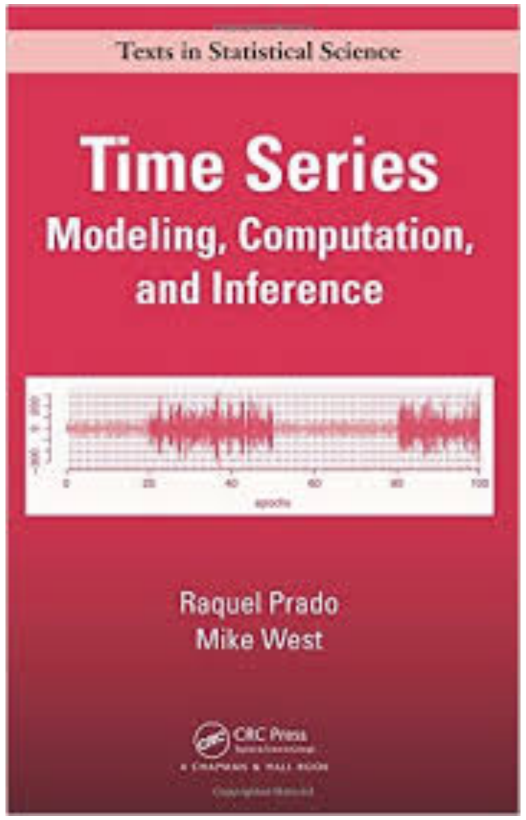
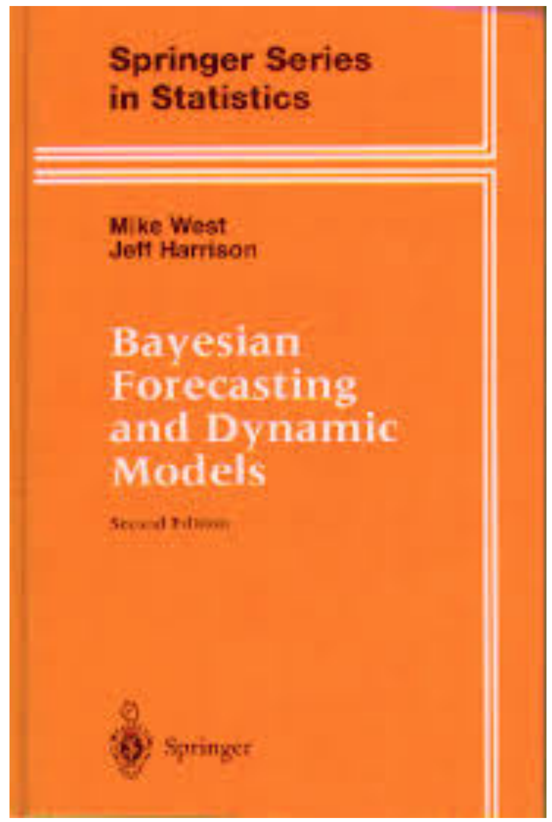
AMS-223: Time Series Analysis, Winter 2017

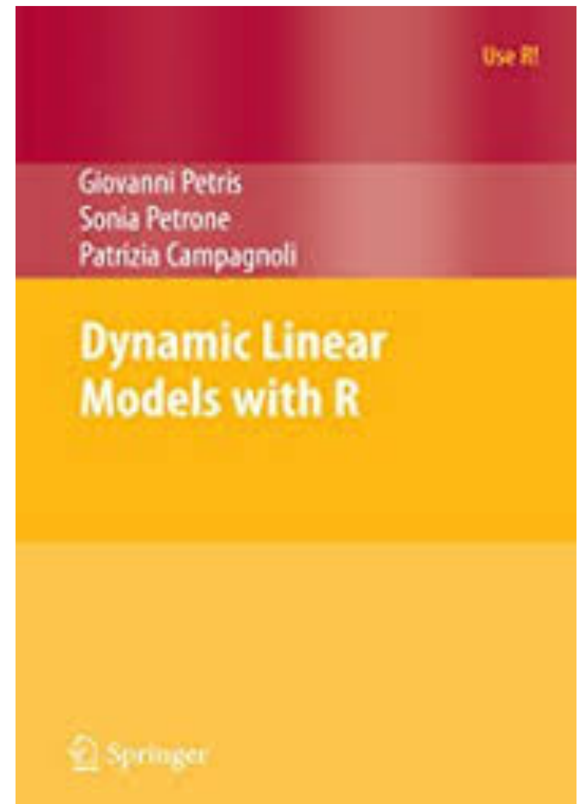
- **Instructor:** Raquel Prado, BE-365C
- **Office Hours:** TBA
- **Website:** https://ams223-winter17-01.courses.soe.ucsc.edu
- **Evaluation:**
  - **Exam 1, 25%, 01/31:** in class + TH (not likely).
  - **Exam 2, 25%, 02/28:** in class + TH (somewhat likely)
  - **Exam 3, 30%, 03/22:** in class (somewhat likely) + TH.
  - **Homework, 20%**: Assignments roughly every other week; will not be graded, subset of the problems will be presented by a subset of the students in class.

