



Análisis de Series de Tiempo con Modelos Dinámicos

Romy Rodriguez Ravines

`romy@dme.ufrj.br`

`dme.ufrj.br/romy`

Lima, 12 a 16 de marzo del 2007

Parte I

Introducción

Análisis de series de tiempo

Serie de tiempo

- Una **serie de tiempo** es una colección de observaciones ordenadas (en el tiempo).
- En los modelos de regresión usuales, el orden de las observaciones es irrelevante, en una serie de tiempo el paso del tiempo tiene un efecto muy importante. **Observaciones adyacentes son dependientes.**
- Ejemplos: índice de precios al consumidor (economía), exportaciones e importaciones (economía), número de casos notificados de dengue (epidemiología), tasa de incidencia de leptospirose (epidemiología), precipitación acumulada diaria (meteorología), etc.

Ejemplo 1: Ventas de un caramelo

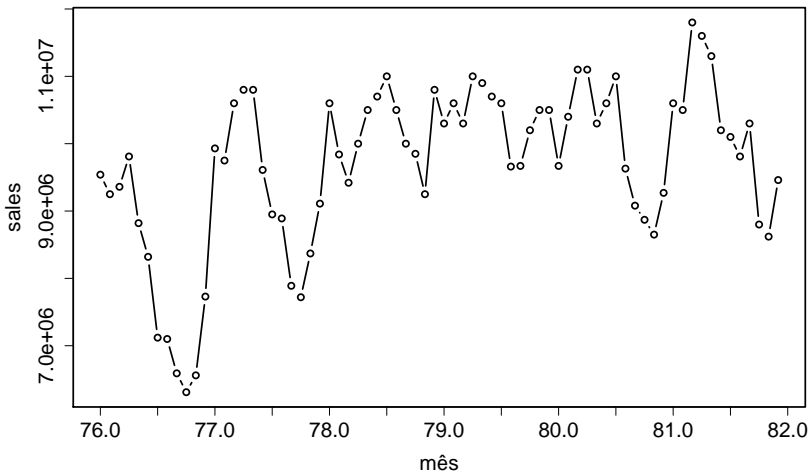


Figura: Ventas mensuales de un caramelo

Ejemplo 2: Número de accidentes graves

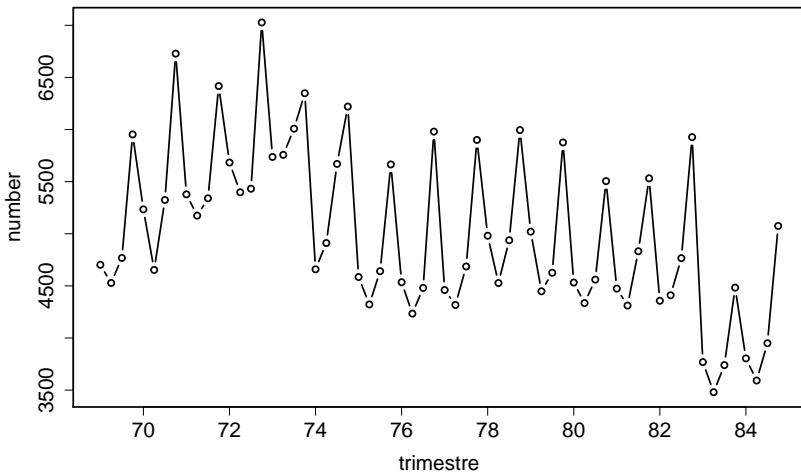


Figura: Número de accidentes graves en carreteras

Ejemplo 3: *Market share*

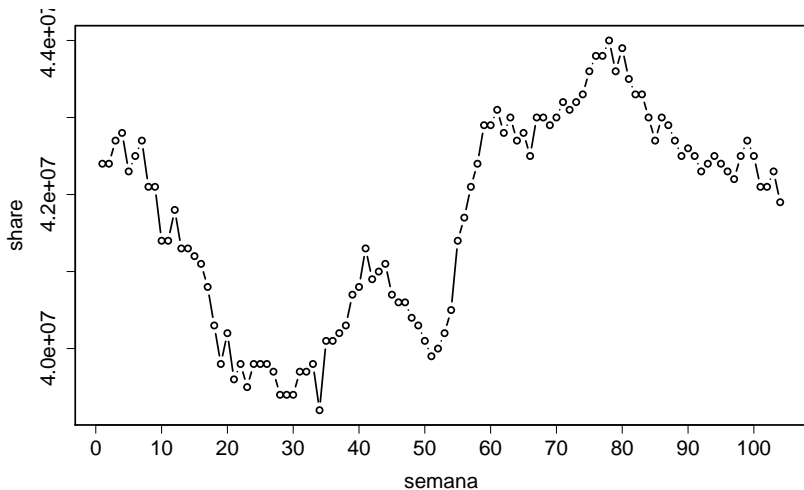


Figura: Serie semanal de *market share*

Objetivos del análisis de series de tiempo

(1) Entender el mecanismo de generación de la serie de tiempo

- describir eficientemente el comportamiento de la serie
- encontrar “razones” o “explicaciones” para el comportamiento de la serie (talvez con otras variables)
- controlar la trayectoria de la serie

(2) Prever valores futuros de las series

- hacer planes de corto, medio y largo plazo
- tomar decisiones apropiadas

- Ambos objetivos están relacionados. Solamente es posible hacer una buena previsión si el modelo es adecuado y viceversa.
- Normalmente el futuro implica incertidumbre, en consecuencia, las previsiones no son perfectas.
- Un objetivo general es **reducir los errores de previsión** lo máximo posible.

Modelando, aprendiendo y pronosticando

Modelo

