Why not the values 7.1, 15.8, 19.24, 71.17 ? Any other scale ?

Assume m categories.

The t'th observation of any categorical time series–regardless of the measurement scale–can be represented by the vector

$$\mathbf{Y}_t = (Y_{t1}, \ldots, Y_{tq})'$$

of length $q = m - 1$, with elements

$$Y_{tj} = \begin{cases} 1, & \text{if the } j\text{th category is observed at time } t \\ 0, & \text{otherwise} \end{cases}$$

for $t = 1, \ldots, N$ and $j = 1, \ldots, q$.

BTS is a special case with $m = 2, q = 1$.

Write for $j = 1, \ldots, q$,

$$\pi_{tj} = \mathsf{E}[Y_{tj} \mid \mathcal{F}_{t-1}] = \mathsf{P}(Y_{tj} = 1 \mid \mathcal{F}_{t-1}),$$

Define:

$$\boldsymbol{\pi}_t = (\pi_{t1}, \ldots \pi_{tq})'$$

$$Y_{tm} = 1 - \sum_{j=1}^{q} Y_{tj}$$

$$\pi_{tm} = 1 - \sum_{j=1}^{q} \pi_{tj}.$$

Let $\{\mathbf{Z}_{t-1}\}$, $t = 1, \ldots, N$, for a $p \times q$ *matrix* that represents a covariate process.

Y_{tj} corresponds to a vector of length p of random time dependent covariates which forms the jth column of $\mathbf{Z}_{t-1}$.

Assume the general regression model:

$$(*) \quad \boldsymbol{\pi}_t(\boldsymbol{\beta}) = \begin{pmatrix} \pi_{t1}(\boldsymbol{\beta}) \\ \pi_{t2}(\boldsymbol{\beta}) \\ \cdots \\ \pi_{tq}(\boldsymbol{\beta}) \end{pmatrix} = \begin{pmatrix} h_1(\mathbf{Z}'_{t-1}\boldsymbol{\beta}) \\ h_2(\mathbf{Z}'_{t-1}\boldsymbol{\beta}) \\ \cdots \\ h_q(\mathbf{Z}'_{t-1}\boldsymbol{\beta}) \end{pmatrix}$$

$$= \mathbf{h}(\mathbf{Z}'_{t-1}\boldsymbol{\beta}).$$

The *inverse* link function $\mathbf{h}$ is defined on R^q and takes values in R^q.

We shall only examine nominal and ordinal time series.

Nominal Time Series.

Nominal categorical variables lack natural ordering.

A model for nominal time series: Multinomial logit model

$$\pi_{tj}(\boldsymbol{\beta}) = \frac{\exp(\boldsymbol{\beta}_j' \mathbf{z}_{t-1})}{1 + \sum_{l=1}^{q} \exp(\boldsymbol{\beta}_l' \mathbf{z}_{t-1})}, \quad j = 1, \ldots, q$$

Note that

$$\pi_{tm}(\boldsymbol{\beta}) = \frac{1}{1 + \sum_{l=1}^{q} \exp(\boldsymbol{\beta}_l' \mathbf{z}_{t-1})}.$$

Multinomial logit model is a special case of (*). Indeed, define $\boldsymbol{\beta}$ to be the $p \equiv qd$-vector

$$\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \ldots, \boldsymbol{\beta}_q')',$$

and $\mathbf{Z}_{t-1}$ the $qd \times q$ matrix

$$\mathbf{Z}_{t-1} = \begin{bmatrix} \mathbf{z}_{t-1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{z}_{t-1} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{z}_{t-1} \end{bmatrix}.$$

Let $\mathbf{h}$ stand for the vector valued function whose components h_j, $j = 1, \ldots, q$, are given by

$$\pi_{tj}(\boldsymbol{\beta}) = h_j(\boldsymbol{\eta}_t) = \frac{\exp(\eta_{tj})}{1 + \sum_{l=1}^{q} \exp(\eta_{tl})}, \; j = 1, \ldots, q$$

with

$$\boldsymbol{\eta}_t = (\eta_{t1}, \ldots, \eta_{tq})' = \mathbf{Z}_{t-1}' \boldsymbol{\beta}.$$

Ordinal Time Series.

Measured on a scale endowed with a natural ordering.