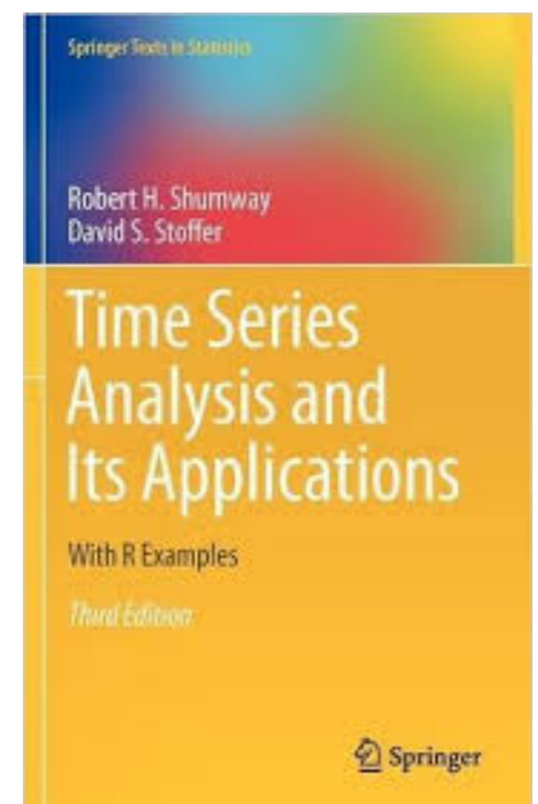
AMS-223: Time Series Analysis, Winter 2017

- **References:**









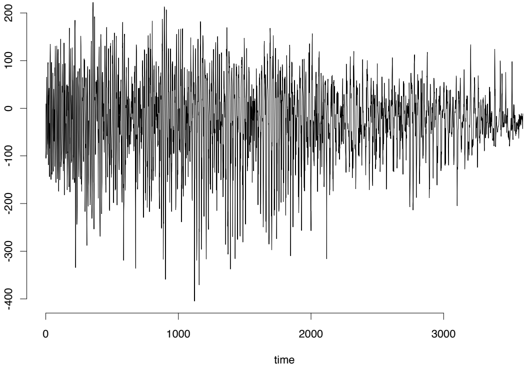
Notation

- **General Case:** $\{\mathbf{y}_t, t \in \mathcal{T}\}$ with $\mathcal{T} = \{t_i, i \in \mathcal{N}\}$
- **Equally-spaced data:** $\mathbf{y}_{1:T}$, i.e., we have observed $\mathbf{y}_1, \ldots, \mathbf{y}_T$.
- Univariate or multivariate processes: y_t or $\mathbf{y}_t = (y_{1,t}, \ldots, y_{k,t})'$.

In this class we will be studying univariate and equally-spaced time series.

Univariate time series analysis

Modeling and inference: Describing the latent structure of a single series



We want to describe the latent components of y_t in terms of their amplitudes, frequencies, etc...

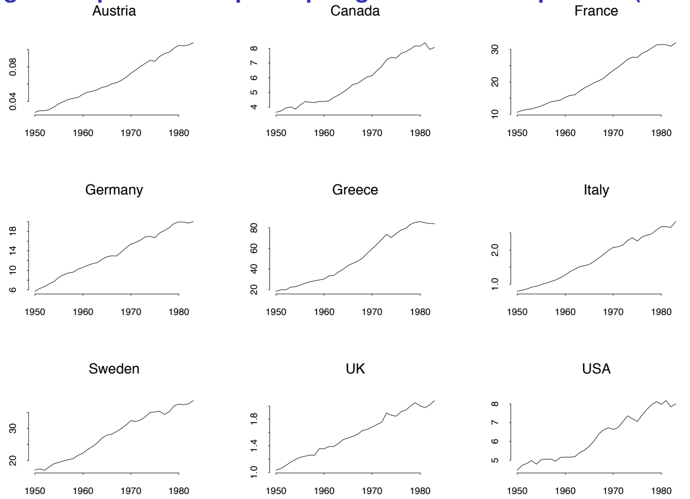
Multivariate time series analysis

What if we have multiple time series or a time series vector, $\mathbf{y}_t = (y_{1,t}, \ldots, y_{k,t})'$, at each time t?

For instance, the electroencephalogram (EEG) time series just displayed is one of several EEGs recorded at different locations over the scalp of a patient. We are interested in discovering common features across multiple EEG signals.

Univariate and multivariate time series analysis

Forecasting. Example: Annual per capita gross domestic product (GDP).



Online monitoring and control

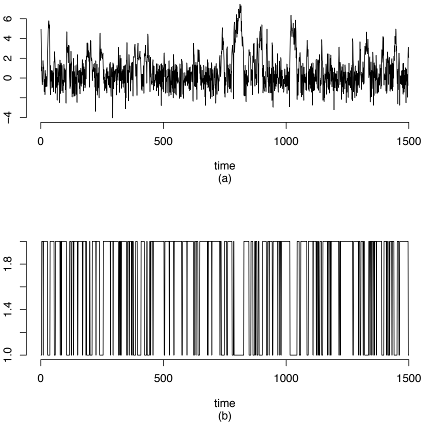
Monitoring a time series to detect possible changes in real time.

