Estimation for Nonlinear State-Space Models

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

Joint work with:
William Dunsmuir, University of New South Wales
Gabriel Rodriguez-Yam, Colorado State University
Ying Wang, Dept of Public Health, W. Virginia
Two CSU Statistics Department Projects

**STARMAP**

*Space-Time Aquatic Resources Modeling and Analysis Program (funded by the EPA-STAR program)*

**PRIMES**

*PRogram for Interdisciplinary Mathematics, Ecology, and Statistics (recommended for funding by NSF IGERT)*

- Student fellowships
- Postdoctoral fellowships
- Short-term visitors
- Workshops/seminars and more.
Nonlinear state-space models
• Observation driven
• Parameter driven

Innovations algorithm (recursive one-step ahead prediction algorithm)
• Applications
  - Gaussian likelihood calculations
  - simulation
  - generalized least squares estimation

Time series of counts
• Examples (asthma data, polio data)
• Generalized linear models (GLM)
• Estimating equations (Zeger)
• MCEM (Chan and Ledolter)
• Importance sampling
  - Durbin and Koopman
• Approximation to the likelihood (Davis, Dunsmuir, and Wang)
• Simulation results

Examples
Generalized State-Space Models (parameter driven)

Observations: \( y_t = (y_1, \ldots, y_t) \)

States: \( \alpha_t = (\alpha_1, \ldots, \alpha_t) \)

Observation equation:
\[
p(y_t | \alpha_t) := p(y_t | \alpha_t, \alpha_{t-1}, y_{t-1})
\]

State equation:
\[
p(\alpha_{t+1} | \alpha_t) := p(\alpha_{t+1} | \alpha_t, \alpha_{t-1}, y_t)
\]

Joint density:
\[
p(y_1, \ldots, y_n, \alpha_1, \ldots, \alpha_n)
\]
\[
= p(y_n | \alpha_n, \alpha_{n-1}, y_{n-1})p(\alpha_n, \alpha_{n-1}, y_{n-1})
\]
\[
= p(y_n | \alpha_n) p(\alpha_n | \alpha_{n-1}, y_{n-1}) p(\alpha_{n-1}, y_{n-1})
\]
\[
= \cdots
\]
\[
= \left( \prod_{j=1}^n p(y_j | \alpha_j) \right) \left( \prod_{j=2}^n p(\alpha_j | \alpha_{j-1}) \right) p(\alpha_1)
\]
Conditional independence:

\[ p(y_1, \ldots, y_n \mid \alpha_1, \ldots, \alpha_n) = \prod_{j=1}^{n} p(y_j \mid \alpha_j) \]

Filtering or posterior density:

\[ p(\alpha_t \mid y_t) = \frac{p(y_t \mid \alpha_t) p(\alpha_t \mid y_{t-1})}{p(y_t \mid y_{t-1})} \]

Predictive densities:

\[ p(\alpha_{t+1} \mid y_t) = \int p(\alpha_t \mid y_t) p(\alpha_{t+1} \mid \alpha_t) d\mu(\alpha_t) \]
\[ p(y_{t+1} \mid y_t) = \int p(y_{t+1} \mid \alpha_{t+1}) p(\alpha_{t+1} \mid y_t) d\mu(\alpha_{t+1}) \]
Examples of parameter driven models

Poisson model for time series of counts

Observation equation:
\[ p(y_t | \alpha_t) = \frac{e^{\alpha_t y_t} e^{-\alpha_t}}{y_t!}, \quad y_t = 0, 1, \ldots, \]

State equation: State variables follow a regression model with Gaussian AR(1) noise
\[ \alpha_t = \beta^T x_t + W_t, \quad W_t = \phi W_{t-1} + Z_t, \quad \{Z_t\} \sim \text{WN}(0, \sigma^2) \]

The resulting transition density of the state variables is
\[ p(\alpha_{t+1} | \alpha_t) = n(\alpha_{t+1} ; \beta^T x_{t+1} + \phi (\alpha_t - \beta^T x_t), \sigma^2) \]

Remark: The case \( \sigma^2 = 0 \) corresponds to a log-linear model with Poisson noise.
Examples of parameter driven models

A stochastic volatility model for financial data (Taylor `86):
Model:

\[ Y_t = \sigma_t Z_t, \{Z_t\} \sim \text{IID} \ N(0,1) \]

\[ \alpha_t = \phi \alpha_{t-1} + W_t, \ \{W_t\} \sim \text{IID} \ N(0,\sigma^2), \]

where \( \alpha_t = \log \sigma_t \).