A model for ordinal time series: Need a latent or auxiliary variable.

Put

$$X_t = -\boldsymbol{\gamma}' \mathbf{z}_{t-1} + e_t,$$

1. $e_t \sim$ cdf F i.i.d.
2. $\boldsymbol{\gamma}$ d-dim vector of parameters.
3. $\mathbf{z}_{t-1}$ covariate d-dim vector.

Define a categorical time series $\{Y_t\}$, from the levels of $\{X_t\}$,

$$Y_t = j \Longleftrightarrow Y_{tj} = 1 \Longleftrightarrow \theta_{j-1} \le X_t < \theta_j$$

$$-\infty = \theta_0 < \theta_1 < \ldots < \theta_m = \infty.$$

Then

$$\begin{aligned}\pi_{tj} &= \mathsf{P}(\theta_{j-1} \le X_t < \theta_j \mid \mathcal{F}_{t-1}) \\ &= F(\theta_j + \boldsymbol{\gamma}'\mathbf{z}_{t-1}) - F(\theta_{j-1} + \boldsymbol{\gamma}'\mathbf{z}_{t-1}),\end{aligned}$$

for $j = 1, \ldots, m$.

There are many possibilities depending on F.

Special case: *Proportional Odds Model*,

$$F(x) = F_l(x) = \frac{1}{1 + \exp(-x)}.$$

Then we have for $j = 1, \ldots, q$,

$$\log\left\{\frac{\mathsf{P}[Y_t \le j|\mathcal{F}_{t-1}]}{\mathsf{P}[Y_t > j|\mathcal{F}_{t-1}}\right\} = \theta_j + \boldsymbol{\gamma}'\mathbf{z}_{t-1}$$

Proportional odds model has the form (*) with $p = (q + d)$:

$$\boldsymbol{\beta} = (\theta_1, \ldots, \theta_q, \boldsymbol{\gamma}')'$$

and $\mathbf{Z}_{t-1}$ the $(q + d) \times q$ matrix

$$\mathbf{Z}_{t-1} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ \mathbf{z}_{t-1} & \mathbf{z}_{t-1} & \cdots & \mathbf{z}_{t-1} \end{bmatrix}.$$

Now set

$$\mathbf{h} = (h_1, \ldots, h_q)',$$

and let for $j = 2, \ldots, q$,

$$\begin{aligned} \pi_{t1}(\boldsymbol{\beta}) = h_1(\boldsymbol{\eta}_t) &= F(\eta_{t1}), \\ \pi_{tj}(\boldsymbol{\beta}) = h_j(\boldsymbol{\eta}_t) &= F(\eta_{tj}) - F(\eta_{t(j-1)}), \end{aligned}$$

where

$$\boldsymbol{\eta}_t = (\eta_{t1}, \ldots, \eta_{tq})' = \mathbf{Z}'_{t-1}\boldsymbol{\beta}.$$

Partial likelihood estimation.

Introduce the multinomial probability

$$f(\mathbf{y}_t; \boldsymbol{\beta} \mid \mathcal{F}_{t-1}) = \prod_{j=1}^{m} \pi_{tj}(\boldsymbol{\beta})^{y_{tj}}.$$

The partial likelihood is a product of the multinomial probabilities,

$$\begin{aligned} \mathrm{PL}(\boldsymbol{\beta}) &= \prod_{t=1}^{N} f(\mathbf{y}_t; \boldsymbol{\beta} | \mathcal{F}_{t-1}) \\ &= \prod_{t=1}^{N} \prod_{j=1}^{m} \pi_{tj}^{y_{tj}}(\boldsymbol{\beta}), \end{aligned}$$

so that the partial log-likelihood is given by

$$l(\boldsymbol{\beta}) \equiv \log \mathrm{PL}(\boldsymbol{\beta}) = \sum_{t=1}^{N} \sum_{j=1}^{m} y_{tj} \log \pi_{tj}(\boldsymbol{\beta}).$$