Example: TAR(1)

$$y_t = \begin{cases} \phi^{(1)} y_{t-1} & + & \epsilon_t^{(1)}, & \theta + y_{t-d} > 0 & (\mathcal{M}_1) \\ \phi^{(2)} y_{t-1} & + & \epsilon_t^{(2)}, & \theta + y_{t-d} \leq 0 & (\mathcal{M}_2), \end{cases}$$

where $\epsilon_t^{(1)} \sim N(0, v_1)$ and $\epsilon_t^{(2)} \sim N(0, v_2)$.

$y_{1:T}$ and $\delta_{1:T}$



Here $\phi^{(1)} = 0.9$, $\phi^{(2)} = -0.3$, $d = 1$, $\theta = -1.5$, and $v_1 = v_2 = 1$. Also, $\delta_t = 1$ if $y_t \sim \mathcal{M}_1$ and $\delta_t = 2$ if $y_t \sim \mathcal{M}_2$.

Time Series Analysis: Goals

- **Description and analysis:** "You can observe a lot by just watching" (Y. Berra)
- **Prediction:** "It is tough to make predictions, specially about the future" (Y. Berra)
- **Grouping/Clustering**
- Time series models as components of other models
- **Tracking and online learning:** "It gets late early out there" (Y. Berra)

Time Series Analysis: Goals

- **Description and analysis:** "You can observe a lot by just watching" (Y. Berra)

- **Prediction:** "It is tough to make predictions, specially about the future" (Y. Berra)

- **Grouping/Clustering**

- Time series models as components of other models

- **Tracking and online learning:** "It gets late early out there" (Y. Berra)

Models and inference are motivated by real applications

Time Series Analysis: Goals

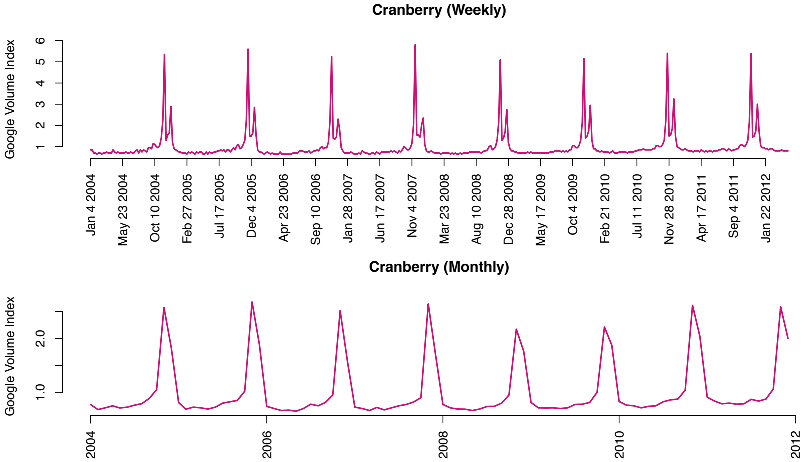
- **Description and analysis:** "You can observe a lot by just watching" (Y. Berra)
- **Prediction:** "It is tough to make predictions, specially about the future" (Y. Berra)
- **Grouping/Clustering**
- Time series models as components of other models
- **Tracking and online learning:** "It gets late early out there" (Y. Berra)

Models and inference are motivated by real applications

"In theory there is no difference between theory and practice, in practice there is"

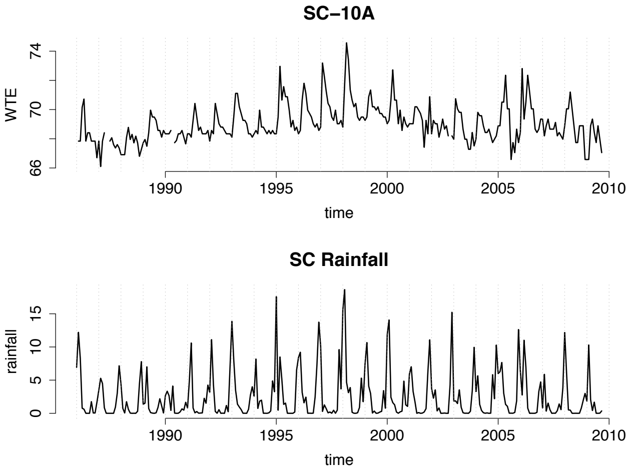
Some Applications

Google Trends



Some Applications

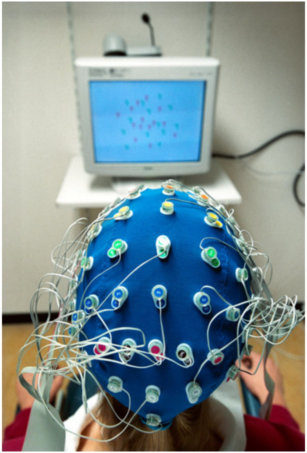
Environmental Data

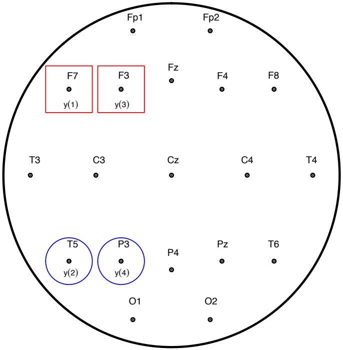


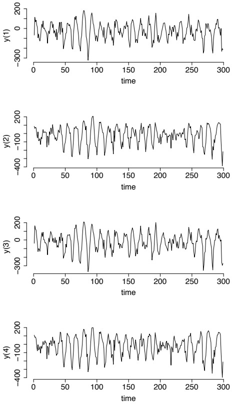
Some Applications

Neuroscience

EEGs measure electrical activity in the brain.