The resulting observation and state transition densities are

\[ p(y_t | \alpha_t) = n(y_t ; 0, \exp(2\alpha_t)) \]

\[ p(\alpha_{t+1} | \alpha_t) = n(\alpha_{t+1} ; \phi \alpha_t, \sigma^2) \]

Properties:

- Martingale difference sequence.
- Stationary.
- Strongly mixing at a geometric rate.
The Innovations Algorithm

Innovations Algorithm (Brockwell and Davis `87): \( \{X_t\} \) is a zero-mean time series with ACVF \( \kappa(i,j) \), then

\[
\hat{X}_{t+1} = P_{sp(1,x_1,\ldots,x_1)} X_{t+1} = \theta_{t1} (X_t - \hat{X}_t) + \cdots + \theta_{tt} (X_1 - \hat{X}_1)
\]

The coefficients \( \theta_{t1}, \ldots, \theta_{tt} \) and prediction errors \( v_{t-1} \) can be computed recursively from the equations,

\[
v_0 = \kappa(1,1)
\]

\[
\theta_{t,t-k} = \left[ \kappa(t+1,k+1) - \sum_{j=0}^{k-1} \theta_{k,k-j} \theta_{t,t-j} v_j \right] v_{k-1}^{-1}, \ k = 0,\ldots,t-1,
\]

and

\[
v_t = \kappa(t+1,t+1) - \sum_{j=0}^{t-1} \theta_{t,t-j}^2 v_j.
\]
Remarks:

- Innovations algorithm expresses one-step predictor in terms of previous innovations, $X_1 - \hat{X}_1, \ldots, X_t - \hat{X}_t$, that are uncorrelated.

- If $\{X_t\}$ is an MA(q) process
  
  $$X_{t+1} = Z_{t+1} + \theta_1 Z_t + \cdots + \theta_q Z_{t-q}, \quad \{Z_t\} \sim WN(0, \sigma^2)$$

  then $(\theta_{t1}, \ldots, \theta_{tt}) = (\theta_{t1}, \ldots, \theta_{tq}, 0, \ldots, 0)$ for all $t$.

- Innovations algorithm is well adapted for ARMA(p,q) models—only need to apply to MA(q) piece (see B&D `96).
The Innovations Algorithm—Applications

Likelihood calculation:

Using the IA representation,

$$\hat{X}_t = \theta_{t-1,1}(X_{t-1} - \hat{X}_{t-1}) + \cdots + \theta_{t-1,t-1}(X_1 - \hat{X}_1)$$

we have

$$\begin{bmatrix}
X_1 \\
X_2 \\
\vdots \\
X_n
\end{bmatrix} = 
\begin{bmatrix}
1 & 0 & 0 & \cdots & 0 \\
\theta_{1,1} & 1 & 0 & \cdots & 0 \\
\theta_{2,2} & \theta_{2,1} & 1 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\theta_{n-1,n-1} & \theta_{n-1,n-2} & \theta_{n-1,n-3} & \cdots & 1
\end{bmatrix}
\begin{bmatrix}
X_1 - \hat{X}_1 \\
X_2 - \hat{X}_2 \\
\vdots \\
X_n - \hat{X}_n
\end{bmatrix}$$

$$X_n = C_n (X_n - \hat{X}_n)$$

By taking covariances of both sides it follows that

$$\Gamma_n = E(X_n X'_n) = C_n D_n C'_n, \quad D_n = \text{diag}(v_0, \ldots, v_{n-1})$$
The Innovations Algorithm—Applications

