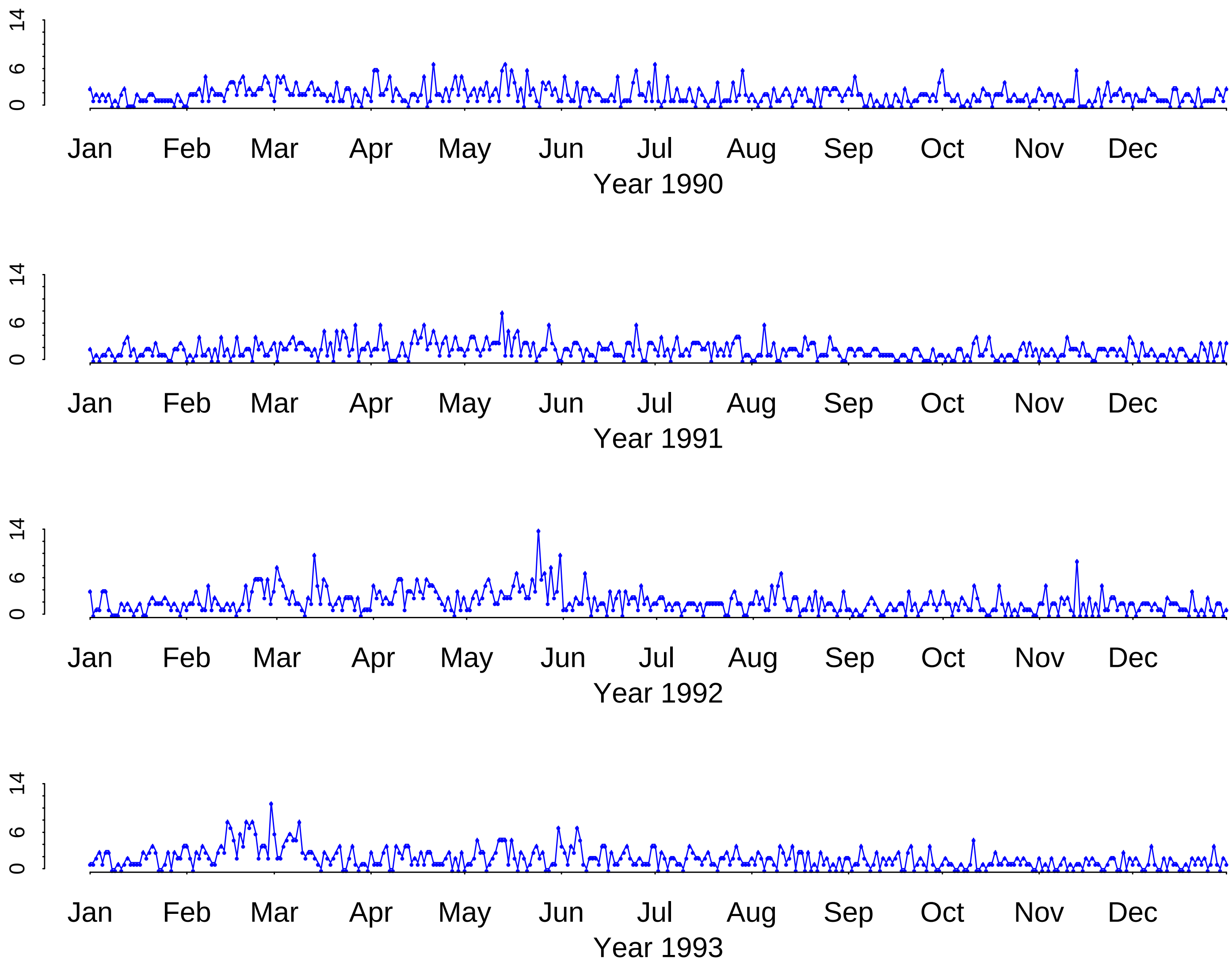
Estimation for State Space Models: an Approximate Likelihood Approach

Richard A. Davis and Gabriel Rodriguez-Yam
Colorado State University

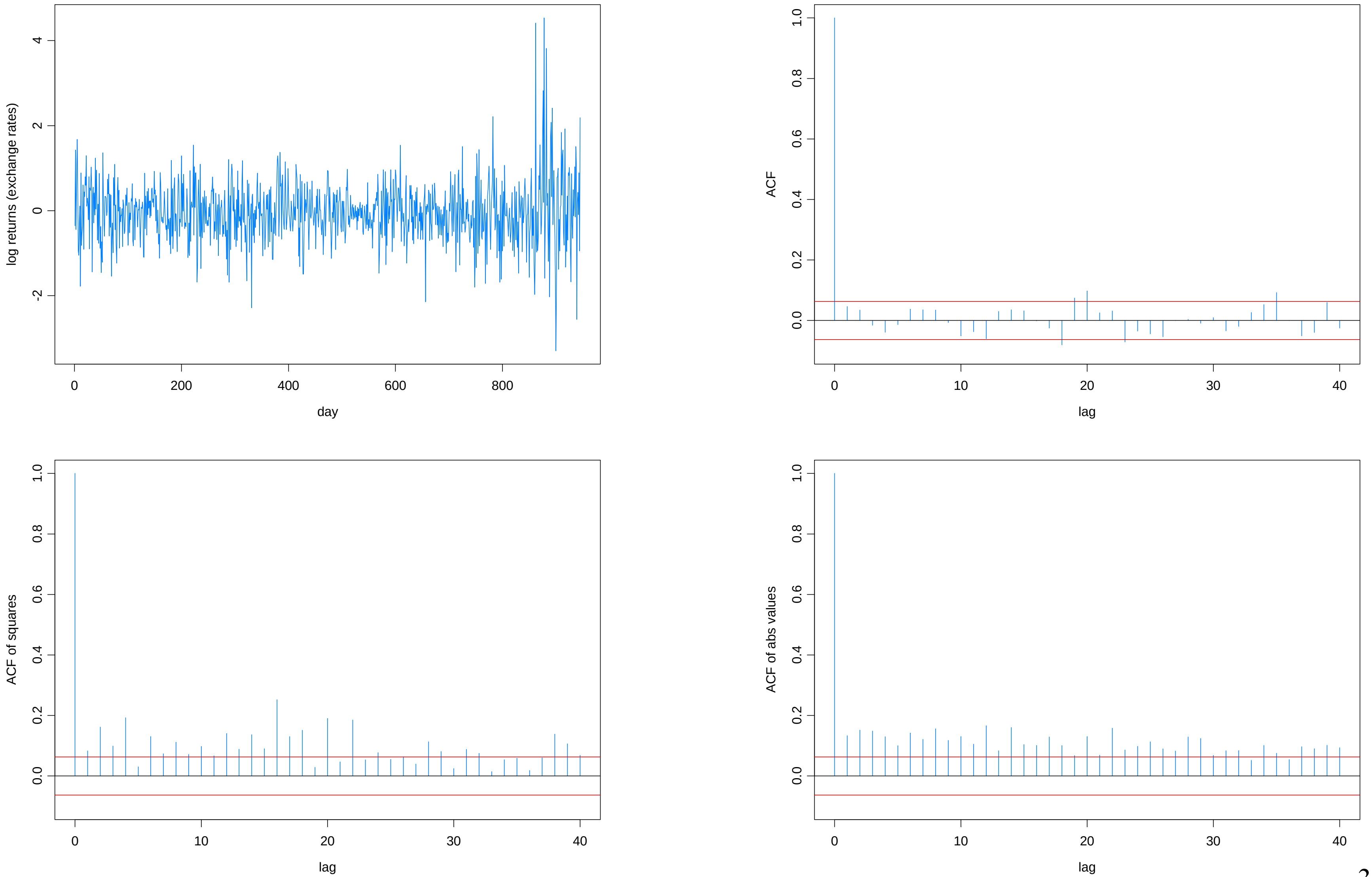
(http://www.stat.colostate.edu/~rdavis/lectures)