Stationarity

Definition

$\{y_t, t \in \mathcal{T}\}$ is **completely or strongly stationary** if, for any sequence of times $t_1, \ldots, t_n$ and any lag h, the distribution of $(y_{t_1}, \ldots, y_{t_n})'$ is identical to the distribution of $(y_{t_1+h}, \ldots, y_{t_n+h})'$.

Definition

$\{y_t, t \in \mathcal{T}\}$ is **weakly or second order stationary** if for any sequence $t_1, \ldots, t_n$, and any h, the first and second joint moments of $(y_{t_1}, \ldots, y_{t_n})'$ and those of $(y_{t_1+h}, \ldots, y_{t_n+h})'$ exist and are identical.

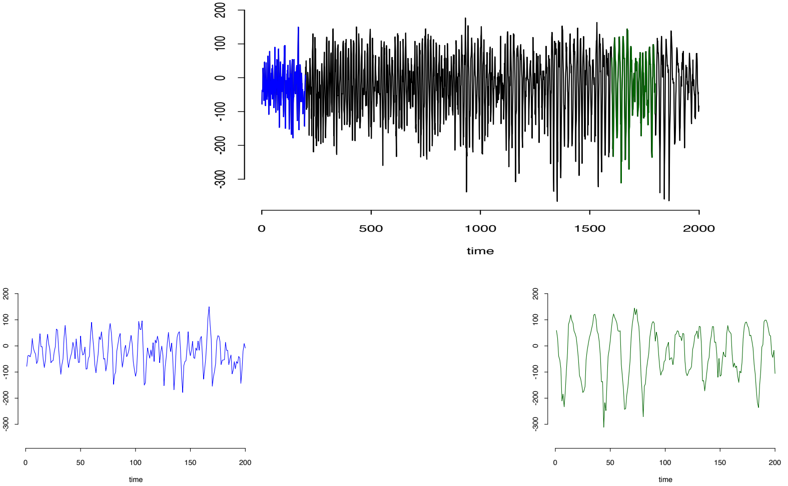
Complete stationarity implies second order stationarity but the converse is not necessarily true.

Stationarity

Second order stationarity

- If $\{y_t\}$ is weakly stationary $\Rightarrow E(y_t) = \mu$, $V(y_t) = v$, and $Cov(y_t, y_s) = \gamma(s - t)$.
- Gaussian time series processes: strong and weak stationarity are equivalent.

Stationarity



The autocorrelation function (ACF)

Definition

The **autocovariance** of $\{y_t\}$ is defined as

$$\gamma(t,s) = Cov(y_t, y_s) = E\{(y_t - \mu_t)(y_s - \mu_s)\}.$$

If $\{y_t\}$ is stationary we can write $\gamma(h) = Cov(y_t, y_{t-h})$.

Definition

The **autocorrelation function** (ACF) is given by

$$\rho(t,s) = \frac{\gamma(t,s)}{\sqrt{\gamma(t,t)\gamma(s,s)}}.$$

For stationary processes we can write $\rho(h) = \gamma(h)/\gamma(0)$.

Cross-covariance and cross-correlation

Definition

The **cross-covariance** is defined as

$$\gamma_{y,z}(t,s) = Cov(y_t, z_s) = E\{(y_t - \mu_{y_t})(z_s - \mu_{z_s})\}.$$

If both processes are stationary we can write $\gamma_{y,z}(h) = Cov(y_t, z_{t-h})$.

Definition

The **cross-correlation** is given by

$$\rho_{y,z}(t,s) = \frac{\gamma_{y,z}(t,s)}{\sqrt{\gamma_{y,y}(t,t)\gamma_{z,z}(s,s)}}.$$

If both processes are stationary $\rho_{y,z}(h) = \gamma_{y,z}(h)/\sqrt{\gamma_y(0)\gamma_z(0)}$.

The sample autocorrelation function

Definition

The **sample autocovariance** (assuming stationarity) is given by

$$\hat{\gamma}(h) = \frac{1}{T} \sum_{t=1}^{T-h} (y_t - \bar{y})(y_{t+h} - \bar{y}),$$

where $\bar{y} = \sum_{t=1}^{T} y_t / T$ is the sample mean.

Definition

The **sample autocorrelation** is given by $\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$.