Quadratic form:

\[
\mathbf{X}_n' \mathbf{\Gamma}_n^{-1} \mathbf{X}_n = (\mathbf{X}_n - \hat{\mathbf{X}}_n)' \mathbf{C}_n' \mathbf{(C}_n'^{-1} \mathbf{D}_n^{-1} \mathbf{C}_n^{-1}) \mathbf{C}_n (\mathbf{X}_n - \hat{\mathbf{X}}_n) \\
= (\mathbf{X}_n - \hat{\mathbf{X}}_n)' \mathbf{D}_n^{-1} (\mathbf{X}_n - \hat{\mathbf{X}}_n) \\
= \sum_{t=1}^{n} (X_t - \hat{X}_t)^2 / \nu_{t-1}
\]

Determinant:

\[
\text{det}(\mathbf{\Gamma}_n) = \text{det}(\mathbf{C}_n \mathbf{D}_n \mathbf{C}_n') = \nu_0 \cdots \nu_{n-1}
\]

Gaussian likelihood:

\[
L(\mathbf{\Gamma}_n) = (2\pi)^{-n/2} (\nu_0 \cdots \nu_{n-1})^{-1/2} \exp\{ -1/2 \sum_{t=1}^{n} (X_t - \hat{X}_t)^2 / \nu_{t-1} \}
\]

Simulation: If \( \{Z_t\} \sim \text{iid N}(0,1) \), put \( X_t = \nu_{t-1}^{1/2} Z_t + \theta_{t-1,1} \nu_{t-2}^{1/2} Z_{t-1} + \cdots + \theta_{t-1,t-1} \nu_0^{1/2} Z_1 \).

Then \( \mathbf{X}_n = (X_1, \cdots, X_n)' = \mathbf{C}_n' \mathbf{D}_n^{-1/2} \mathbf{Z}_n \)

has covariance matrix \( \mathbf{\Gamma}_n \).
Count data: $Y_1, \ldots, Y_n$

Regression (explanatory) variable: $x_t$

Model: Distribution of the $Y_t$ given $x_t$ and a stochastic process $\alpha_t$ are independent. Poisson distributed with mean

$$\mu_t = \exp(x_t^T \beta + \alpha_t).$$

The distribution of the stochastic process $\alpha_t$ may depend on a vector of parameters $\gamma$.

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

Primary objective: Inference about $\beta$. 
Polio Data With Estimated Regression Function
Parameter-Driven Model for the Mean Function $\mu_t$

**Parameter-driven specification:** (Assume $Y_t | \mu_t$ is Poisson($\mu_t$))

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

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

**e.g.** (AR(1) process)

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

Advantages of this model specification:

- Properties of model (ergodicity and mixing) easy to derive.
- Interpretability of regression parameters

$$E(Y_t) = \exp(x_t^T \beta) \exp(\alpha_t) = \exp(x_t^T \beta), \quad \text{if } E\exp(\alpha_t) = 1.$$ 

Disadvantages:

- Estimation is difficult-likelihood function not easily calculated (MCEM, importance sampling, estimating eqns).
- Model building can be laborious

**Remark:** See Davis, Dunsmuir, and Wang (1999) for testing of the existence of a latent process and estimating its ACF.
Estimation Methods — Importance Sampling (Durbin and Koopman)

Model:

\[ Y_t \mid \alpha_t, x_t \sim \text{Pois}(\exp(x_t^T \beta + \alpha_t)) \]
\[ \alpha_t = \phi \alpha_{t-1} + \varepsilon_t, \quad \{\varepsilon_t\} \sim \text{IID N}(0, \sigma^2) \]

Relative Likelihood: Let \( \psi = (\beta, \phi, \sigma^2) \) and suppose \( g(y_n, \alpha_n; \psi_0) \) is an approximating joint density for \( Y_n = (Y_1, \ldots, Y_n)' \) and \( \alpha_n = (\alpha_1, \ldots, \alpha_n)' \).