Esquema de descripción (y explicación) que organiza la información (y la experiencia) para las etapas de aprendizaje y pronóstico.



Aprendizaje

Procesamiento de la información con el modelo



Pronóstico

Es una hipótesis, conjetura o especulación sobre el futuro.

Sobre el modelo

- Un buen modelo permite el aprendizaje para realizar pronósticos adecuados.
- Debe ser **parsimonioso**
- La descripción debe ser simple y flexible para adaptar el futuro (incertidumbre) y facilitar el aprendizaje
- En la presencia de incertidumbre, el modelo es **probabilístico**.
- La naturaleza dinámica de las series de tiempo requiere modelos que se adapten con el tiempo
- La estructura del modelo debe **permitir cambios locales**
- Los cambios pueden ser modelados de forma estocástica, i.e., usando probabilidad.

Sobre el modelo (cont.)

- Idea básica: definir modelos representando la estructura de la serie (tendencia, ciclo, estacionalidad, etc.). De allí viene el nombre de **modelos estructurales**.
- Principal característica: **los parámetros del modelo varían probabilísticamente con el tiempo**.
- Por ejemplo, si se tiene interés en pronosticar Y , y se sabe que está influenciada por X . Se puede establecer la relación

$$Y_t = X_t\theta + \epsilon.$$

En este caso, θ debe variar con el tiempo.

- Cuando se usa inferencia bayesiana se denominan **modelos dinámicos**.

Sobre el modelo: Modelos lineales dinámicos

- En un modelo de regresión usual:

$$Y_t = \mathbf{X}_t' \boldsymbol{\theta} + v_t, \quad v_t \sim N(0, \sigma^2), \quad t = 1, \dots, T. \quad (1)$$

$\boldsymbol{\theta}$ es un vector $p \times 1$ de coeficientes asociados a un conjunto de p variables explicativas, \mathbf{X}_t .

- En un modelo dinámico, los **parámetros varían con el tiempo**:

$$Y_t = \mathbf{X}_t' \boldsymbol{\theta}_t + v_t.$$

El único cambio con relación a (1) es el **índice de $\boldsymbol{\theta}$** .