Joint work with:
William Dunsmuir, University of New South Wales
Ying Wang, Dept of Public Health, W. Virginia

Example: Daily Asthma Presentations (1990:1993)



Example: Pound-Dollar Exchange Rates
(Oct 1, 1981 - June28, 1985; Koopman website)



- Motivating Examples
  - Time series of counts
  - Stochastic volatility
- Generalized state-space models
  - Observation driven
  - Parameter driven
- Model setup and estimation
  - Generalized linear models (GLM)
  - Estimating equations (Zeger)
  - MCEM (Chan and Ledolter)
  - Importance sampling
    - Durbin and Koopman
  - Approximation to the likelihood (Davis, Dunsmuir, and Wang)
- Application
  - Time series of counts
  - Stochastic volatility

Generalized State-Space Models

Observations: $\mathrm{y}^{(t)} = (y_1, \ldots, y_t)$

States: $\alpha^{(t)} = (\alpha_1, \ldots, \alpha_t)$

Observation equation:

$$p(y_t \mid \alpha_t) := p(y_t \mid \alpha_t, \alpha^{(t-1)}, \mathrm{y}^{(t-1)})$$

State equation:

-observation driven

$$p(\alpha_{t+1} \mid \mathrm{y}^{(t)}) := p(\alpha_{t+1} \mid \alpha_t, \alpha^{(t-1)}, \mathrm{y}^{(t)})$$

-parameter driven

$$p(\alpha_{t+1} \mid \alpha_t) := p(\alpha_{t+1} \mid \alpha_t, \alpha^{(t-1)}, \mathrm{y}^{(t)})$$

Examples of observation driven models

Poisson model for time series of counts

Observation equation:

$$p(y_t \mid \alpha_t) = \frac{e^{\alpha_t y_t - e^{-\alpha_t}}}{y_t!}, \quad y_t = 0, 1, \ldots,$$

State equation:

$$\alpha_{t+1} = \mu + \theta(Y_t - \exp\{\alpha_t\})/\exp\{\alpha_t/2\}$$

where the equation is defined recursively as a function of the past of the y_t's.

Remarks:

- This is an example of a GLARMA(0,1) (see Davis, Dunsmuir, Streett (2003) for more details).
- Estimation is relatively straightforward (can calculate the likelihood in closed form).
- Stability behavior, such as stationarity and ergodicty, is difficult to derive.

Examples of observation driven models-cont

An observation driven model for financial data:

Model (GARCH(p,q)):

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

$$\sigma_t^2 = \alpha_0 + \alpha_1 Y_{t-1}^2 + \cdots + \alpha_p Y_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_q \sigma_{t-q}^2$$

Special case (ARCH(1)=GARCH(1,0)): The resulting observation and state transition density/equations are

$$p(y_t | \sigma_t) = \text{n}(y_t; 0, \sigma_t^2)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 Y_{t-1}^2$$

Properties:

- Martingale difference sequence.
- Stationary for $\alpha_1 \in [0, 2e^E)$, E-Euler's constant.
- Strongly mixing at a geometric rate.
- For general ARCH (GARCH), properties are difficult to establish.

Examples of parameter driven models

<u>Poisson model for time series of counts</u>

Observation equation:

$$p(y_t \mid \alpha_t) = \frac{e^{\alpha_t y_t} e^{-e^{\alpha_t}}}{y_t!}, \quad y_t = 0, 1, \ldots,$$

State equation: State variables follow a regression model with Gaussian AR(1) noise

$$\alpha_t = \beta^T x_t + W_t , \quad W_t = \phi W_{t-1} + Z_t , \quad \{Z_t\} \sim \mathrm{WN}(0,\sigma^2)$$

The resulting transition density of the state variables is

$$p(\alpha_{t+1} \mid \alpha_t) = n(\alpha_{t+1} ; \beta^T x_{t+1} + \phi\,(\alpha_t - \beta^T x_t), \sigma^2)$$

<u>Remark:</u> The case $\sigma^2 = 0$ corresponds to a log-linear model with Poisson noise.

Examples of parameter driven models-cont

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

Model:

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

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

where $\alpha_t = 2 \log \sigma_t$.

The resulting observation and state transition densities are

$$p(y_t | \alpha_t) = \text{n}(y_t; 0, \exp(2\alpha_t))$$

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

Properties:

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

Exponential Family Setup for Parameter-Driven Model

Time series data: $Y_1, \ldots, Y_n$

Regression (explanatory) variable: x_t

Observation equation:

$$p(y_t \mid \alpha_t) = \exp\{(\alpha_t + \beta^T x_t)\, y_t - b(\alpha_t + \beta^T x_t) + c(y_t)\}.$$

State equation: $\{\alpha_t\}$ follows an autoregressive process satisfying the recursions

$$\alpha_t = \gamma + \phi_1 \alpha_{t-1} + \phi_2 \alpha_{t-2} + \ldots + \phi_p \alpha_{t-p} + \varepsilon_t\,,$$

where $\{\varepsilon_t\} \sim$ IID $N(0,\sigma^2)$.

Note: $\alpha_t = 0$ corresponds to standard generalized linear model.

Original primary objective: Inference about β.

Estimation Methods for Parameter Driven Models

- GLM (ignores the presence of the latent process, i.e., $\alpha_t = 0$.)
- Estimating equations (Zeger `88): Let $\hat{\beta}$ be the solution to the equation

$$\frac{\partial\mu}{\partial\beta}\Gamma_n(y_n - \mu) = 0,$$

where $\mu = \exp(X\,\beta)$ and $\Gamma_n = \mathrm{var}(Y_n)$.

- Monte Carlo EM (Chan and Ledolter `95)
- Importance sampling (Durbin & Koopman `01, Kuk `99, Kuk & Chen `97):
- Approximate likelihood (Davis, Dunsmuir & Wang '98)

Estimation Methods Specialized to Poisson Example— GLM estimation

Model: $Y_t \mid \alpha_t, \mathrm{x}_t \sim Pois(\exp(\mathrm{x}_t^{\mathrm{T}}\beta + \alpha_t))$.

GLM log-likelihood:

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

(This *likelihood* ignores presence of the latent process.)

Assumptions on regressors:

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

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

Theory of GLM Estimation in Presence of Latent Process

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

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

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. The x_t can depend on the sample size n.