\[
L(\psi) = \int p(y_n \mid \alpha_n) p(\alpha_n) d\alpha_n \\
= \int \frac{p(y_n \mid \alpha_n) p(\alpha_n)}{g(y_n, \alpha_n; \psi_0)} g(y_n, \alpha_n; \psi_0) d\alpha_n \\
= \int \frac{p(y_n \mid \alpha_n) p(\alpha_n)}{g(y_n, \alpha_n; \psi_0)} g(\alpha_n \mid y_n; \psi_0) g(y_n; \psi_0) d\alpha_n \\
\frac{L(\psi)}{L_g(\psi_0)} = \int \frac{p(y_n \mid \alpha_n) p(\alpha_n)}{g(y_n, \alpha_n; \psi_0)} g(\alpha_n \mid y_n; \psi_0) d\alpha_n
\]
Importance Sampling (cont)

\[
\frac{L(\psi)}{L_g(\psi_0)} = \int \frac{p(y_n | \alpha_n) p(\alpha_n)}{g(y_n, \alpha_n; \psi_0)} g(\alpha_n | y_n; \psi_0) d\alpha_n \\
= E_g \left[ \frac{p(y_n | \alpha_n) p(\alpha_n)}{g(y_n, \alpha_n; \psi_0)} \right] y_n; \psi_0 \right] \\
\sim \frac{1}{N} \sum_{j=1}^{N} \frac{p(y_n | \alpha_n^{(j)}) p(\alpha_n^{(j)})}{g(y_n, \alpha_n^{(j)}; \psi_0)},
\]

where \( \{\alpha_n^{(j)}; j = 1, \ldots, N\} \sim \text{iid } g(\alpha_n | y_n; \psi_0) \).

Notes:

- This is a “one-sample” approximation to the relative likelihood. That is, for one realization of the \( \alpha \)'s, we have, in principle, an approximation to the whole likelihood function.

- Approximation is only good in a neighborhood of \( \psi_0 \). Geyer suggests maximizing ratio wrt \( \psi \) and iterate replacing \( \psi_0 \) with \( \hat{\psi} \).
Importance Sampling — example

Simulation example: \( Y_t | \alpha_t \sim Pois(\exp(0.7 + \alpha_t)) \),

\[ \alpha_t = 0.5 \alpha_{t-1} + \varepsilon_t, \quad \{\varepsilon_t\} \sim \text{IID } N(0, 0.3), \quad n = 200, \quad N = 1000 \]
Simulation example: $Y_t \mid \alpha_t \sim \text{Pois}(\exp(0.7 + \alpha_t))$, $\phi = 0.5$, $\sigma^2 = 0.3$, $n = 200$, $N = 1000$
Importance Sampling (cont)

Choice of importance density $g$:

Durbin and Koopman suggest a linear state-space approximating model

$$Y_t = \mu_t + x_t^T \beta + \alpha_t + Z_t, \quad Z_t \sim N(0, H_t),$$

with

$$\mu_t = y_t - \hat{\alpha}_t - x_t' y_t e^{-\hat{\alpha}_t + x_t' \beta} + 1,$$

$$H_t = e^{-(\hat{\alpha}_t + x_t' \beta)},$$

where the $\hat{\alpha}_t = E_g (\alpha_t \mid y_n)$ are calculated recursively under the approximating model until convergence.

With this choice of approximating model, it turns out that

$$g(\alpha_n \mid y_n; \psi_0) \sim N(\Gamma_n^{-1} \tilde{y}_n, \Gamma_n^{-1}),$$

where

$$\tilde{y}_n = y_n - e^{x_\beta + \hat{\alpha}_n} + e^{x_\beta + \hat{\alpha}_n} \hat{\alpha}_n,$$

$$\Gamma_n = \text{diag}(e^{x_\beta + \hat{\alpha}_n} + (E(\alpha_n \alpha_n'))^{-1}).$$
Importance Sampling (cont)

Components required in the calculation.