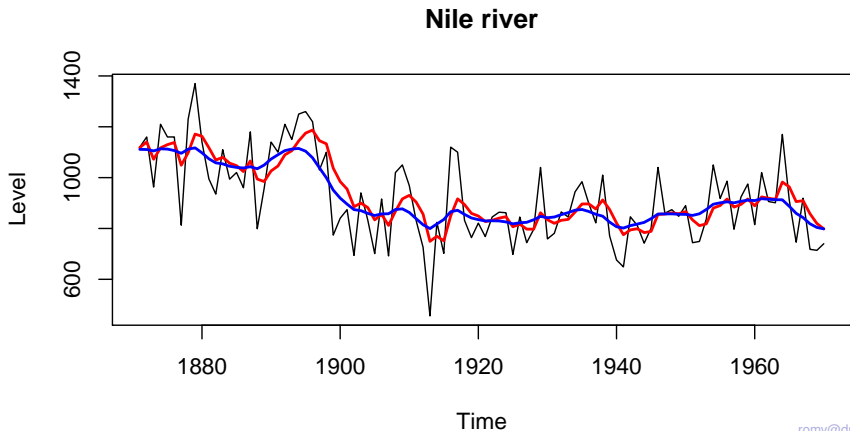
- Si el sistema es estático: $\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1} = \boldsymbol{\theta}$.
- Si el sistema es dinámico:

$$\boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \mathbf{w}_t.$$

\mathbf{G}_t es una matriz de valores conocidos y \mathbf{w}_t es un error aleatorio.

Sobre el modelo: ejemplo

Figura: Ejemplo de modelo dinámico de primer orden



Sobre el aprendizaje

- Regla de aprendizaje: **Teorema de Bayes**
- Modelos posibles: M_1, M_2, \dots, M_M , cada uno con probabilidad a priori $p(M), M = M_1, M_2, \dots, M_M$.
- Después de observar $Y = y$ se tiene $P(M|Y = y)$ dada por

$$\begin{aligned} P(M|Y = y) &= \frac{P(Y = y, M)}{P(Y = y)} \\ &= \frac{P(Y = y|M) \times P(M)}{P(Y = y)} \end{aligned}$$

Luego,

$$\underbrace{P(M|Y = y)}_{\text{posteriori}} \propto \underbrace{P(Y = y|M)}_{\text{verosimilitud}} \times \underbrace{P(M)}_{\text{priori}}$$

Sobre el pronóstico

Intervención

Modificar el modelo de acuerdo con la información disponible. Por ejemplo, serie de ventas de pasajes aéreos \leftarrow 11S

Monitoramiento

Ajuste del modelo no es bueno \rightarrow se activa una señal (monitor) \rightarrow cambios en el modelo

Análisis Retrospectivo

El pasado da información sobre el futuro + el futuro da información sobre el pasado \rightarrow revisar estimación

Objetivos del curso

- Presentar a los usuarios los conceptos y técnicas fundamentales del análisis de Series Temporales con Modelos Dinámicos, bajo la perspectiva bayesiana.
- Ilustrar los tópicos abordados con casos prácticos utilizando los softwares `BATS`, `R` y `WinBUGS`.
- Al final del curso, los alumnos tendrán habilidad para analizar y criticar los resultados obtenidos con este tipo de análisis en las series de tiempo de su interés.

Referencias principales

- Pole, West & Harrison (1994). *Applied Bayesian Forecasting and Time Series Analysis*. New York: Chapman-Hall.
- West & Harrison (1997). *Bayesian Forecasting and Dynamic Models*. New York: Springer.

Softwares principales:

BATS `ftp.stat.duke.edu/pub/bats/`

R `http://www.r-project.org`
(The dlm package: Petris, 2006)

WinBUGS `http://www.mrc-bsu.cam.ac.uk/bugs/`

BTS `http://lib.stat.cmu.edu/DOS/S/`
(SPLUS for Windows functions)

Organización del curso I

Introducción

- 1 Introducción
- 2 Contenido
- 3 Revisión de probabilidad e inferencia bayesiana

Organización del curso II

Modelos Lineales Dinámicos Normales: Definición

- 4 Modelo lineal dinámico
- 5 Modelos de tendencia
- 6 Análisis secuencial
- 7 Evolución del sistema
- 8 Análisis retrospectivo

Organización del curso III

Modelos Lineales Dinámicos Normales: Aspectos Prácticos

- 9 Modelos de regresión dinámicos
- 10 Modelos con estacionalidad
- 11 Superposición de modelos
- 12 Monitoramiento
- 13 Intervención

Organización del curso IV

Inferencia con Métodos de Simulación Estocástica

- 14 Métodos de simulación estocástica
- 15 Monte Carlo via cadenas de Markov (MCMC)
- 16 Inferencia en MLD con MCMC