Under a modified Assumption A:

$$\frac{\mathbf{G}_N(\boldsymbol{\beta})}{N} \to \int_{R^{p\times q}} \mathbf{Z}\mathrm{U}(\boldsymbol{\beta})\boldsymbol{\Sigma}(\boldsymbol{\beta})\mathrm{U}'(\boldsymbol{\beta})\mathbf{Z}'\nu(d\mathbf{Z}) = \mathbf{G}(\boldsymbol{\beta})$$

$$\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \to \mathcal{N}_p\left(\mathbf{0}, \mathbf{G}^{-1}(\boldsymbol{\beta})\right)$$

$$\sqrt{N}\left(\boldsymbol{\pi}_t(\widehat{\boldsymbol{\beta}}) - \boldsymbol{\pi}_t(\boldsymbol{\beta})\right) \to \mathcal{N}_q\left(\mathbf{0}, \mathbf{Z}_{t-1}\mathbf{D}_t(\boldsymbol{\beta})\mathbf{G}^{-1}(\boldsymbol{\beta})\mathbf{D}_t'(\boldsymbol{\beta})\mathbf{Z}_{t-1}'\right)$$

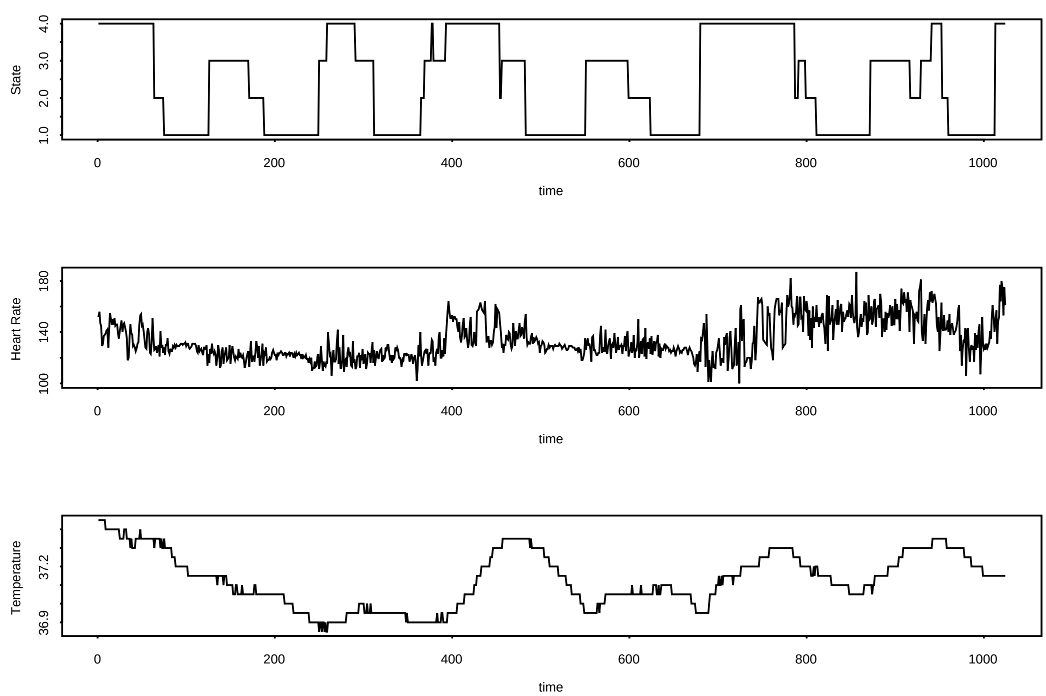
Example: Sleep State.

Covariates: Heart Rate, Temperature.

$N = 700$.

$$\begin{array}{rl} 1: & \text{quiet sleep,} \\ 2: & \text{indeterminate sleep,} \\ 3: & \text{active sleep,} \\ 4: & \text{awake.} \end{array}$$

Ordinal CTS: "4" $<$ "1" $<$ "2" $<$ "3".



Fit proportional odds models.