- $g(y_n, \alpha_n)$
  - $\tilde{y}'_n \Gamma^{-1}_n \tilde{y}_n$
  - $\det(\Gamma_n)$
- simulate from $N(\Gamma^{-1}_n \tilde{y}_n, \Gamma^{-1}_n)$
  - compute $\Gamma^{-1}_n \tilde{y}_n$
  - simulate from $N(0, \Gamma^{-1}_n)$
Importance Sampling (cont)

Details.

\[
(E(\alpha_n, \alpha'_n))^{-1} = \sigma^{-2} \begin{pmatrix}
1 & -\phi & 0 & \ldots & 0 \\
-\phi & 1 + \phi^2 & -\phi & \ldots & 0 \\
0 & -\phi & 1 + \phi^2 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \ldots & 1 + \phi^2
\end{pmatrix}
\]

\[
\Gamma_n = \text{diag}(e^{\hat{\alpha} + \chi}) + \sigma^{-2} \begin{pmatrix}
1 & -\phi & 0 & \ldots & 0 \\
-\phi & 1 + \phi^2 & -\phi & \ldots & 0 \\
0 & -\phi & 1 + \phi^2 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \ldots & 1 + \phi^2
\end{pmatrix}
\]

This is the covariance function of a 1-dependent sequence, so that \( \Gamma_n = C_n D_n C'_n \), where

\[
C_n = \begin{pmatrix}
1 & 0 & 0 & \ldots & 0 \\
\theta_{1,1} & 1 & 0 & \ldots & 0 \\
0 & \theta_{2,1} & 1 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \ldots & 1
\end{pmatrix}
\]
Importance Sampling (cont)

It follows that

\[
\tilde{y}_n^r \Gamma_n^{-1} \tilde{y}_n = \sum_{t=1}^{n} (\tilde{y}_t - \hat{y}_t)^2 / v_{t-1}
\]

and

\[
\Gamma_n^{-1} \tilde{y}_n = C_n^{-1} D_n^{-1} C_n (\tilde{y}_n - \hat{y}_n)
\]

\[
= C_n^{-1} (D_n^{-1} (\tilde{y}_n - \hat{y}_n))
\]

which can be solved for the vector \( \Gamma_n^{-1} \tilde{y}_n \) via the recursion

\[
C_n^r \Gamma_n^{-1} \tilde{y}_n = D_n^{-1} (\tilde{y}_n - \hat{y}_n).
\]

All of these calculations can be carried out quickly using the innovations algorithm.

To simulate from \( N(0, \Gamma_n^{-1}) \) note that

\[
U_n = C_n^{-1} D_n^{-1} Z_n,
\]

where \( Z_n \sim N(0,1) \), has covariance matrix \( \Gamma_n^{-1} \).
Importance Sampling — example

Simulation example: $\beta = .7$, $\phi = .5$, $\sigma^2 = .3$, $n = 200$, $N = 1000$, 50 realizations plotted
Estimation Methods — Approximation to the likelihood

Joint density function:

\[
p(y_n, \alpha_n) \propto \frac{\det(G)^{1/2}}{\prod_{t=1}^{n} y_t !} \exp\{-y_n^T(\alpha_n + X\beta) - e^{1^T(\alpha_n + X\beta)} - \alpha_n^T G_n \alpha_n / 2\},
\]

where \(G_n^{-1} = E(\alpha_n^T \alpha_n)\).

Conditional density function:

\[
p(\alpha_n \mid y_n) \propto \exp\{-y_n^T \alpha_n - e^{1^T(\alpha_n + X\beta)} - \alpha_n^T G_n \alpha_n / 2\},
\]

which, by expanding the term, \(e^{1^T(\alpha_n + X\beta)}\) in a neighborhood of \(\alpha_n^*\), and ignoring third-order + terms yields the approximation

\[
p_{a}(\alpha_n \mid y_n) \propto \exp\{-y_n^T(\alpha_n + X\beta) - e^{1^T(\alpha_n^* + X\beta)} + (\alpha_n - \alpha_n^*)^T e^{\alpha_n^* + X\beta} + \frac{1}{2} (\alpha_n - \alpha_n^*)^T \text{diag}(e^{\alpha_n^* + X\beta})(\alpha_n - \alpha_n^*) - \alpha_n^T G_n \alpha_n / 2\}.
\]
After simplification, we find

\[
p_a(\alpha_n \mid y_n) \propto \exp\{-y_n^T (\alpha_n + X^T \beta) - e_1^T (\alpha_n^* + X^T \beta) + (\alpha_n - \alpha_n^*)^T e_1^{\alpha_n^* + X^T \beta}
\]

\[
+ \frac{1}{2} (\alpha_n - \alpha_n^*)^T \text{diag}(e_1^{\alpha_n^* + X^T \beta})(\alpha_n - \alpha_n^*) - \alpha_n^T G_n \alpha_n / 2 \}
\]

\[
\sim N(\Gamma_n^{-1} \tilde{y}_n, \Gamma_n^{-1})
\]