Modelos Dinámicos No-Lineales No-Normales

- 17 Modelos lineales dinámicos generalizados (MLDG)
- 18 Modelos no lineales dinámicos generalizados (MNLDG)

Revisión de probabilidad e inferencia bayesiana

Probabilidad

- **Probabilidad** es un número entre 0 y 1. Representa la creencia en una determinada afirmación. Por ejemplo,

$$Pr(\text{cara en el lanzamiento de una moneda}) = 0,5$$

$$Pr(\text{hoy llueve}) = 0,1$$

- **Probabilidad total** (0 ó 1) representa la creencia en la veracidad o falsedad de una afirmación.
- **Probabilidad condicional** es la probabilidad con base en el conocimiento previo de la veracidad de una afirmación. Por ejemplo,

$$Pr(\text{resultado par en un dado lanzado} | \text{resultado fue } \leq 5).$$

- Tenemos que $Pr(A|B) = \frac{Pr(AB)}{Pr(B)}$. Si $A = \{2, 4, 6\}$, $B = \{1, 2, 3, 4, 5\}$ e $AB = \{2, 4\}$, $Pr(AB) = 2/6$ y $Pr(B) = 5/6$. Luego, $Pr(A|B) = \frac{2/6}{5/6} = \frac{2}{5}$

Teorema de Bayes (cont.)

- También tenemos que $Pr(B|A) = \frac{Pr(AB)}{Pr(A)}$, entonces,
 $Pr(AB) = Pr(B|A) \times P(A)$. Luego,

$$Pr(A|B) = \frac{Pr(AB)}{Pr(B)} = \frac{Pr(B|A) \times Pr(A)}{Pr(B)} \propto Pr(B|A)Pr(A)$$

Este resultado es conocido como **Teorema de Bayes**.

- Es la base de la inferencia bayesiana pues **enseña como actualizar nuestra creencia en A después de recibir nuevas informaciones, por ejemplo, B**

Teorema de Bayes (cont.)

Teorema (Teorema de Bayes)

Si B_1, B_2, \dots, B_n son eventos mutuamente excluyentes de los cuales uno debe ocurrir, entonces,

$$P(B_r | A) = \frac{P(B_r)P(A | B_r)}{\sum_{i=1}^n P(B_i)P(A | B_i)} \quad (2)$$

para $r = 1, 2, \dots, \text{o } n$.

Variables aleatorias

- **Variable o cantidad aleatoria** es aquella cuyo valor no conocemos. Por ejemplo,
 - X : Número resultado del lanzamiento de un dado
 - Y : Nivel de glicosís en la sangre de una persona.
- Su incertidumbre es representada probabilísticamente, p.e., $Pr(X = x) = \frac{1}{6}, x = 1, 2, 3, 4, 5, 6.$
- Las variables aleatorias tienen varias clasificaciones:
 - **Discreta** o **Continua**
 - **Observable** o **No observable**

Ejemplos:

Discreta	observable	:	X
Discreta	no observable	:	Indicador de enfermedad en una persona
Continua	observable	:	Ventas de un producto en gran escala
Continua	no observable	:	Y

Distribución de probabilidad

- **Variables discretas** son representadas por la **función de probabilidad** $p(x) = Pr(X = x)$, que caracteriza completamente la incertidumbre con relación a X pues

$$Pr(X \in A) = \sum_{x \in A} Pr(X = x) = \sum_{x \in A} p(x)$$

- **Variables continuas** son representadas por la función de densidad de probabilidad o **densidad** $p(x)$, que caracteriza la incertidumbre con relación a X pues

$$Pr(X \in A) = \int_A p(x) dx$$

- Una función de probabilidad o una densidad definen una **distribución de probabilidad**.
- Características importantes: medidas de **posición** (media, mediana, moda) y medidas de **dispersión** (varianza, desvío estándar)

Distribución normal univariada

Definición

La v.a. (continua) X tiene una distribución normal con media μ y varianza σ^2 , denotada por $X \sim N[\mu, \sigma^2]$, si su densidad es

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2 \right\}, x \in \mathbb{R}.$$