Model	Covariates	AIC
1	$1+\mathbf{Y}_{t-1}$	401.56
2	$1+\mathbf{Y}_{t-1}+\log R_t$	401.51
3	$1+\mathbf{Y}_{t-1}+\log R_t+T_t$	403.32
4	$1+\mathbf{Y}_{t-1}+T_t$	403.52
5	$1+\mathbf{Y}_{t-1}+\mathbf{Y}_{t-2}+\log R_t$	407.28
6	$1+\mathbf{Y}_{t-1}+\log R_{t-1}$	403.40
7	$1+\log R_t$	1692.31

Model 2: $1 + \mathbf{Y}_{t-1} + \log R_t$

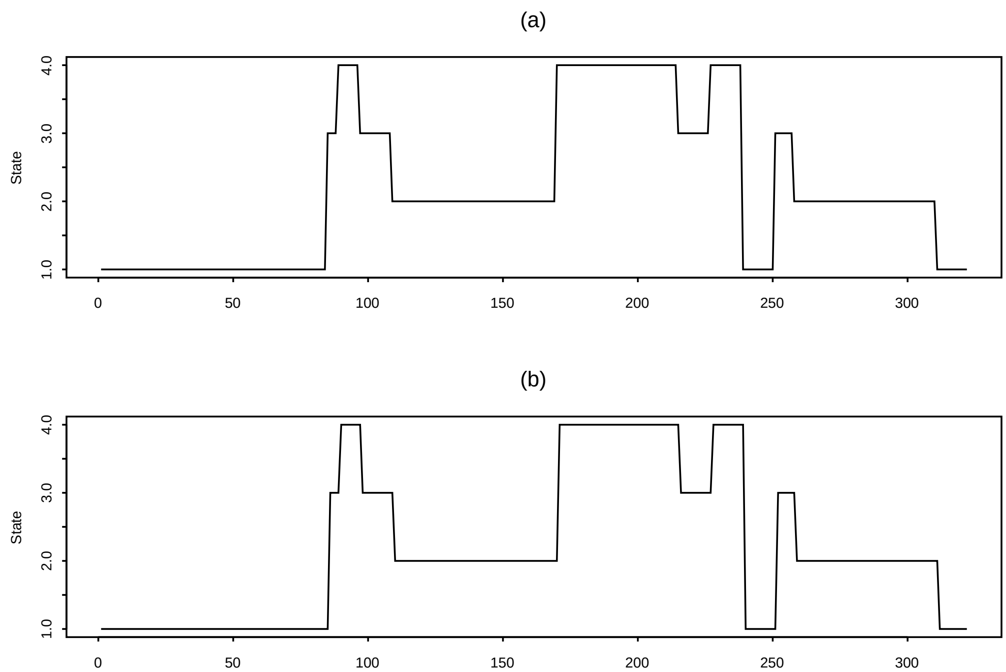
$$\log \left[\frac{\mathsf{P}(Y_t \le "4" \mid \mathcal{F}_{t-1})}{\mathsf{P}(Y_t > "4" \mid \mathcal{F}_{t-1})} \right] =$$
$$\theta_1 + \gamma_1 Y_{(t-1)1} + \gamma_2 Y_{(t-1)2} + \gamma_3 Y_{(t-1)3} + \gamma_4 \log R_t,$$

$$\log \left[\frac{\mathsf{P}(Y_t \le "1" \mid \mathcal{F}_{t-1})}{\mathsf{P}(Y_t > "1" \mid \mathcal{F}_{t-1})} \right] =$$
$$\theta_2 + \gamma_1 Y_{(t-1)1} + \gamma_2 Y_{(t-1)2} + \gamma_3 Y_{(t-1)3} + \gamma_4 \log R_t,$$

$$\log \left[\frac{\mathsf{P}(Y_t \le "2" \mid \mathcal{F}_{t-1})}{\mathsf{P}(Y_t > "2" \mid \mathcal{F}_{t-1})} \right] =$$
$$\theta_3 + \gamma_1 Y_{(t-1)1} + \gamma_2 Y_{(t-1)2} + \gamma_3 Y_{(t-1)3} + \gamma_4 \log R_t,$$

$\widehat{\theta}_1 = -30.352$, $\widehat{\theta}_2 = -23.493$, $\widehat{\theta}_3 = -20.349$, $\widehat{\gamma}_1 = 16.718$, $\widehat{\gamma}_2 = 9.533$, $\widehat{\gamma}_3 = 4.755$, $\widehat{\gamma}_4 = 3.556$.

The corresponding standard errors are 12.051, 12.012, 11.985, 0.872, 0.630, 0.501 and 2.470.



(a) Observed versus (b) predicted sleep states for Model 2 of Table applied to the testing data set. $N = 322$.