Approximate likelihood:

\[
p_a(y_n; \psi) = \frac{p(y_n, \alpha_n)}{p_a(\alpha_n \mid y_n)} \propto \frac{\det(G_n)^{1/2}}{\det(\Gamma_n)^{1/2}} \exp\{y_n^T X^T \beta + 0.5 \tilde{y}_n^T \Gamma_n^{-1} \tilde{y}_n \},
\]

\[
\tilde{y}_n = y_n - \exp\{X^T \beta\} \exp\{\alpha_n^*\} + \exp\{\alpha_n^*\} \exp\{X^T \beta\} \alpha_n^*
\]

(component-wise multiplication for vectors)

Note: We actually expand the joint density for \( Y_n \) and \( \alpha_n \) in a neighborhood of \( \alpha^* \).
Implementation:

1. Let $\alpha^* = \alpha^*(\psi)$ be the converged value of $\alpha^{(j)}(\psi)$, where
   \[
   \alpha^{(j+1)}(\psi) = \Gamma_n^{-1} \tilde{y}_n(\psi)
   \]

2. Maximize $p_a(y_n; \psi)$ with respect to $\psi$. 
Model: $Y_t \mid \alpha_t \sim Pois(\exp(.7 + \alpha_t))$, $\alpha_t = .5 \alpha_{t-1} + \varepsilon_t$, $\{\varepsilon_t\} \sim$IID $N(0, .3)$, $n = 200$

Estimation methods:

- Importance sampling ($N=1000$, $\psi_0$ updated a maximum of 10 times)

<table>
<thead>
<tr>
<th>beta</th>
<th>phi</th>
<th>sigma2</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>0.6982</td>
<td>0.4718</td>
</tr>
<tr>
<td>std</td>
<td>0.1059</td>
<td>0.1476</td>
</tr>
</tbody>
</table>

- Approximation to likelihood

<table>
<thead>
<tr>
<th>beta</th>
<th>phi</th>
<th>sigma2</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>0.7036</td>
<td>0.4579</td>
</tr>
<tr>
<td>std</td>
<td>0.0951</td>
<td>0.1365</td>
</tr>
</tbody>
</table>
Model: $Y_t \mid \alpha_t \sim Pois(\exp(0.7 + \alpha_t))$, $\alpha_t = 0.5 \alpha_{t-1} + \epsilon_t$, $\{\epsilon_t\} \sim IID N(0, .3)$, $n = 200$

Approx likelihood

Importance Sampling
Model for \( \{ \alpha_t \} \):
\[
\alpha_t = \phi \alpha_{t-1} + \epsilon_t, \quad \{ \epsilon_t \} \sim \text{IID } \mathcal{N}(0, \sigma^2). 
\]

- Importance sampling (\( \psi_0 \) updated 5 times for each \( N=100, 500, 1000, \))
- Simulation based on 1000 replications and the fitted AL model.

<table>
<thead>
<tr>
<th></th>
<th>Import Sampling</th>
<th>Approx Like</th>
<th>GLM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \hat{\beta}_{IS} )</td>
<td>Simulation</td>
<td>Simulation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Intercept</td>
<td>0.203</td>
<td>0.223</td>
<td>0.381</td>
</tr>
<tr>
<td>Trend((\times 10^{-3}))</td>
<td>-2.675</td>
<td>-2.778</td>
<td>3.979</td>
</tr>
<tr>
<td>cos(2(\pi t/12))</td>
<td>0.110</td>
<td>0.103</td>
<td>0.124</td>
</tr>
<tr>
<td>sin(2(\pi t/12))</td>
<td>-0.456</td>
<td>-0.456</td>
<td>0.151</td>
</tr>
<tr>
<td>cos(2(\pi t/6))</td>
<td>0.399</td>
<td>0.401</td>
<td>0.123</td>
</tr>
<tr>
<td>sin(2(\pi t/6))</td>
<td>0.015</td>
<td>0.024</td>
<td>0.118</td>
</tr>
<tr>
<td>(\phi)</td>
<td>0.865</td>
<td>0.777</td>
<td>0.198</td>
</tr>
<tr>
<td>(\sigma^2)</td>
<td>0.088</td>
<td>0.100</td>
<td>0.068</td>
</tr>
</tbody>
</table>
Application to Model Fitting for the Polio Data (cont)