Application to Model for Polio Data

Assume the $\{\alpha_t\}$ follows a log-normal AR(1), 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	0.17	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

Estimation Methods — Importance Sampling (Durbin and Koopman)

Model:

$$Y_t \mid \alpha_t , x_t \sim Pois(\exp(x_t^T \beta + \alpha_t))$$

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

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)

$$\frac{L(\psi)}{L_g(\psi_0)} = \int \frac{p(\mathrm{y}_n \mid \alpha_n)\, p(\alpha_n)}{g(\mathrm{y}_n, \alpha_n; \psi_0)}\, g(\alpha_n \mid \mathrm{y}_n; \psi_0) d\alpha_n$$

$$= E_g\left[\frac{p(\mathrm{y}_n \mid \alpha_n)\, p(\alpha_n)}{g(\mathrm{y}_n, \alpha_n; \psi_0)} \mid \mathrm{y}_n; \psi_0) \right]$$

$$\sim \frac{1}{N} \sum_{j=1}^{N} \frac{p(\mathrm{y}_n \mid \alpha_n^{(j)})\, p(\alpha_n^{(j)})}{g(\mathrm{y}_n, \alpha_n^{(j)}; \psi_0)},$$

where $\{\alpha_n^{(j)}; j = 1, \ldots, N\} \sim \text{iid g}(\alpha_n \mid \mathrm{y}_n; \psi_0).$

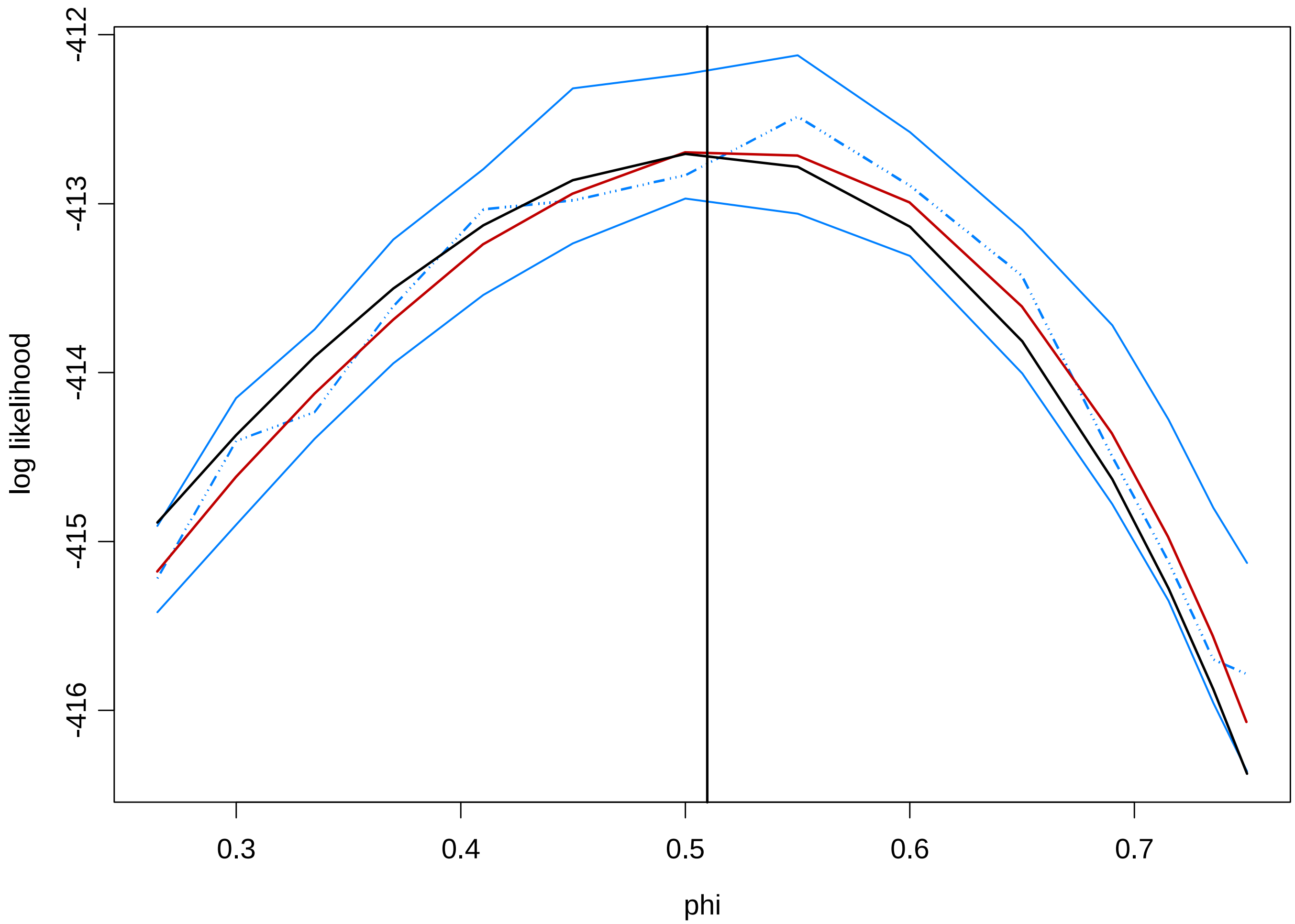
Notes:

- This is a "one-sample" approximation to the relative likelihood. That is, for one realization of the α's, we have, in principle, an approximation to the whole likelihood function.
- Approximation is only good in a neighborhood of ψ_0. Geyer suggests maximizing ratio wrt ψ and iterate replacing ψ_0 with $\hat{\psi}$.

Importance Sampling — example

Simulation example: $Y_t \mid \alpha_t \sim Pois(\exp(.7 + \alpha_t))$,

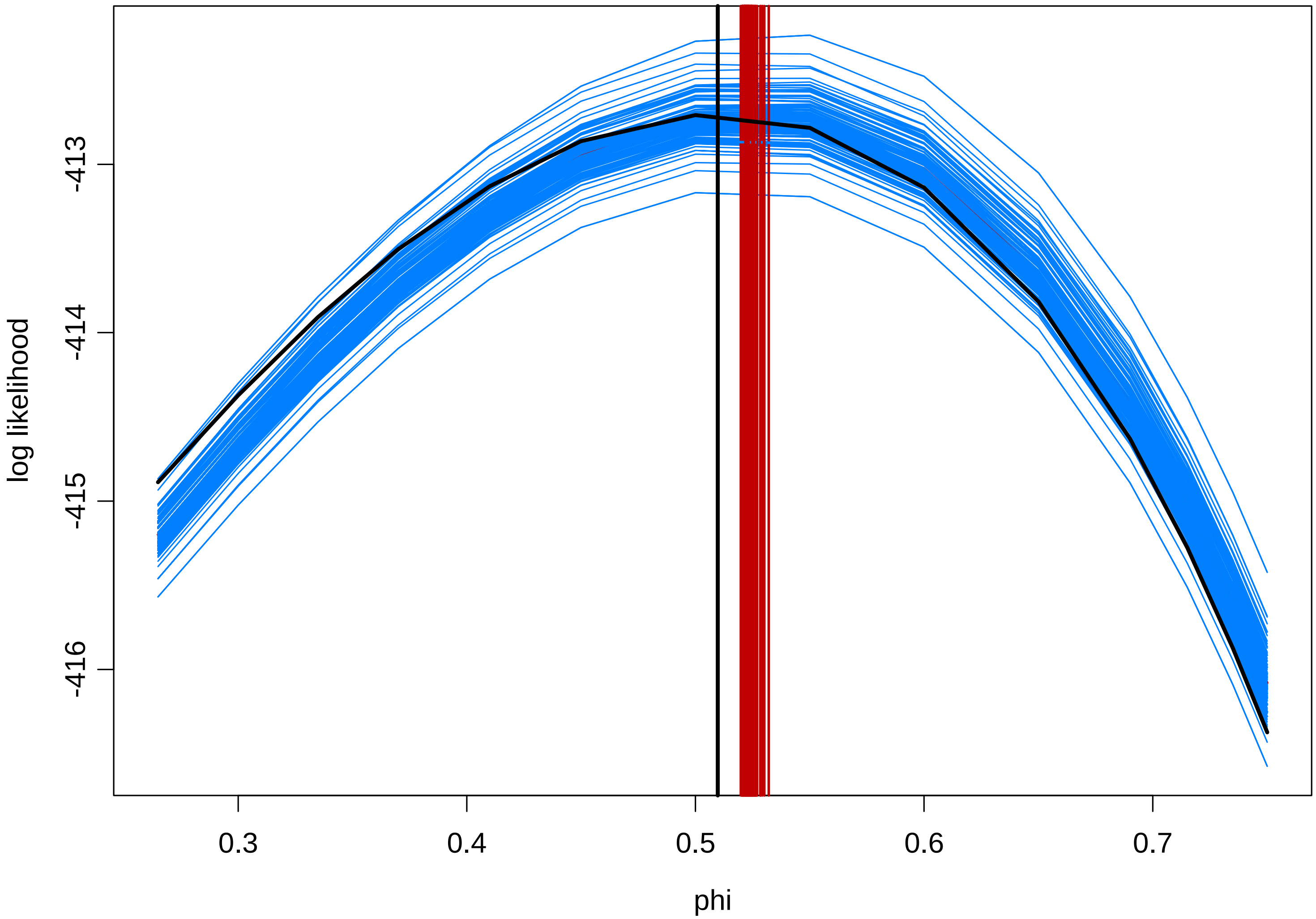
$\alpha_t = .5\,\alpha_{t-1} + \varepsilon_t$, $\{\varepsilon_t\}$~IID N(0, .3), $n = 200$, $N = 1000$