Sample cross-correlation and auto-correlation

Definition

The **sample cross-covariance** (assuming stationarity) is

$$\hat{\gamma}_{y,z}(h) = \frac{1}{T}\sum_{t=1}^{T-h}(y_{t+h} - \bar{y})(z_t - \bar{z}).$$

Definition

The **sample cross-correlation** is

$$\hat{\rho}_{y,z}(h) = \frac{\hat{\gamma}_{y,z}(h)}{\sqrt{\hat{\gamma}_y(0)\hat{\gamma}_z(0)}}.$$

The ACF

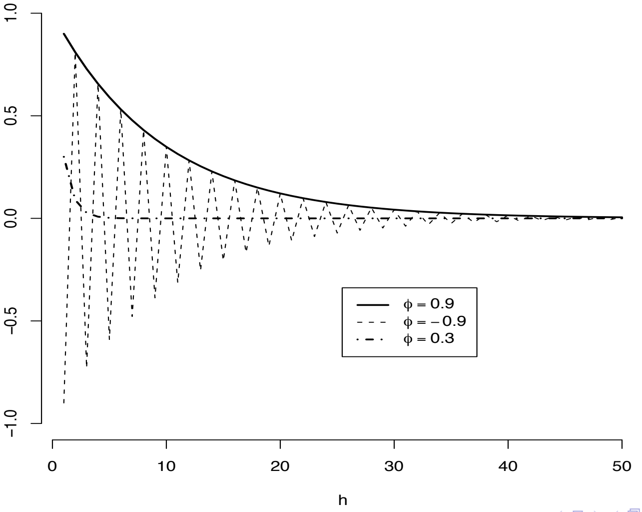
Examples

White Noise. Let $y_t \sim N(0, v)$ for all t, with $Cov(y_t, y_s) = 0$ if $t \neq s$. Then, $\gamma(0) = v$, $\gamma(h) = 0$ for all $h \neq 0$, and so, $\rho(0) = 1$ and $\rho(h) = 0$ for all $h \neq 0$.

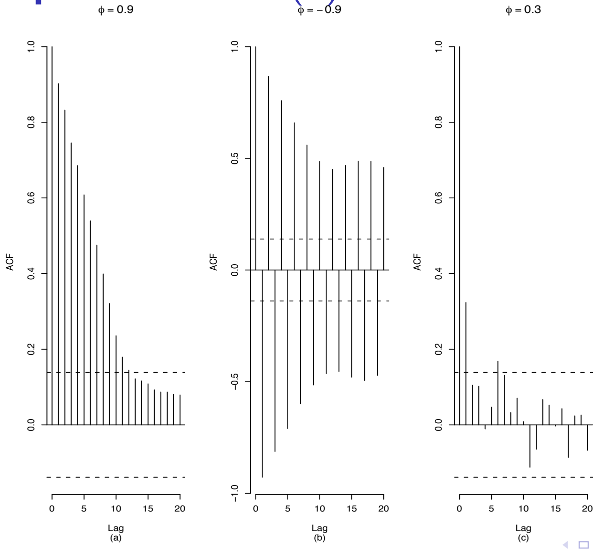
AR(1)**.** Let $y_t = \phi y_{t-1} + \epsilon_t$, $\epsilon_t \sim N(0, v)$. Then,

$$\begin{aligned} \gamma(0) &= \frac{v}{(1-\phi^2)}, \\ \gamma(h) &= \phi^{|h|}\gamma(0). \end{aligned}$$

ACF of AR(1)



Sample ACF of AR(1)



R functions

```
ts.plot()

arima.sim(model,n,rand.gen = rnorm, innov = rand.gen(n,
...),...)

acf(x,lag.max=NULL,type=c("correlation","covariance",
"partial"),plot=TRUE,na.action=na.fail,demean=TRUE,...)
```

Example:

```
ts.plot(arima.sim(model=list(ar=c(0.9)),n=100))
```

Smoothing

Decompose the data as "signal" + "noise" where the signal is a "smooth" component.

- *Moving averages.* Apply a linear filter of the form

$$z_t = \sum_{j=-q}^{p} a_j y_{t+j},$$

with $\sum_{j=-q}^{p} a_j = 1$. It is generally assumed that $p = q$, $a_j \geq 0$ and $a_j = a_{-j}$. In this case the order of the moving average is $2p + 1$.

Useful `R` functions:

```
filter(x,filter, method,...)
```

Example:
$z_t = 0.125y_{t-2} + 0.25y_{t-1} + 0.25y_t + 0.25y_{t+1} + 0.125y_{t+2}$:

```
z=filter(y,filter=c(0.125,0.25,0.25,0.25,0.125),
sides=2)
```

Smoothing

- *Polynomial and harmonic regressions.* Fit models of the form:

$$\begin{aligned} y_t &= \beta_0 + \beta_1 t + \ldots + \beta_p t^p + \epsilon_t, \\ y_t &= a_1 \cos(2\pi\omega_1 t) + b_1 \sin(2\pi\omega_1 t) + \ldots \\ &\quad + a_p \cos(2\pi\omega_p t) + b_p \sin(2\pi\omega_p t) + \epsilon_t, \end{aligned}$$

- *Kernel smoothing.*

$$z_t = \sum_{i=1}^{T} w_t(i) y_t, \quad w_i(t) = K\left(\frac{t-i}{b}\right) / \sum_{j=1}^{T} K\left(\frac{t-j}{b}\right),$$

  where $K(\cdot)$ is a kernel function, such as a normal kernel. The parameter b is a bandwidth. The larger the value of b, the smoother z_t is.

