Asymptotics for hidden Markov models with covariates

Jens Ledet Jensen

Department of Mathematical Sciences
Aarhus University

New Frontiers in Applied Probability, August 2011

Outline of talk

- Two examples of correlated count data with covariates
- Results: old and new
- Outline of proof:
  - Mixing properties
  - Central limit theorem for “score”
  - Uniform convergence of “information”

Result: asymptotic normality of parameter estimate

Hay and Pettitt

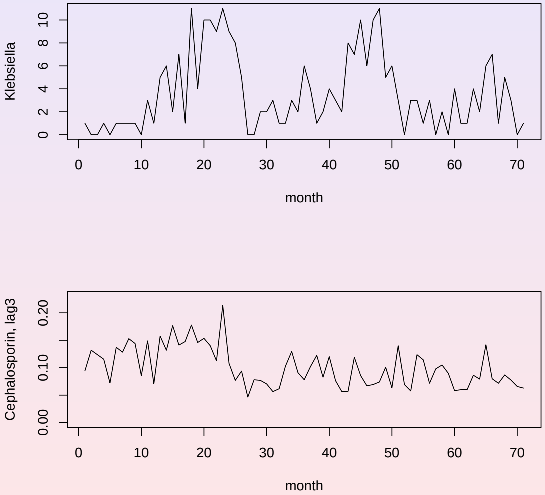
Bayesian analysis of a time series of counts with covariates: an application to the control of an infectious disease

Biostatistics 2001

Counts: Monthly ESBL bacteria producing Klebsiella pneumonia in an Australian hospital
(resistant to many antibiotics, first outbreak in Denmark: 2007)

Covariate: the amount of antibiotic cephalosporins used, lagged 3 months

Plot of data


Klebsiella
0
2
4
6
8
10
0
10
20
30
40
50
60
70
month
Cephalosporin, lag3
0.00
0.10
0.20
0
10
20
30
40
50
60
70
month

Model

y_i: count, z_i:covariate

$$y_i|\mu_i \sim \text{Poisson}(\mu_i)$$

$$\mu_i = \exp(\beta z_i + x_i), \qquad x_i = \phi x_{i-1} + N(0, \sigma^2)$$

Homogeneous hidden Markov and non-homogeneous emission probabilitites

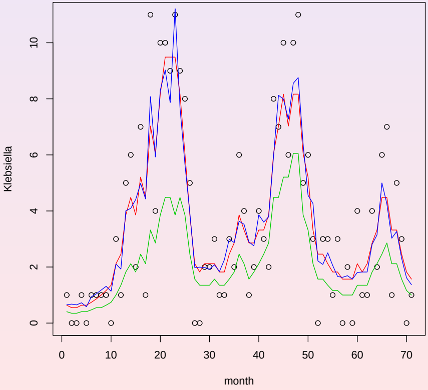
Fully bayesian analysis using MCMC: posterior quantities:

parameter	mean	2.5%	97.5%
β	6.9	1.7	11.1

My run 1

Discretize hidden state space (x_i): 41 points (truncated AR(1))

$\hat{\beta} = 5.0$, likelihood ratio test, $\beta = 0$: 5.5%



red: $\beta = 0$, MAP of $\exp(x_i)$

blue: $\beta = 5$, MAP of $\exp(\beta z_i + x_i)$

green: $\beta = 5$, MAP of $\exp(x_i)$

My run 2

$x_i = (\tilde{x}_i, u_i)$ is a Markov chain on $\{0.5, 1, 2, \ldots, 9\} \times \{-1, 0, 1\}$

first coordinate: hidden mean
second coordinate: increase or decrease at last step

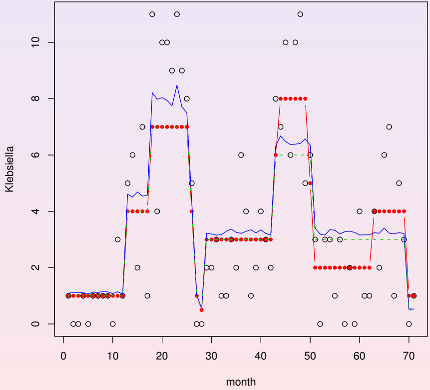
u_i :

	-1	0	1
-1	$1-\alpha$	α	0
0	$\rho(1-\gamma)$	γ	$(1-\rho)(1-\gamma)$
1	0	α	$1-\alpha)$