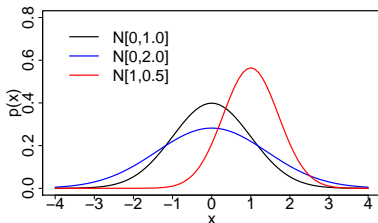


Figura: Ejemplos de $N[\mu, \sigma^2]$

- $Pr(X \in A) = \int_A p(x)dx$ sólo puede ser calculada numéricamente.
- Resultados para la normal estandar, $N[0, 1]$:
 - $Pr(X \in [-2, 2]) = 0,95$
 - $Pr(X \in [-1,64, 1,64]) = 0,90$

Teorema de Bayes para variables aleatorias

- μ es una cantidad desconocida. La **incertidumbre** sobre μ es representada por $p(\mu)$.
- X es una cantidad relacionada con μ (por ejemplo, una medición) cuya incertidumbre esta representada por $p(x|\mu)$.
- Si observamos $X = x$, la incertidumbre sobre μ pasa a ser $p(\mu|x)$. Por el teorema de Bayes,

$$p(\mu|x) = \frac{p(x|\mu)p(\mu)}{p(x)} \propto p(x|\mu)p(\mu)$$

- **El teorema de Bayes enseña como actualizar nuestro conocimiento después de recibir información relevante.**

Teorema de Bayes para variables normales

- Se sabe que el nivel de glucosa en una persona normal, μ , puede ser descrito por una distribución $N[120, 100]$:

$$p(\mu) = \frac{1}{\sqrt{200\pi}} \exp \left\{ -\frac{1}{2} \left(\frac{\mu - 120}{10} \right)^2 \right\}$$

- Entonces, $Pr(\mu \text{ estar entre } 100 \text{ y } 140) = 0,95$
- Una observación, y , es realizada en un laboratorio. Si $p(y|\mu)$ es $N(\mu, 25)$, i.e.

$$p(y|\mu) = \frac{1}{\sqrt{50\pi}} \exp \left\{ -\frac{1}{2} \left(\frac{y - \mu}{5} \right)^2 \right\}$$

Entonces, las chances de y no representar (equivocarse) el nivel de glucosa en más de 10ml son de 5 %.

Teorema de Bayes para variables normales (cont.)

- Se recoge una muestra y **el valor observado es 127ml.**
- Ahora la incertidumbre sobre el nivel de glucosa esta dada por

$$\begin{aligned}
 p(\mu|y = 127) &\propto p(127 | \mu)p(\mu) \\
 &\propto \exp \left\{ -\frac{1}{2} \left(\frac{127 - \mu}{5} \right)^2 \right\} \exp \left\{ -\frac{1}{2} \left(\frac{\mu - 120}{10} \right)^2 \right\} \\
 &\propto \exp \left\{ -\frac{1}{2} \left(\frac{\mu - 125,6}{4,5} \right)^2 \right\}
 \end{aligned}$$

- Luego, $p(\mu|y = 127)$ es $N(125,6, 20,25)$ y

$$Pr(\mu \text{ estar entre } 116.6 \text{ y } 134.6) = 0,95$$

Teorema de Bayes para variables normales (cont.)

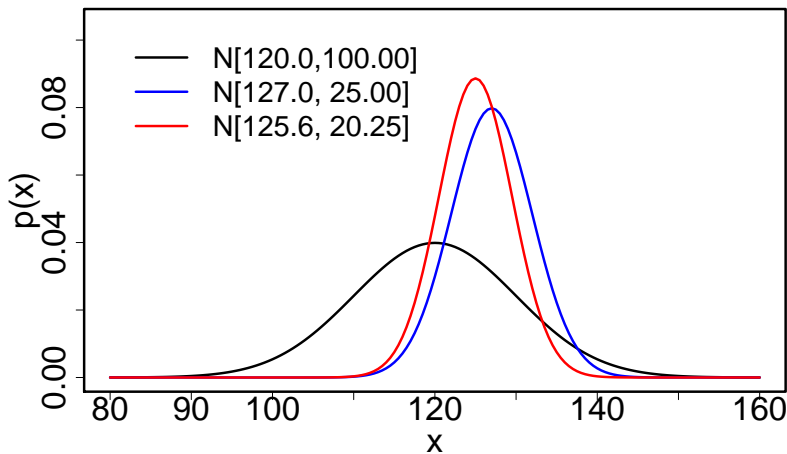


Figura: Ejemplo de la aplicación del teorema de Bayes para variables normales

Inferencia bayesiana: ejemplo

Deseamos estimar μ . Realizamos n mediciones tales que:
 $x_i = \mu + e_i$, $i = 1, \dots, n$ y $e_i \sim N(0, \sigma^2)$, σ^2 conocida. Entonces,