Smoothing

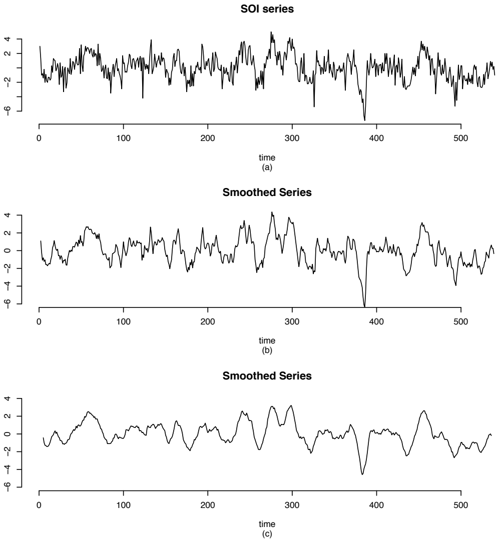
- *Splines.*
- *Lowess function.*

`R` functions:

```
kernel()

spline()

lowess()
```


SOI series
4
2
−2
−6
0
100
200
300
400
500
time
(a)
Smoothed Series
4
2
0
−2
−4
−6
0
100
200
300
400
500
time
(b)
Smoothed Series
4
2
0
−2
−4
−6
0
100
200
300
400
500
time
(c)

Differencing

- First order difference:

$$Dy_t = (1 - B)y_t = y_t - y_{t-1},$$

  with B the backward operator.
- Higher order differences:

$$D^d y_t = (1 - B)^d y_t.$$

R function: `diff()`.

Bayes' theorem: Independent Observations

$$p(\theta|y_{1:T}) = \frac{\overbrace{p(y_{1:T}|\theta)}^{\text{likelihood}} \times \overbrace{p(\theta)}^{\text{prior}}}{\underbrace{p(y_{1:T})}_{\text{predictive}}},$$

with

$$p(y_{1:T}) = \int p(y_{1:T}|\theta)p(\theta)d\theta.$$

Alternatively, we can write:

$$\begin{aligned} p(\theta|y_{1:T}) &\propto p(y_T|y_{1:(T-1)}, \theta) \times p(\theta|y_{1:(T-1)}) \\ &\propto \underbrace{p(y_T|\theta)}_{\text{likelihood}} \times \underbrace{p(\theta|y_{1:(T-1)})}_{\text{prior}}. \end{aligned}$$

Bayes' theorem: Dependence on $(t-1)$

$$p(\theta|y_{1:T}) \quad \propto \quad \underbrace{p(\theta)}_{\text{prior}} \underbrace{p(y_1|\theta) \prod_{t=2}^{T} p(y_t|y_{t-1},\theta)}_{\text{likelihood}} .$$

Bayes' theorem: Dependence on $(t-1)$

Example

AR(1)**:** $y_t = \phi y_{t-1} + \epsilon_t$, $\epsilon_t \sim N(0, v)$, and so $\boldsymbol{\theta} = (\phi, v)$.

- Conditional likelihood: $p(y_t | y_{t-1}, \boldsymbol{\theta}) = N(y_t | \phi y_{t-1}, v)$;
- $p(y_1 | \boldsymbol{\theta}) = N(0, v/(1-\phi^2))$;

Then,

$$p(\boldsymbol{\theta} | y_{1:T}) \propto p(\boldsymbol{\theta}) \frac{\sqrt{(1-\phi^2)}}{(2\pi v)^{T/2}} \exp\left\{ -\frac{Q^*(\phi)}{2v} \right\},$$

with

$$Q^*(\phi) = y_1^2(1-\phi^2) + \underbrace{\sum_{t=2}^{T} (y_t - \phi y_{t-1})^2}_{Q(\phi)}$$

Bayes' theorem: Dependence on $(t-1)$

Example

AR(1) **(cont.):** We can also use the *conditional likelihood* $p(y_{2:T}|\theta, y_1)$ as an approximation to the full likelihood and obtain the posterior

$$p(\theta|y_{1:T}) \propto p(\theta) v^{-(T-1)/2} \exp\left\{-\frac{Q(\phi)}{2v}\right\}$$

Estimation

- *Maximum likelihood estimation (MLE):* Find $\hat{\theta} = \theta_{\text{MLE}}$ that maximizes $p(y_{1:T}|\theta)$.
- *Maximum a posteriori estimation (MAP):* Find $\hat{\theta} = \theta_{\text{MAP}}$ that maximizes $p(\theta|y_{1:T})$.
- *Least squares estimation (LSE):* Write the model as

$$\mathbf{y} = \mathbf{F}'\beta + \epsilon, \quad \epsilon \sim N(\mathbf{0}, v\mathbf{I}).$$

with $\dim(\mathbf{y}) = n$ and $\dim(\beta) = p$ so that

$$p(\mathbf{y}|\mathbf{F}', \beta, v) = (2\pi v)^{-n/2}\exp(-Q(\mathbf{y}, \beta)/2v)\,,$$

and find $\hat{\beta}$ that minimizes $Q(\mathbf{y}, \beta)$.

Bayesian Estimation

Consider again the model $\mathbf{y} = \mathbf{F}'\beta + \epsilon$, with $\epsilon \sim N(\mathbf{0}, v\mathbf{I})$. The posterior density is given by

$$\begin{aligned} p(\beta, v|\mathbf{y}) &\propto p(\beta, v) \times p(\mathbf{y}|\beta, v) \\ &\propto p(\beta, v) \times (2\pi v)^{-n/2} \exp(-Q(\mathbf{y}, \beta)/2v) \end{aligned}$$

where

$$Q(\beta, \mathbf{y}) = (\mathbf{y} - \mathbf{F}'\beta)'(\mathbf{y} - \mathbf{F}'\beta) = (\beta - \hat{\beta})'(\mathbf{F}\mathbf{F}')(\beta - \hat{\beta}) + R,$$

with $\hat{\beta} = (\mathbf{F}\mathbf{F}')^{-1}\mathbf{F}\mathbf{y}$ and $R = (\mathbf{y} - \mathbf{F}'\hat{\beta})'(\mathbf{y} - \mathbf{F}'\hat{\beta})$.