Importance Sampling — example

Simulation example: $Y_t \mid \alpha_t \sim Pois(\exp(.7 + \alpha_t))$,

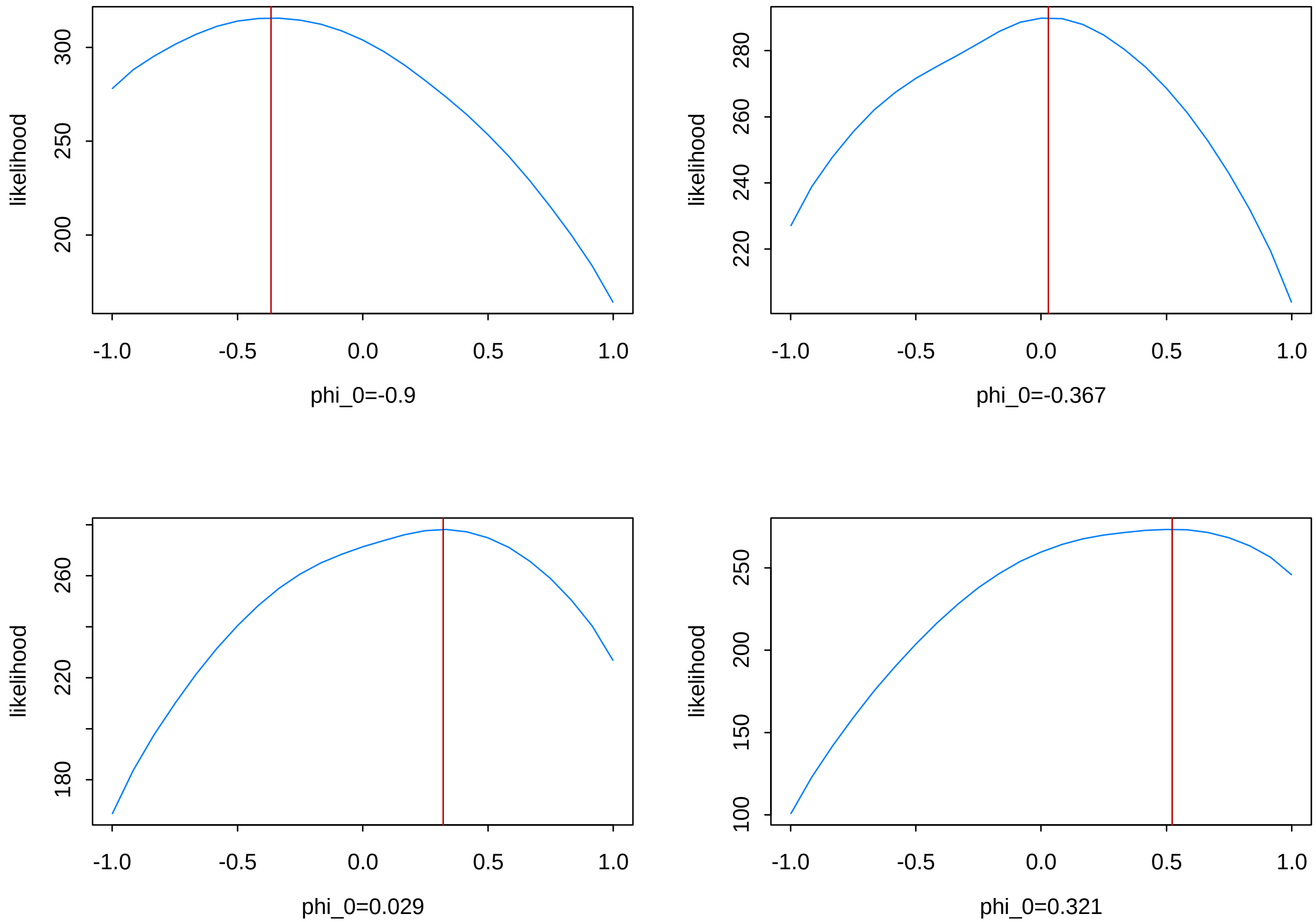
$\alpha_t = .5\,\alpha_{t-1} + \varepsilon_t$, $\{\varepsilon_t\} \sim \text{IID N}(0, .3)$, $n = 200$, $N = 1000$



Importance Sampling — example

Simulation example: $Y_t \mid \alpha_t \sim Pois(\exp(.7 + \alpha_t))$,

$$\alpha_t = .5\,\alpha_{t-1} + \varepsilon_t\,,\quad \{\varepsilon_t\} \sim \text{IID N}(0, .3),\ n = 200,\ N = 1000$$



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

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

$$\Gamma_n = \mathrm{diag}(e^{X\beta + \hat{\alpha}_n}) + (E(\alpha_n \alpha'_n))^{-1}.$$

Importance Sampling (cont)