$$p(x|\mu) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2} \frac{(x_i - \mu)^2}{\sigma^2} \right\}.$$

Usando el teorema de Bayes tenemos

$$\underbrace{p(\mu|x)}_{\text{posteriori}} \propto \underbrace{p(x|\mu)}_{\text{verosimilitud}} \times \underbrace{p(\mu)}_{\text{priori}}$$

$$\underbrace{\exp \left\{ -\frac{1}{2} \frac{(\mu - a_2)^2}{b_2} \right\}}_{\text{posteriori}} \propto \underbrace{\exp \left\{ -\frac{n}{2} \frac{(\bar{x} - \mu)^2}{\sigma^2} \right\}}_{\text{verosimilitud}} \times \underbrace{\exp \left\{ -\frac{1}{2} \frac{(\mu - a_1)^2}{b_1} \right\}}_{\text{priori}}$$

Luego, la media y varianza de la distribución a posteriori de μ son:

$$a_2 = b_2(n\sigma^{-2}\bar{x} + b_1^{-1}a_1) \quad \text{e} \quad b_2 = n\sigma^{-2} + b_1^2$$

Inferencia bayesiana: *remarks*

- El **proceso de inferencia** está basado en la distribución de μ dada la observación de $x =$ la **distribución a posteriori** $p(\mu|x)$, obtenida a través del teorema de Bayes:

$$\underbrace{p(\mu|x)}_{\text{posteriori}} \propto \underbrace{p(x|\mu)}_{\text{verosimilitud}} \times \underbrace{p(\mu)}_{\text{priori}}$$

- Para completar el modelo es necesario especificar la distribución a priori** \Rightarrow Informativa / Poco Informativa / Vaga - Propia / Impropia.
- Problema: generalmente conocemos la **forma funcional de** $p(\mu|x)$. En la mayoría de los casos esa forma es **compleja**, de **dimensión alta** y es difícil evaluarla analíticamente.
- La “constante” $p(x)$ no necesita ser calculada pues $p(\mu|x)$ es una densidad y podemos reconocer su forma sin las constantes.

Vectores aleatorios

- Una **colección de variables aleatorias** puede ser agrupada en un **vector aleatorio**.
- Si todos los componentes de un vector aleatorio son discretos, el vector también es **discreto** con función de probabilidad :

$$f(x_1, \dots, x_p) = Pr(X_1 = x_1, \dots, X_p = x_p)$$

$$\text{y } Pr((X_1, \dots, X_p) \in A) = \sum_{(x_1, \dots, x_p) \in A} f(x_1, \dots, x_p)$$

- Si todos los componentes de un vector son continuos, el vector será **continuo** con densidad $f(x_1, \dots, x_p)$ y

$$Pr((X_1, \dots, X_p) \in A) = \int_A f(x_1, \dots, x_p) dx_1 \dots dx_p$$

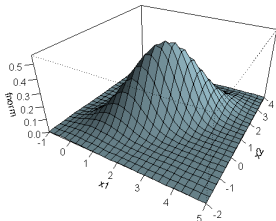
Distribución normal multivariada

Definición

$\mathbf{X} = (X_1, \dots, X_p)$ tiene una distribución normal multivariada con vector de medias $\boldsymbol{\mu}$ y matriz de varianzas-covarianzas $\boldsymbol{\Sigma}$, denotada por $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, si su densidad es

$$f_p(\mathbf{x}) = (2\pi)^{p/2} |\boldsymbol{\Sigma}|^{1/2} \exp \left\{ -\frac{1}{2} \left((\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right)^2 \right\}.$$

Si $(X_1, \dots, X_p) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, entonces cada componente $X_j \sim N(\mu_j, \sigma_{jj})$ donde μ_j es el j -ésimo componente del vector $\boldsymbol{\mu}$ y σ_{jj} es el j -ésimo componente de la diagonal de la matriz $\boldsymbol{\Sigma}$.