$\tilde{x}_i$: decrease: $i \to i-1$ or $i-2$ or $i-3$
increase: $i \to i+1$ or $i+2$ or $i+3$

My run 2

$\hat{\beta} = 0.9$, likelihood ratio test, $\beta = 0$: 40%



red: $\beta = 0$, MAP of $\tilde{x}_i$

blue: $\beta = 0.9$, MAP of $\tilde{x}_i \exp(\beta z_i)$

green: $\beta = 0.9$, MAP of $\tilde{x}_i$

Model problem: roles of covariate and hidden variable are not well separated

Jørgensen, Lundbye-Christensen, Song, Sun

A longitudinal study of emergency room visits and air pollution for Prince Gorge, British Columbia

Statistics in Medicine 1996

Counts: daily counts of emergency room visits for four repiratory diseases

Covariates: 4 meteorological ($\tilde{z}$) and 2 air pollution (z)

Plot of Data

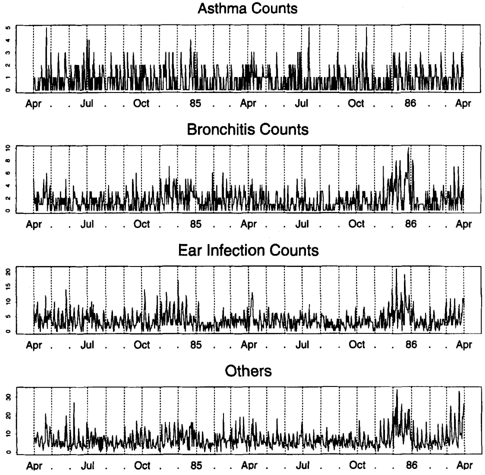


Figure 1. Emergency room visits for four respiratory diseases

Plot of Data

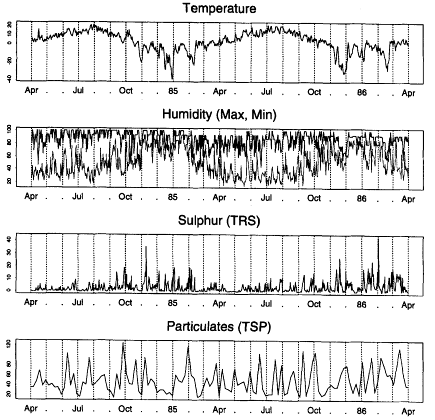


Figure 2. Meteorological and air pollution variables

Model

$$y_{it}|x_t \sim \text{Poisson}(a_{it}x_t)$$

$$a_{it} = \exp(\alpha_i \tilde{z}_i)$$

$$x_t|x_{t-1} \sim \text{Gamma}(E = b_t x_{t-1}, \text{Var} = b_t^2 x_{t-1}\sigma^2)$$

$$b_t = \exp(\beta(z_t - z_{t-1}))$$

Non-homogeneous hidden Markov and non-homogeneous emission probabilitites

Analysis via approximate Kalman filter

General model in this talk

x_i: non-homogeneous Markov chain, not observed

transition density: $p_i(x_i|x_{i-1};\theta)$

y_i: conditionally independent given $(x_1,\ldots,x_n)$, observed
conditional distribution depends on x_i only

emission density: $g_i(y_i|x_i;\theta)$

Covariates: enters through the index i on p and g

Papers: setup

	state spaces: hidden	observed
Baum and Petrie 1966	finite	finite
Bickel, Ritov and Rydén 1998	finite	general
Jensen and Petersen 1999	$\sim$general	general
Douc, Moulines and Rydén 2004	$\sim$general	general, AR(1)
Jensen 2005	finite	finite
Fuh 2006	(general)	(general)

All except J 2005: homogeneous Markov, homogeneous emission

Result: there exists solution $\hat{\theta}$ to likelihood equations with $\sqrt{n}(\hat{\theta} - \theta)$ asymptotically normal

Fuh, Ann.Statist. 2006

Appears very general

Example from paper: x_i is AR(1), $y_i = x_i + N(0, 1)$

But: there are serious errors in the paper

results cannot be trusted

Conditions on hidden variable

All papers and here:

$$0 < \sigma_- \leq p_i(\cdot|\cdot;\theta) \leq \sigma_+ < \infty, \theta \in B_0$$

upper bounds on log derivatives of $p_i(\cdot|\cdot;\theta)$

moments of upper bound of log derivatives of $g_i(y_i|\cdot;\theta)$

Not covered: x_i is an AR(1)