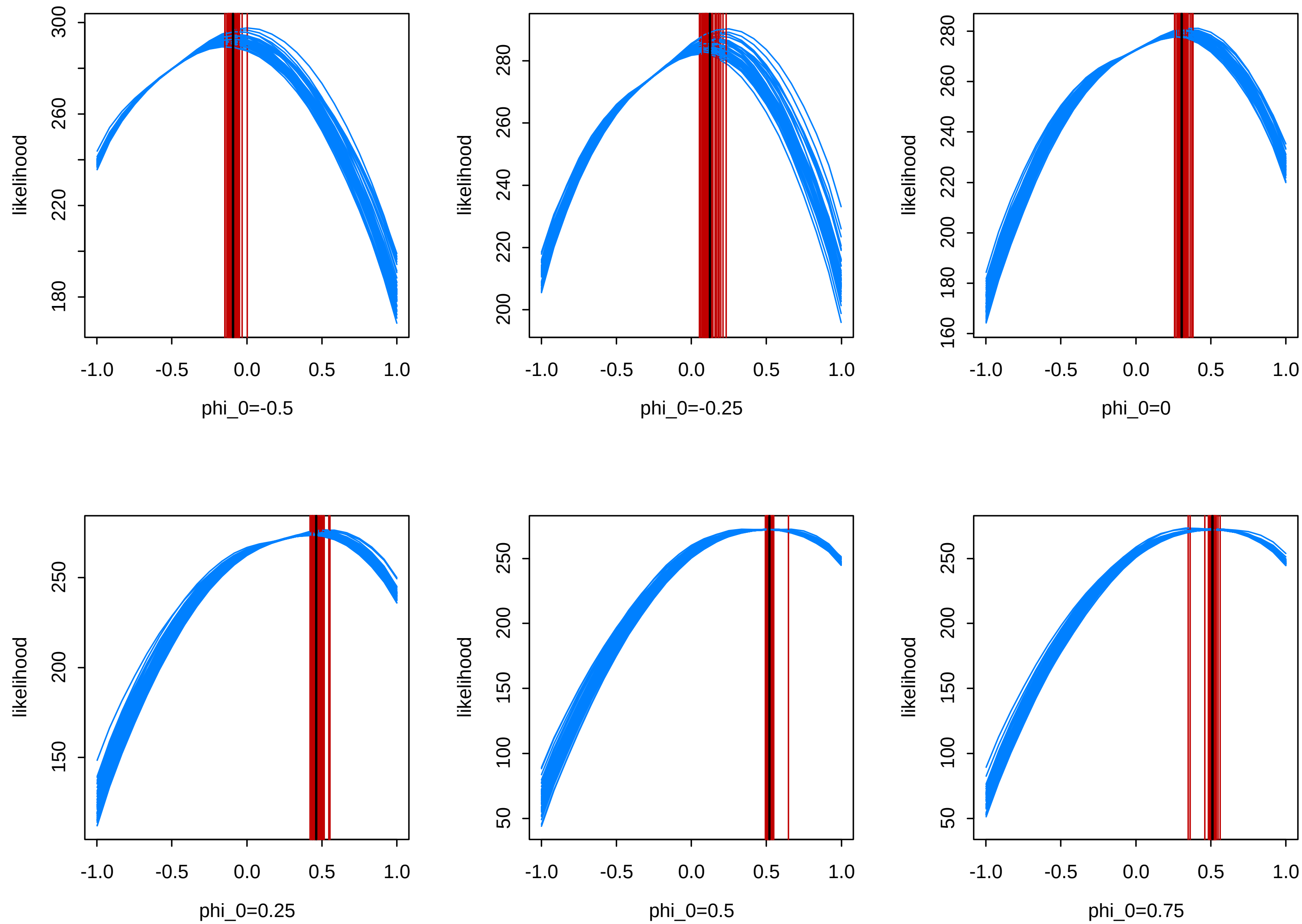
Components required in the calculation.

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

Remark: These quantities can be computed quickly using a version of the innovations algorithm or the Kalman smoothing recursions.

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

General setup:

$$p(\mathrm{y}_n, \alpha_n) \propto p(\mathrm{y}_n \mid \alpha_n) \det(G_n)^{1/2} \exp\{-(\alpha_n - \mu)^T G_n (\alpha_n - \mu)/2\}$$

where

$$G_n^{-1} = E(\alpha_n - \mu)^T (\alpha_n - \mu)$$

Likelihood:

$$L(\psi) = \int p(\mathrm{y}_n \mid \alpha_n) p(\alpha_n) d\alpha_n$$

Consider a Gaussian approximation $p_{\mathrm{a}}(\alpha_{\mathrm{n}} \mid \mathrm{y}_{\mathrm{n}}) = \phi(\alpha_{\mathrm{n}} ; \mu_0 , \Sigma_0)$ to the posterior

$$p(\alpha_{\mathrm{n}} \mid \mathrm{y}_{\mathrm{n}}) \propto p(\alpha_{\mathrm{n}} \mid \mathrm{y}_{\mathrm{n}})\, p(\alpha_{\mathrm{n}})$$

Setting equal the respective posterior modes α_{a}^* and α^* of $p_{\mathrm{a}}(\alpha_{\mathrm{n}} \mid \mathrm{y}_{\mathrm{n}})$ and $p(\alpha_{\mathrm{n}} \mid \mathrm{y}_{\mathrm{n}})$, we have $\mu_0 = \alpha^*$, where α^* is the solution of the equation

$$\frac{\partial}{\partial \alpha_n} \log p(\mathrm{y}_n \mid \alpha_n, \psi) - G_n(\alpha_n - \mu) = 0$$

Estimation Methods — Approximation to the likelihood (cont)

Matching Fisher information matrices:

$$\Sigma_0 = \left(-\frac{\partial^2}{\partial\alpha\partial\alpha^T} \log p(\mathrm{y}_n \mid \alpha_n, \psi) \Big|_{\overset{*}{\alpha}} + G_n \right)^{-1}$$

Approximating posterior:

$$p_a(\alpha_n \mid \mathrm{y}_n, \psi) = \phi\left(\alpha_n; \alpha^*, \left(-\frac{\partial^2}{\partial\alpha\partial\alpha^T} \log p(\mathrm{y}_n \mid \alpha_n, \psi) \Big|_{\overset{*}{\alpha}} + G_n \right)^{-1}\right)$$

Notes:

1. This approximating posterior is identical to the importance sampling density used by Durbin and Koopman.

2. In traditional Bayesian setting, posterior is approximately p_a for large n (see Bernardo and Smith, 1994).

Estimation Methods — Approximation to the likelihood (cont)

Approximate likelihood: Note that

$$p(\alpha_n \mid \mathrm{y}_n) = \frac{p(\mathrm{y}_n \mid \alpha_n)\, p(\alpha_n)}{L(\psi; \mathrm{y}_n)},$$

which by solving for L in the expression,

$$p_{\mathrm{a}}(\alpha_{\mathrm{n}}^{*} \mid \mathrm{y}_{\mathrm{n}}, \psi) = p(\alpha_{\mathrm{n}}^{*} \mid \mathrm{y}_{\mathrm{n}}, \psi),$$

we obtain

$$L_a(\psi; \mathrm{y}_{\mathrm{n}}) = p(\mathrm{y}_{\mathrm{n}} \mid \alpha^{*}, \psi)\, p(\alpha^{*}, \psi) / p_a(\alpha^{*} \mid \mathrm{y}_{\mathrm{n}}, \psi)$$

$$= \frac{|G_n|^{1/2}\; p(\mathrm{y}_{\mathrm{n}} \mid \alpha^{*}, \psi) \exp\{-(\alpha^{*} - \mu)^T G_n (\alpha^{*} - \mu)/2\}}{\det\left(-\frac{\partial^2}{\partial\alpha\partial\alpha^T} \log p(\mathrm{y}_n \mid \alpha_{\mathrm{n}}, \psi)\Big|_{\alpha^{*}} + G_n \right)^{1/2}}$$

Estimation Methods — Approximation to the likelihood (cont)

Case of exponential family:

$$L_a(\psi;\mathrm{y_n}) = \frac{|G_n|^{1/2}}{(K+G_n)^{1/2}} \exp\{\mathrm{y}_n^T \alpha^* - 1^T\{b(\alpha^*) - c(\mathrm{y}_n)\} - (\alpha^* - \mu)^T G_n (\alpha^* - \mu)/2\},$$

where

$$K = \mathrm{diag}\{\frac{\partial^2}{\partial \alpha_t^2} b_t(\alpha_t)\Big|_{\alpha_t^*}\},$$

and α^* is the solution to the equation

$$\mathrm{y}_n - \frac{\partial}{\partial \alpha_n} \mathrm{b}(\alpha_\mathrm{n}) - \mathrm{G_n}(\alpha_\mathrm{n} - \mu) = 0.$$

Using a Taylor expansion, the latter equation can be solved iteratively.