Approx Likelihood

Importance Sampling
Polio Data: observed and conditional mean (approx like)
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 $\cos(2\pi t/365), \sin(2\pi t/365)$
- 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.
## Results for Asthma Data—(IS & AL)

<table>
<thead>
<tr>
<th>Term</th>
<th>IS</th>
<th>AL</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.590</td>
<td>0.591</td>
<td>0.593</td>
<td>0.0658</td>
</tr>
<tr>
<td>Sunday effect</td>
<td>0.138</td>
<td>0.138</td>
<td>0.139</td>
<td>0.0531</td>
</tr>
<tr>
<td>Monday effect</td>
<td>0.229</td>
<td>0.231</td>
<td>0.230</td>
<td>0.0495</td>
</tr>
<tr>
<td>cos(2\pi t/365)</td>
<td>-0.218</td>
<td>-0.218</td>
<td>-0.217</td>
<td>0.0415</td>
</tr>
<tr>
<td>sin(2\pi t/365)</td>
<td>0.200</td>
<td>0.179</td>
<td>0.181</td>
<td>0.0437</td>
</tr>
<tr>
<td>Term 1, 1990</td>
<td>0.188</td>
<td>0.198</td>
<td>0.194</td>
<td>0.0638</td>
</tr>
<tr>
<td>Term 2, 1990</td>
<td>0.183</td>
<td>0.130</td>
<td>0.129</td>
<td>0.0664</td>
</tr>
<tr>
<td>Term 1, 1991</td>
<td>0.080</td>
<td>0.075</td>
<td>0.070</td>
<td>0.0733</td>
</tr>
<tr>
<td>Term 2, 1991</td>
<td>0.177</td>
<td>0.164</td>
<td>0.157</td>
<td>0.0665</td>
</tr>
<tr>
<td>Term 1, 1992</td>
<td>0.223</td>
<td>0.221</td>
<td>0.214</td>
<td>0.0667</td>
</tr>
<tr>
<td>Term 2, 1992</td>
<td>0.243</td>
<td>0.239</td>
<td>0.237</td>
<td>0.0620</td>
</tr>
<tr>
<td>Term 1, 1993</td>
<td>0.379</td>
<td>0.397</td>
<td>0.394</td>
<td>0.0625</td>
</tr>
<tr>
<td>Term 2, 1993</td>
<td>0.127</td>
<td>0.111</td>
<td>0.108</td>
<td>0.0682</td>
</tr>
<tr>
<td>Humidity H_t/20</td>
<td>0.009</td>
<td>0.010</td>
<td>0.007</td>
<td>0.0032</td>
</tr>
<tr>
<td>NO(_2) max</td>
<td>-0.125</td>
<td>-0.107</td>
<td>-0.108</td>
<td>0.0347</td>
</tr>
<tr>
<td>AR(1), (\phi)</td>
<td>0.385</td>
<td>0.788</td>
<td>0.468</td>
<td>0.3790</td>
</tr>
<tr>
<td>(\sigma^2)</td>
<td>0.053</td>
<td>0.010</td>
<td>0.018</td>
<td>0.0153</td>
</tr>
</tbody>
</table>
Asthma Data: observed and conditional mean

1990

Counts

Day of Year

1991

Counts

Day of Year

1992

Counts

Day of Year

1993

Counts

Day of Year
Summary Remarks

1. Importance sampling offers a nice clean method for estimation in parameter driven models.

2. The innovations algorithm allows for quick implementation of importance sampling. Extends easily to higher-order AR structure.

3. Relative likelihood approach is a one-sample based procedure.

4. Approximation to the likelihood is a non-simulation based procedure which may have great potential especially with large sample sizes and/or large number of explanatory variables.