Conditions on observed variable

BRR 1998, JP 1999:

condition on $\max_{a,b} \frac{g(y|a;\theta)}{g(y|b;\theta)}$

to control mixing properties of $x|y$

Conditions on observed variable

BRR 1998, JP 1999:

condition on $\max_{a,b} \frac{g(y|a;\theta)}{g(y|b;\theta)}$

to control mixing properties of $x|y$

DMR 2004: simple trick to avoid this

(choosing a different dominating measure, dependent on i)

same trick used here:

for all $i, y_i, \theta \in B_0$: $0 < \int g_i(y_i|x_i;\theta)\mu(dx_i)) < \infty$

Covered: $y_i|x_i \sim \text{poisson}(\exp(\beta z_i + x_i))$,

x: finite state space, z: bounded

Conditions: Estimating equation

Previous papers: $\hat{\theta} =$ MLE

Here: Find $\hat{\theta}$ by solving

$$S_n(\theta) = \sum_{i=1}^{n} E_\theta\{\psi_i(\theta; \bar{x}_i, y_i)|y_1, \dots, y_n\} = 0$$

$$\bar{x}_i = (x_{i-1}, x_i, x_{i+1}), \qquad E_\theta \psi_i(\theta) = 0$$

MLE: $\psi_i(\theta; \bar{x}_i, y_i) = D^1 \log(p_i(x_i|x_{i-1}; \theta) g_i(y_i|x_i; \theta))$

moments of upper bound of $\psi_i(\cdot; \cdot, y_i)$ and $D\psi_i(\cdot; \cdot, y_i)$

Example: estimating equation

x_i: finite state

$y_i|x_i$: Ising lattice field on $\{1, 2, \ldots, k\}^2$

$$g_i(y_i|x_i) = c(\beta(x_i)) \exp\left[\beta(x_i) \sum_{u\sim v} y_{iu} y_{iv}\right], \quad y_{iu} \in \{-1, 1\}$$

$c(\beta)$ is unknovn: use pseudolikelihood $\to \psi$

$$D_1 \log g_i(y_i|x_i) \to D_1 \log \prod_u p_i(y_{iu}|y_{i,(-u)}, x_i)$$

Estimation: Quasilikelihood

Zeger: 1988

Solve $M(\theta)(y - \mu(\theta))$

Asymptotics is ‘simple’: $\sum_i h(y_i)$: mixing of y_i’s

Here: MLE or Estimating equation: each term in sum depends on all y_i’s!

Way of thinking (arbitrary silly covariate sequence)

$$J_n(\theta) = -DS_n(\theta), \quad \gamma(n,\delta) = \sup_{\theta \in B(\delta)} \left| \frac{1}{n}(J_n(\theta) - J_n(\theta_0)) \right|$$

Assume:

(i) $\frac{1}{n} J_n(\theta_0) - F_n \xrightarrow{P} 0, \quad F_n$ nonrandom, $\text{eigen}(F_n) > c_0$

(ii) $\gamma(n, \delta_n) \xrightarrow{P} 0$ for any $\delta_n \to 0$

(iii) $\frac{1}{\sqrt{n}} S_n(\theta_0) G_n^{-1/2} \xrightarrow{D} N_p(0, I), \quad c_1 < \text{eigen}(G_n) < c_2$

Result: $\sqrt{n}(\hat{\theta}_n - \theta_0)(\frac{1}{n} J_n) G_n^{-1/2} \xrightarrow{D} N_p(0, I)$ for any consistent $\hat{\theta}$.

Mixing: basic

Conditional process $(x_1, \ldots, x_n)|(y_1, \ldots, y_n)$

General: density $c \prod_{k=1}^{n} p_k(x_k|x_{k-1}) g_k(x_k)$

transition density wrt μ: $p_k(x_k|x_{k-1}) g_k(x_k) a_k(x_k)/a_{k-1}(x_{k-1})$

define μ_k by $\frac{d\mu_k}{d\mu}(x_k) = g_k(x_k) a_k(x_k)/\int g_k(z) a_k(z) \mu(dz)$

transition density $q_k(x_k|x_{k-1})$ wrt μ_k:

$$p_k(x_k|x_{k-1})/\int p_k(z|x_{k-1}) \mu_k(dz)$$

Bounds: $\frac{\sigma_-}{\sigma_+} \leq q_k(x_k|x_{k-1}) \leq \frac{\sigma_+}{\sigma_-}$ from

$$\sigma_- \leq p_k(x_k|x_{k-1}) \leq \sigma_+$$

Two sided: $\left(\frac{\sigma_-}{\sigma_+}\right)^2 \leq q_k(x_k|x_{k-1}, x_{k+1}) \leq \left(\frac{\sigma_+}{\sigma_-}\right)^2$