Estimation Methods — Approximation to the likelihood

Implementation:

1. Let $\alpha^* = \alpha^*(\psi)$ be the converged value of $\alpha^{(j)}(\psi)$, where

$$\alpha^{(j+1)}(\psi) = (\ddot{b}^{j} + G_n)^{-1}\tilde{y}_n^{j}(\psi),$$

   and

$$\tilde{y}_n^{j} = y_n - \dot{b}^{j} + \ddot{b}^{j}\alpha^{(j)} + G_n\mu.$$

2. Maximize $p_a(y_n;\psi)$ with respect to ψ.

Simulation Results

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, ψ_0 updated a maximum of 10 times)

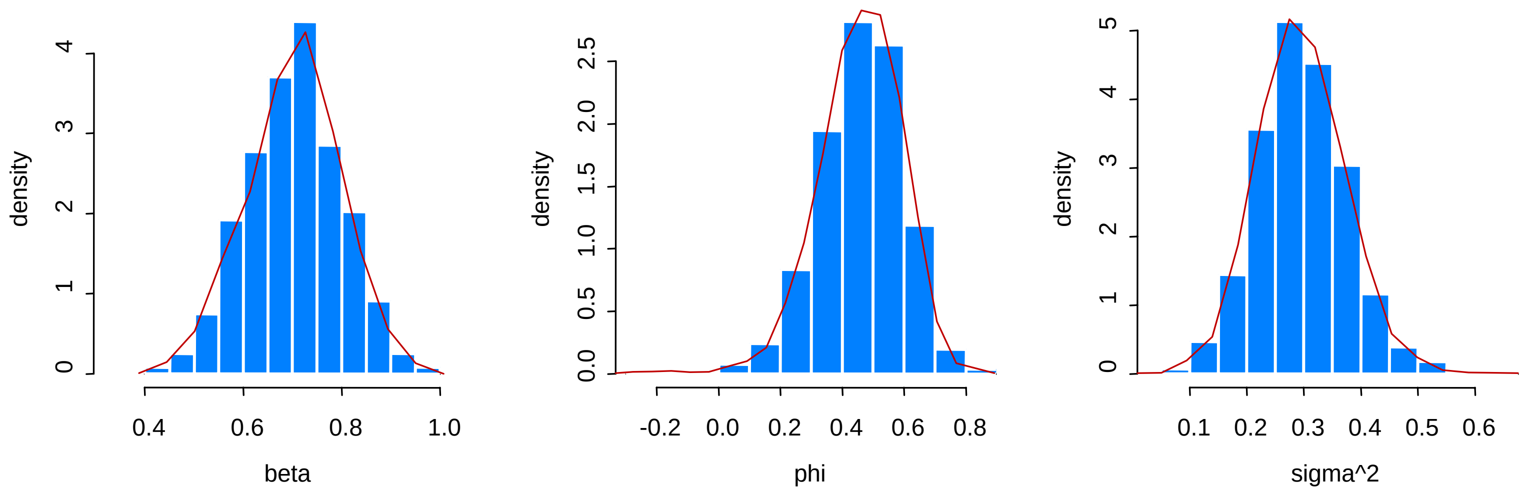
	beta	phi	sigma2
mean	0.6982	0.4718	0.3008
std	**0.1059**	**0.1476**	**0.0899**

- Approximation to likelihood

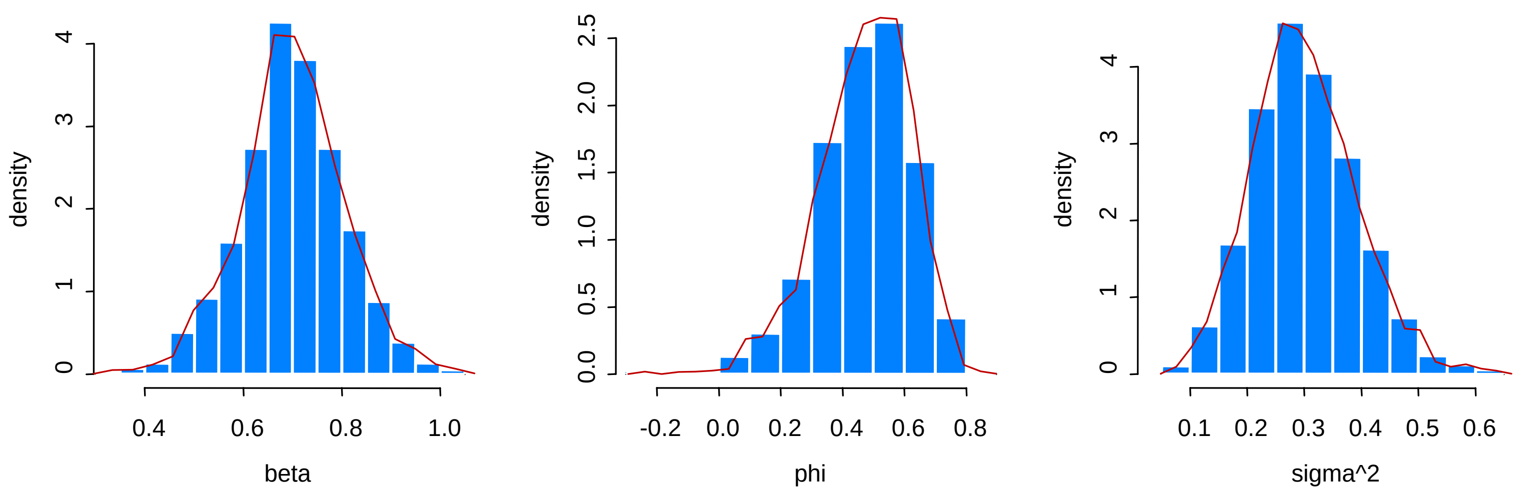
	beta	phi	sigma2
mean	0.7036	0.4579	0.2962
std	**0.0951**	**0.1365**	**0.0784**

Model: $Y_t \mid \alpha_t \sim Pois(\exp(.7 + \alpha_t))$, $\alpha_t = .5\, \alpha_{t-1} + \varepsilon_t$, $\{\varepsilon_t\} \sim \mathrm{IID}\ \mathrm{N}(0, .3)$, $n = 200$