- The MLE of β is $\hat{\beta}$;
- The MLE of v is R/n, however, $s^2 = R/(n-p)$ is used instead.

Bayesian Estimation

Reference prior: $p(\beta, v) \propto 1/v \Rightarrow$

- $p(\beta|\mathbf{y}, \mathbf{F})$ is Student-t with $n-p$ degrees of freedom, mode $\hat{\beta}$ and density

$$p(\beta|\mathbf{y}, \mathbf{F}) \propto |\mathbf{F}\mathbf{F}'|^{1/2} \left\{1 + (\beta - \hat{\beta})'\mathbf{F}\mathbf{F}'(\beta - \hat{\beta})/(n-p)s^2\right\}^{-n/2}$$

- When $n \to \infty$ $p(\beta|\mathbf{y}, \mathbf{F}) \approx N(\beta|\hat{\beta}, s^2(\mathbf{F}\mathbf{F}')^{-1})$.
- $p(v|\mathbf{y}) = \text{IG}\left(\frac{(n-p)}{2}, \frac{(n-p)s^2}{2}\right)$.

Bayesian Estimation

Conjugate Prior:

$$p(\beta, v) = p(\beta|v)p(v) = N(\beta|\mathbf{m}_0, v\mathbf{C}_0) \times IG(v|n_0/2, d_0/2)$$

$\Rightarrow$

$$p(\beta, v|\mathbf{F}, \mathbf{y}) \propto v^{-\{(p+n+n_0)/2+1\}} \times e^{-\frac{(\beta-\mathbf{m}_0)'\mathbf{C}_0^{-1}(\beta-\mathbf{m}_0)+(\mathbf{y}-\mathbf{F}'\beta)'(\mathbf{y}-\mathbf{F}'\beta)+d_0}{2v}}$$

- $(\mathbf{y}|\mathbf{F}, v) \sim N(\mathbf{F}'\mathbf{m}_0, v(\mathbf{F}'\mathbf{C}_0\mathbf{F} + \mathbf{I}_n))$;
- $(\beta|\mathbf{F}, v) \sim N(\mathbf{m}, v\mathbf{C})$, with

$$\begin{aligned}
\mathbf{m} &= \mathbf{m}_0 + \mathbf{C}_0\mathbf{F}[\mathbf{F}'\mathbf{C}_0\mathbf{F} + \mathbf{I}_n]^{-1}(\mathbf{y} - \mathbf{F}'\mathbf{m}_0) \\
\mathbf{C} &= \mathbf{C}_0 - \mathbf{C}_0\mathbf{F}[\mathbf{F}'\mathbf{C}_0\mathbf{F} + \mathbf{I}_n]^{-1}\mathbf{F}'\mathbf{C}_0,
\end{aligned}$$

Bayesian Estimation (Conjugate prior)

- $(v|\mathbf{F}, \mathbf{y}) \sim \text{IG}(n^*/2, d^*/2)$ with $n^* = n + n_0$ and $d^* = \mathbf{e}'\mathbf{Q}^{-1}\mathbf{e} + d_0$, with

$$\mathbf{e} = (\mathbf{y} - \mathbf{F}'\mathbf{m}_0), \quad \text{and} \quad \mathbf{Q} = (\mathbf{F}'\mathbf{C}_0\mathbf{F} + \mathbf{I}).$$

- $(\beta|y_{1:n}, \mathbf{F}) \sim T_{n^*}[\mathbf{m}, d^*\mathbf{C}/n^*]$.

Estimation

Example

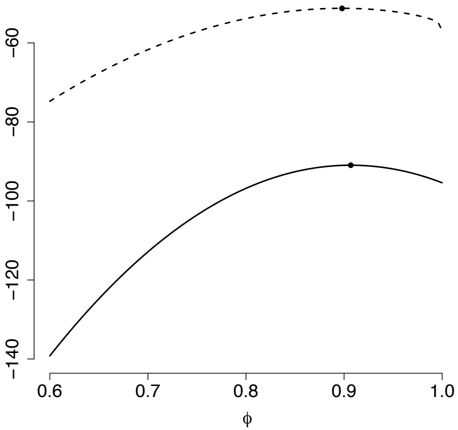
ML, MAP, and LS estimators for the AR(1) model.
$y_t = \phi y_{t-1} + \epsilon_t$, with $\epsilon_t \sim N(0, 1)$. In this case $\theta = \phi$.