Mixing

Chain: $c \prod_{k=1}^{n} p_k(x_k|x_{k-1}) g_k(x_k)$

Let $r < s$ and $\rho = 1 - \sigma_-/\sigma_+$, then
$\sup_u P(x_s \in A|x_r = u) - \inf_v P(x_s \in A|x_r = v) \leq \rho^{s-r}$,

Let $r < s_1 \leq s_2 < t$ and $\tilde{\rho} = 1 - (\sigma_-/\sigma_+)^2$, then
$\sup_{a,b} P(x_{s_1}^{s_2} \in B|x_r = a, x_t = b)$
$- \inf_{u,v} P(x_{s_1}^{s_2} \in B|x_r = u, x_t = v) \leq \tilde{\rho}^{s_1-r} + \tilde{\rho}^{t-s_2}$

Iterative argument: Doob 1953!

(Generalization: perhaps read and understand Meyn and Tweedie: Markov Chains and Stochastic Stability)

Central limit theorem

$S_n = \sum_{i=1}^{n} E(\psi_i | y_1, \ldots, y_n)$

Mixing properties of summands ? Not so obvious

Instead:

$$E(\psi_i | y_1^n) - E(\psi_i | y_{i-l}^{i+l})| \leq 4(\sup_{\bar{x}_i} \psi_i)\tilde{\rho}^{l-1}$$

General CLT based on Göetze and Hipp, 1982

$S_n = \sum_{i=1}^n Z_i, \quad E(Z_i) = 0, \quad E|Z_i|^{2+\epsilon} \leq K_0$

σ-algebras $\mathcal{D}_j$:

$|P(A_1 \cap A_2) - P(A_1)P(A_2)| \leq \gamma_0 |I_1|^{\gamma_1} |I_2|^{\gamma_2} \text{dist}(I_1, I_2)^{-\lambda}$

for $A_i \in \sigma(\mathcal{D}_j : j \in I_i)$

$E|Z_j - Z_j(m)| \leq K_1 m^{-\lambda}, \quad z_j(m)$ is $\sigma(\mathcal{D}_i : |i-j| \leq m)$-measurable

$\text{eigen}(\frac{1}{n}\text{Var}S_n) \geq c_0$

Then: $S_n \text{Var}(S_n)^{-1/2} \xrightarrow{D} N_p(0, I)$

Uniform convergence of information

$$J_n(\theta) = -\frac{\partial}{\partial\theta} S_n(\theta), \qquad \omega_i = \log[p_i(x_i|x_{i-1})g_i(y_i|x_i)]$$

$$J_n(\theta) = -\sum_{i=1}^n E_\theta\left[\frac{\partial}{\partial\theta}\psi_i(\theta)|y_1^n\right]$$

$$-\sum_{i,j=1}^n \mathrm{Cov}_\theta\left[\psi_i(\theta), \frac{\partial}{\partial\theta}\omega_j(\theta)|y_1^n\right]$$

Difference of two conditional means

$$|E_\theta[b(x_r^s)|y_1^n] - E_{\theta_0}[b(x_r^s)|y_1^n]|$$

$$\leq b^0\{2p|\theta - \theta_0| \textstyle\sum_{i=r-l+1}^{s+l} h_i(y_i) + 8\tilde{\rho}^l\}, \quad \tilde{\rho} = 1 - \left(\frac{\sigma_-}{\sigma_+}\right)^2$$

b^0: upper bound on $b(x_r^s)$

$h_i(y_i) = \sup_{x_{i-1}, x_i, \theta \in B_0, r} |\frac{\partial}{\partial \theta_r}\omega_i(\theta)|$