Approx likelihood



Importance Sampling



Application to Model Fitting for the Polio Data

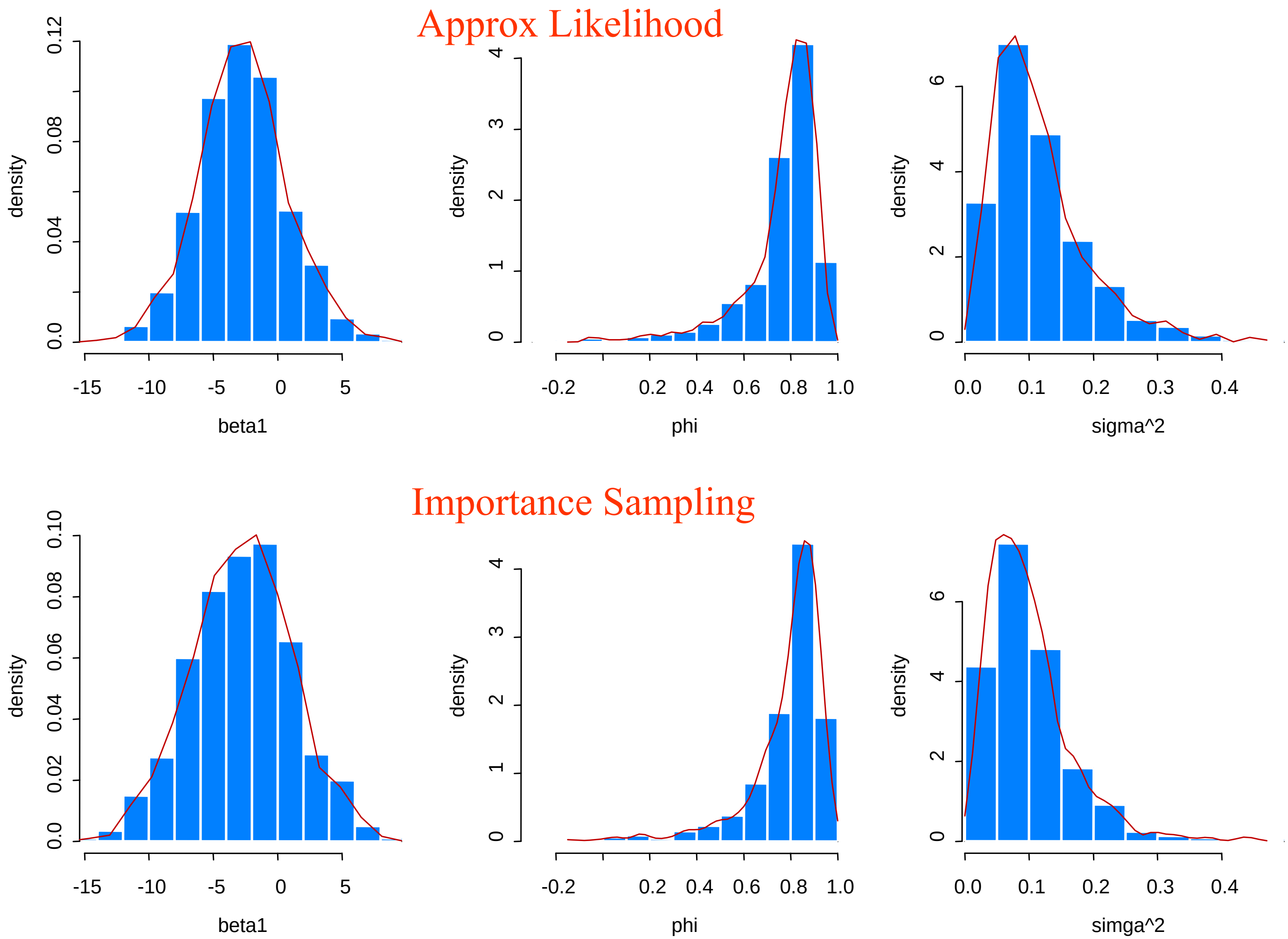
Model for $\{\alpha_t\}$:

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

- Importance sampling (ψ_0 updated 5 times for each N=100, 500, 1000,)
- Simulation based on 1000 replications and the fitted AL model.

	Import Sampling			Approx Like			GLM	
		Simulation			Simulation			
	$\hat{\beta}_{IS}$	Mean	SD	$\hat{\beta}_{AL}$	Mean	SD	$\hat{\beta}_{GLM}$	SD
Intercept	**0.203**	0.223	0.381	**0.202**	0.210	0.343	**.207**	0.078
Trend($\times 10^{-3}$)	**-2.675**	-2.778	3.979	**-2.690**	-2.720	3.415	**-4.18**	1.400
$\cos(2\pi t/12)$	**0.110**	0.103	0.124	**0.113**	0.111	0.123	**-.152**	0.097
$\sin(2\pi t/12)$	**-0.456**	-0.456	0.151	**-0.454**	-0.454	0.143	**-.532**	0.109
$\cos(2\pi t/6)$	**0.399**	0.401	0.123	**0.396**	0.400	0.114	**.169**	0.098
$\sin(2\pi t/6)$	**0.015**	0.024	0.118	**0.016**	0.012	0.110	**-.432**	0.101
ϕ	**0.865**	0.777	0.198	**0.845**	0.764	0.165		
σ^2	**0.088**	0.100	0.068	**0.104**	0.114	0.075		

Application to Model Fitting for the Polio Data (cont)



Simulation Results

Stochastic volatility model:

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

$$\alpha_t = \gamma + \phi\alpha_{t-1} + \varepsilon_t, \quad \{\varepsilon_t\} \sim \text{IID } N(0,\sigma^2), \ \text{where } \alpha_t = 2 \log \sigma_t \text{ ; } n=1000, \ NR=500$$

CV=10

	True	AL	RMSE	IS	RMSE
γ	–.411	–.491	.210	–.490	.216
ϕ	0.950	0.940	.025	0.940	.026
σ	0.484	**0.478**	.065	**0.481**	.073

CV=1

	True	AL	RMSE	IS	RMSE
γ	–.368	–.499	.341	–.485	.324
ϕ	0.950	0.932	.046	0.934	.043
σ	0.260	**0.270**	.068	**0.268**	.068

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(2πt/365), sin(2πt/365)
- seasonal effect modeled by

$$P_{ij}(t) = \frac{1}{B(2.5,5)} \left(\frac{t - T_{ij}}{100} \right)^{2.5} \left(1 - \frac{t - T_{ij}}{100} \right)^{5}$$

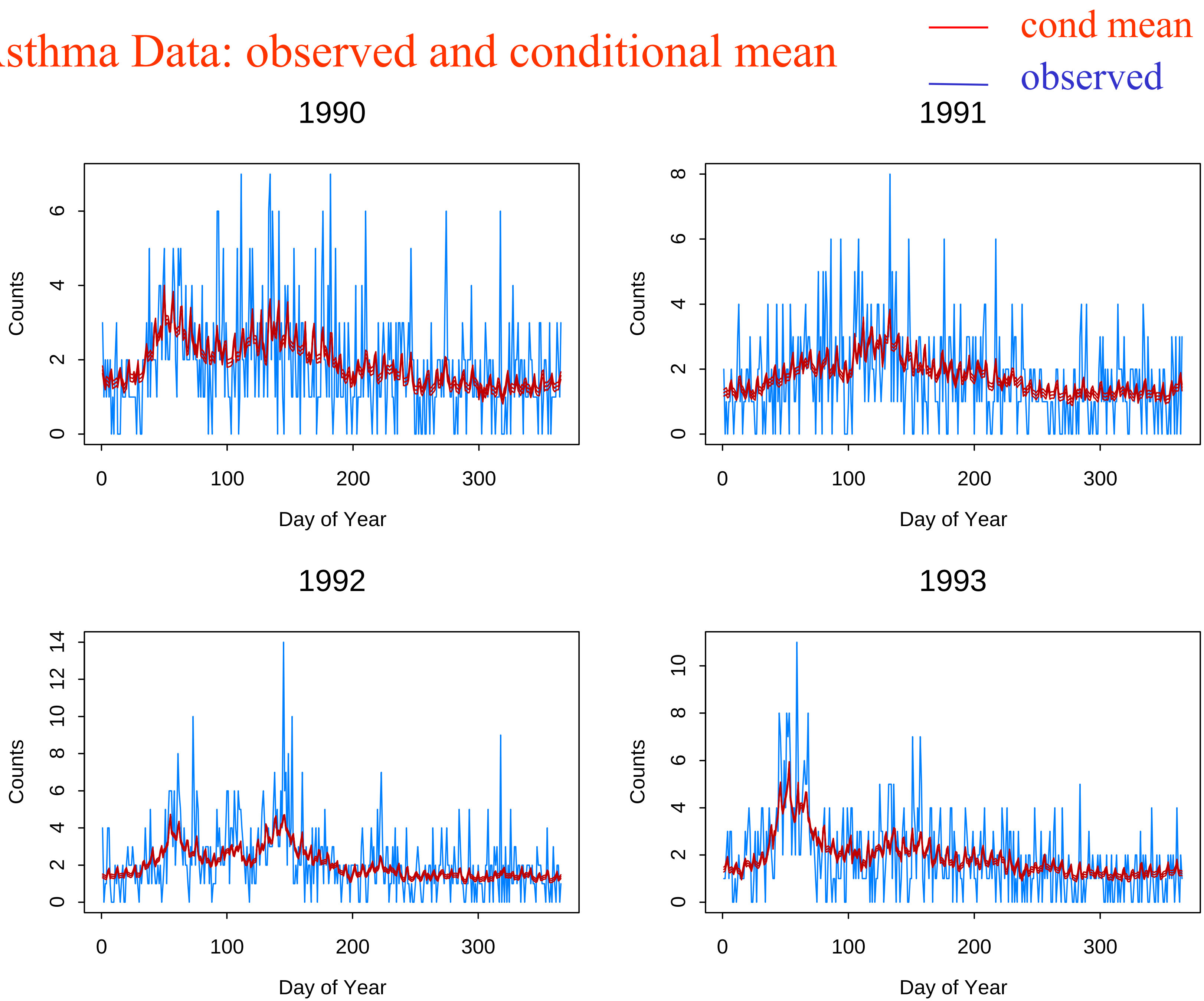
  where $B(2.5,5)$ is the beta function and T_{ij} is the start of the j^{th} school term in year i.