- The conditional MLE is found by maximizing $\exp\{-Q(\phi)/2\}$ (or by minimizing $Q(\phi)$). Therefore, $\hat{\phi} = \phi_{\mathrm{ML}} = \sum_{t=2}^{T} y_t y_{t-1} / \sum_{t=2}^{T} y_{t-1}^2$.
- MLE of unconditional likelihood is obtained by maximizing $p(y_{1:T}|\phi)$ or by minimizing

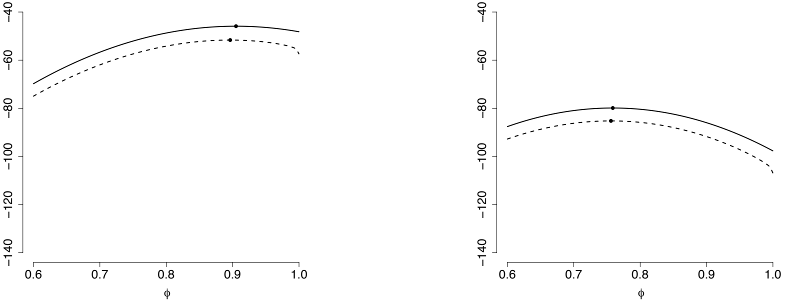
$$-0.5[\log(1-\phi^2) - Q^*(\phi)].$$

We need methods such as Newton-Raphson or scoring to find ϕ_{ML}.

AR(1) conditional and unconditional likelihoods; simulated data with $\phi = 0.9$; MLEs $\hat{\phi} = 0.9069$ and $\hat{\phi} = 0.8979$.



AR(1) **conditional and unconditional posteriors with priors** $N(0, c)$, $c = 1$ **and** $c = 0.01$



Bayesian Estimation

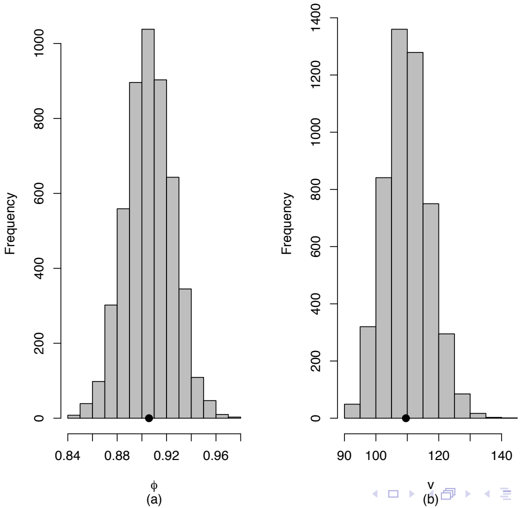
Reference analysis in the AR(1) model.

- For the conditional likelihood $\phi_{\text{ML}} = \sum_{t=2}^{T} y_{t-1} y_t / \sum_{t=1}^{T-1} y_t^2$.
- Under the reference prior $\phi_{\text{MAP}} = \phi_{\text{ML}}$.
- Also,

$$R = \sum_{t=2}^{T} y_t^2 - \frac{(\sum_{t=2}^{T} y_t y_{t-1})^2}{\sum_{t=1}^{T-1} y_t^2},$$

  and so $s^2 = R/(T-2)$ estimates v.
- Marginal posterior for ϕ : Student-t with $T-2$ degrees of freedom, centered at ϕ_{ML} with scale $s^2(\mathbf{F}\mathbf{F}')^{-1}$.
- Marginal posterior for v : $Inv - \chi^2(v|T-2, s^2)$ or, equivalently, $IG(v|(T-2)/2, (T-2)s^2/2)$.

AR(1) reference analysis; 500 simulated observations with $\phi = 0.9$ and $v = 100$.



Bayesian Estimation: Non-Conjugate Analysis

AR(1) **with full likelihood:** The prior $p(\phi, v) \propto 1/v$ does not lead to a closed form posterior distribution when the full likelihood is used. We obtain

$$p(\phi, v|y_{1:n}) \propto v^{-(n/2+1)}(1-\phi^2)^{1/2}\exp\left\{\frac{-Q^*(\phi)}{2v}\right\}.$$

How can we summarize posterior inference in this case?

- Via simulation-based methods such as Markov chain Monte Carlo...

MCMC: The Metropolis Hastings Algorithm

Creates a sequence of random draws, $\theta^{(1)}, \theta^{(2)}, \ldots,$ whose distributions converge to the target distribution, $p(\theta|y_{1:n})$.

1. Draw $\theta^{(0)}$ with $p(\theta^{(0)}|y_{1:n}) > 0$ from $p_0(\theta)$.
2. For $m = 1, 2, \ldots,$ (until convergence):
   (a) Sample $\theta^* \sim J(\theta^*|\theta^{(m-1)})$
   (b) Compute the importance ratio

$$r = \frac{p(\theta^*|y_{1:n})/J(\theta^*|\theta^{(m-1)})}{p(\theta^{(m-1)}|y_{1:n})/J(\theta^{(m-1)}|\theta^*)}.$$

   (c) Set

$$\theta^{(m)} = \begin{cases} \theta^* & \text{with probability} = \min(r, 1) \\ \theta^{(m-1)} & \text{otherwise.} \end{cases}$$

MCMC: AR(1) case

MCMC for AR(1) with full likelihood.

- Sample $v^{(m)}$ from $(v|\phi, y_{1:n}) \sim IG(n/2, Q^*(\phi)/2)$ (Gibbs step, every draw will be accepted)
- Sample $\phi^* \sim N\left(\phi^{(m-1)}, c\right)$.

MCMC: AR(1) example