$\omega_i = \log[p_i(x_i|x_{i-1})g_i(y_i|x_i)]$

Difference of two conditional covariances

$$E_\theta(a_u b_v | y_1^n) - E_{\theta_0}(a_u b_v | y_1^n) \le$$

$$a_u^0 b_u^0 \left[2p|\theta - \theta_0| \sum_{i=u-l}^{v+1+l} h_i(y_i) + 8\tilde{\rho}^l \right]$$

$$E_\theta(a_u | y_1^n) E_\theta(b_v | y_1^n) - E_{\theta_0}(a_u | y_1^n) E_{\theta_0}(b_v | y_1^n)$$

$$\le a_u^0 b_u^0 \left[2p|\theta - \theta_0| \{ \sum_{i=u-l}^{u+1+l} h_i(y_i) + \sum_{i=v-l}^{v+1+l} h_i(y_i) \} + 16\tilde{\rho}^l \right]$$

Use this when u and v are close
otherwise: bound of Ibragimov and Linnik
on covariances for mixing sequences

Nonrandom limit of observed information

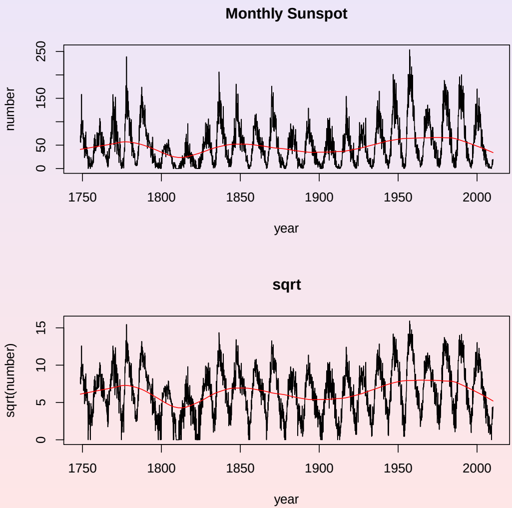
$$J_n(\theta) = -\sum_{i=1}^n E_\theta\left[\frac{\partial}{\partial\theta}\psi_i(\theta)|y_1^n\right]$$
$$-\sum_{i,j=1}^n \mathrm{Cov}_\theta\left[\psi_i(\theta), \frac{\partial}{\partial\theta}\omega_j(\theta)|y_1^n\right]$$

$$\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^n E(a_u|y_1^n)\right) = O(1/n)$$

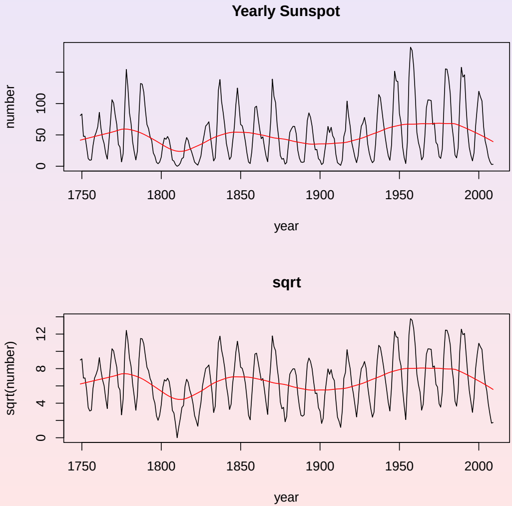
$$\mathrm{Var}\left(\frac{1}{n}\sum_{u,v=1}^n \mathrm{Cov}(a_u, b_v|y_1^n)\right) \to 0$$

End of proof!

Sunspot numbers: monthly



Sunspot numbers: yearly



Model

$$y_i|x_i \sim N(h(x_i), \sigma^2)$$

$$x_i = (t_i, w_i), \quad t_i \in \{0, 2, \ldots, 53\} \quad w_i \in \{3, 4, 5, 6, 7\}$$

$$t_{i+1} = t_i + w_{i+1} (\text{mod } 54)$$

$p(w_{i+1}|w_i)$ some persistence (slow period / fast period)

$$h(x_i) = h(t_i) = \begin{cases} 2 + t_i \frac{3}{10} & 0 \leq t_i \leq 20, \\ 8 - (t_i - 20)\frac{3}{17} & 20 < t_i < 54. \end{cases}$$

Model

$p(w_{i+1}|w_i)$:

	3	4	5	6	7
3	ρ	$(1-\rho)/2$	$(1-\rho)/2$	0	0
4	$(1-\rho)/3$	ρ	$(1-\rho)/3$	$(1-\rho)/3$	0
5	$(1-\rho)/4$	$(1-\rho)/4$	ρ	$(1-\rho)/4$	$(1-\rho)/4$
6	0	$(1-\rho)/3$	$(1-\rho)/3$	ρ	$(1-\rho)/3$
7	0	0	$(1-\rho)/2$	$(1-\rho)/2$	ρ

stationary: $(\frac{2}{14}, \frac{3}{14}, \frac{4}{14}, \frac{3}{14}, \frac{2}{14})$

Simulations

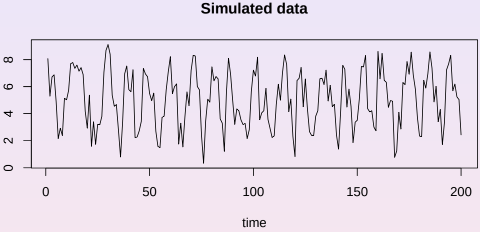
Simulate $n = 200$ observations — Find $(\hat{\rho}, \hat{\sigma})$

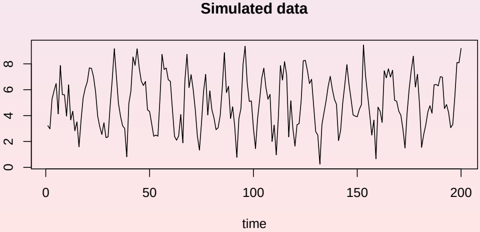
We use $\theta = \log(\hat{\rho}/(1-\hat{\rho}))$ and $\log(\hat{\sigma})$

Repeat this 500 times

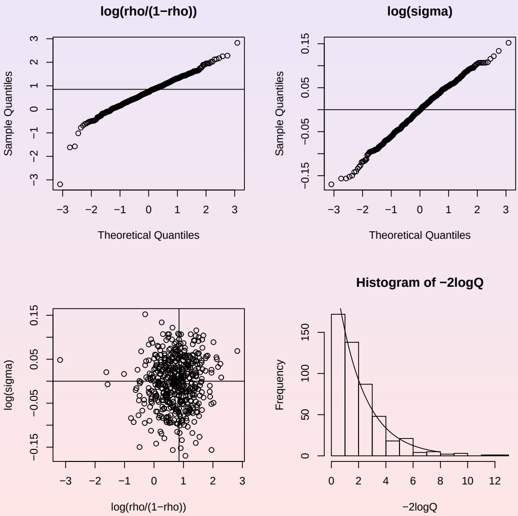
Simulations: $\rho = 0.7$ ($\theta = 0.85$), $\sigma = 1$

Simulated data





Asymptotic normality ?



Sqrt of yearly sunspot numbers

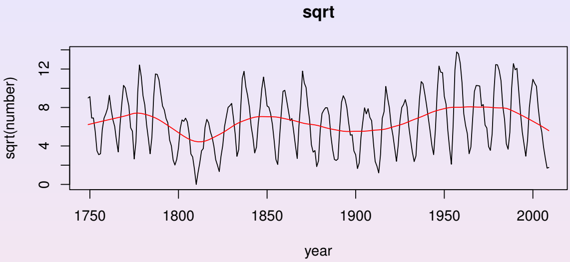
Trend: (Gleissberg cycle)

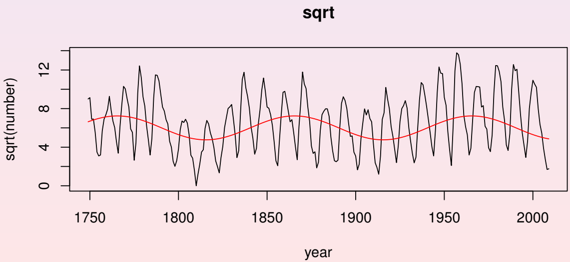
$$E(y_i|x_i) = h(x_i) + \beta_1 \cos(2\pi t/100) + \beta_2 \sin(2\pi t/100)$$

$$\hat{\rho} = 0.38, \qquad \hat{\sigma} = 1.22$$

$$\hat{\beta}_1 = -1.19, \qquad \hat{\beta}_2 = 0.35$$

Sunspot numbers: yearly





End of talk

Questions:

- Remove compactness assumption on state space
- How to do model check for hidden Markov model ?
- Interplay between hidden variable and covariates ?
- Interplay between flexibility in hidden variable and σ^2 ? $(y_i|x_i \sim N(h(x_i), \sigma^2))$