- 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)

Term	IS	AL	Mean	SD
Intercept	0.590	0.591	0.593	.0658
Sunday effect	0.138	0.138	0.139	.0531
Monday effect	0.229	0.231	0.230	.0495
$\cos(2\pi t/365)$	-0.218	-0.218	-0.217	.0415
$\sin(2\pi t/365)$	0.200	0.179	0.181	.0437
Term 1, 1990	0.188	0.198	0.194	.0638
Term 2, 1990	0.183	0.130	0.129	.0664
Term 1, 1991	0.080	0.075	0.070	.0733
Term 2, 1991	0.177	0.164	0.157	.0665
Term 1, 1992	0.223	0.221	0.214	.0667
Term 2, 1992	0.243	0.239	0.237	.0620
Term 1, 1993	0.379	0.397	0.394	.0625
Term 2, 1993	0.127	0.111	0.108	.0682
Humidity $H_t/20$	0.009	0.010	0.007	.0032
NO_2 max	-0.125	-0.107	-0.108	.0347
AR(1), ϕ	0.385	0.788	0.468	.3790
σ^2	0.053	0.010	0.018	.0153

Asthma Data: observed and conditional mean



Is the posterior distribution close to normal?

Compare posterior mean with posterior mode: Can compute the posterior mean using *SIR* (sampling importance-resampling)

Posterior mode: The mode of $p(\alpha_n \mid y_n)$ is α^* found at the last iteration.

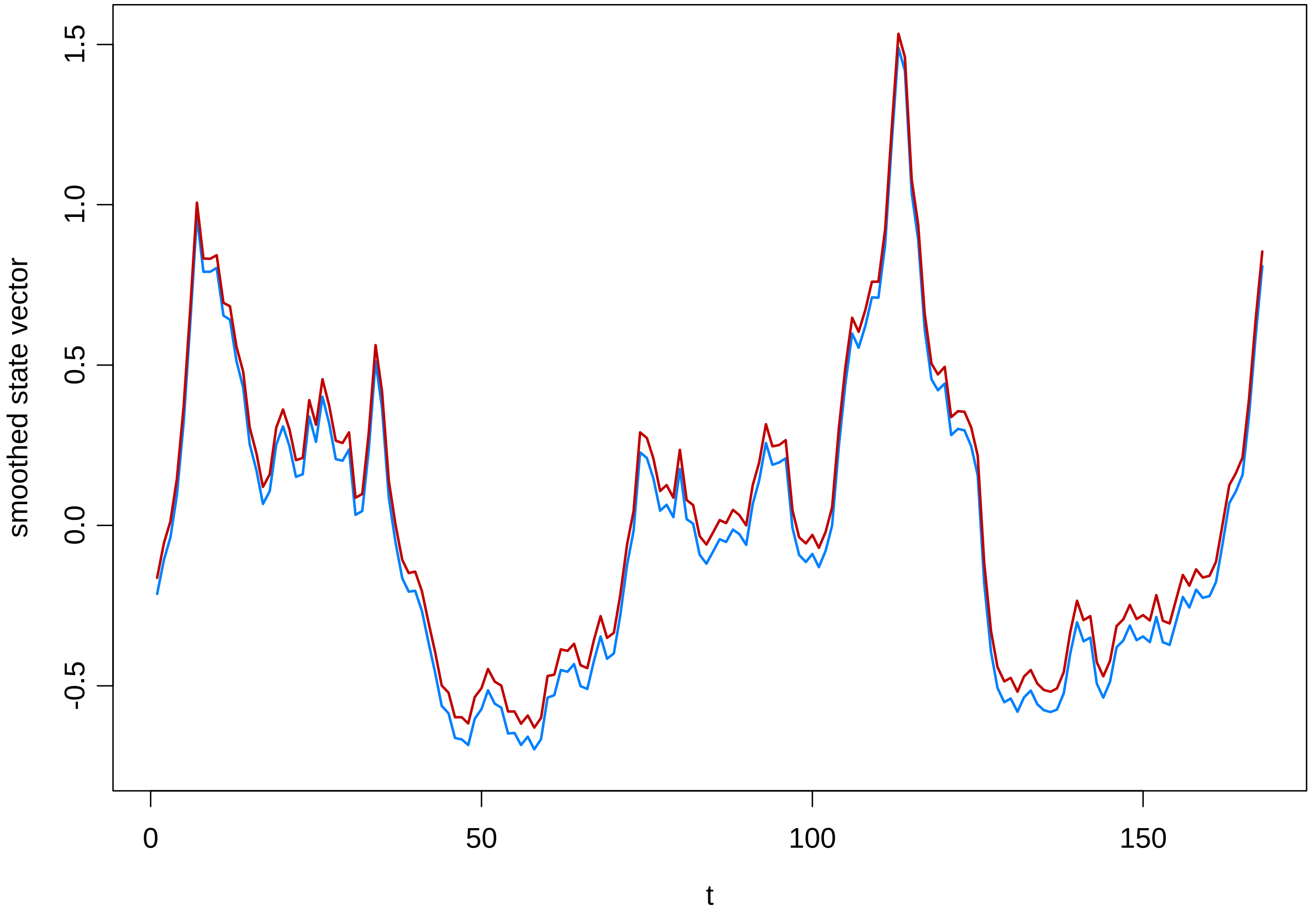
Posterior mean: The mean of $p(\alpha_n \mid y_n)$ can be found using SIR.

Let $\alpha^{(1)}, \alpha^{(2)}, \ldots, \alpha^{(N)}$ be independent draws from the multivariate distr $p_a(\alpha_n \mid y_n)$. For N large, an approximate iid sample from $p(\alpha_n \mid y_n)$ can be obtained by drawing a random sample from $\alpha^{(1)}, \alpha^{(2)}, \ldots, \alpha^{(N)}$ with probabilities

$$p_i = \frac{w_i}{\sum_{i=1}^{N} w_i}, \quad w_i = \frac{p(\alpha^{(i)} \mid y_n)}{p_a(\alpha^{(i)} \mid y_n)} \propto \frac{L(\psi; y_n, \alpha^{(i)})}{p_a(\alpha^{(i)} \mid y_n)}, \quad i = 1, \ldots, N.$$

Posterior mean vs posterior mode?

Polio data: blue = mean, red = mode



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.

5. Approximation likelihood approach is amenable to boostrapping procedures for bias correction.