Resultados principales

- **Teorema de Bayes para variables aleatorias:**

$$p(\mu|x) = \frac{f(x|\mu)p(\mu)}{f(x)} \propto f(x|\mu)p(\mu)$$

- **Normal Multivariada - Distribución Condicional**

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \Big| \boldsymbol{\mu}, \boldsymbol{\Sigma} \sim N \left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_1 & \Sigma_{12} \\ \Sigma_{12} & \Sigma_2 \end{pmatrix} \right]$$

$$Y_1|Y_2, \boldsymbol{\mu}, \boldsymbol{\Sigma} \sim N[\mu_1 + \Sigma_{12}\Sigma_2^{-1}(Y_2 - \mu_2), \Sigma_1 - \Sigma_{12}\Sigma_2^{-1}\Sigma_{21}]$$

- **Normal- Gamma Multivariada**

Si $Y|\boldsymbol{\mu}, \phi, \boldsymbol{\Sigma} \sim N[\boldsymbol{\mu}, \boldsymbol{\Sigma}\phi^{-1}]$ y $\phi \sim \text{Gamma}(n/2, d/2)$ entonces:

$$Y \sim t_n(\boldsymbol{\mu}, d/n\boldsymbol{\Sigma})$$

Parte II

Modelos Lineales Dinámicos Normales: Definición e Inferencia Secuencial

Modelo lineal dinámico

Modelo de regresión usual

- En un modelo de regresión tenemos una variable respuesta Y que es explicada por un conjunto de variables explicativas X_1, \dots, X_p a través de la relación

$$Y = \theta_0 + \theta_1 X_1 + \dots + \theta_p X_p + v \quad (3)$$

En general se asume que v tiene distribución $N(0, \sigma^2)$.

- La ecuación (3) puede ser escrita como:

$$Y = \mathbf{X}'\boldsymbol{\theta} + v, \quad (4)$$

donde:

$$\mathbf{X} = \begin{pmatrix} 1 \\ X_1 \\ \vdots \\ X_p \end{pmatrix} \quad \text{y} \quad \boldsymbol{\theta} = \begin{pmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_p \end{pmatrix}$$

Modelo de regresión usual (cont.)

- La naturaleza de las variables explicativas o regresoras es bastante amplia. De esta manera se puede utilizar cualquier variable cuantificable.
- Los **parámetros** o **coeficientes** de regresión $\theta_1, \dots, \theta_p$ informan sobre la influencia que las regresoras tienen sobre la respuesta Y . En la práctica, sus valores son **desconocidos**.
- Los coeficientes son estimados a partir de una colección de observaciones realizadas sobre o modelo (3). Son observadas respuestas Y_1, \dots, Y_n con sus respectivos regresoras X_1, \dots, X_n , y simbólicamente se tiene:

$$Y_t = \mathbf{X}_t' \boldsymbol{\theta} + v_t, \quad t = 1, \dots, n. \quad (5)$$

Modelo de regresión dinámico

- En un modelo dinámico, los parámetros varían con el tiempo. El modelo de regresión es extendido para

$$Y_t = \mathbf{X}'_t \boldsymbol{\theta}_t + v_t, \quad t = 1, \dots, n. \quad (6)$$

donde el único cambio con relación al modelo en (5) es el índice en $\boldsymbol{\theta}$

- El modelo en (6) tiene un gran número de parámetros para estimar. Este modelo necesita de más información.
- Los parámetros sucesivos están íntimamente relacionados. En general, un parámetro es igual a su antecesor más una pequeña perturbación causada por los cambios a los que el sistema está sujeto.
- Si el sistema es estático, como en (3), se tiene:

$$\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1} = \boldsymbol{\theta}.$$

Modelo de regresión dinámico (cont.)

- En los modelos dinámicos, se tiene una forma más general

$$\boldsymbol{\theta}_t = \mathbf{G}_{t-1}\boldsymbol{\theta}_{t-1} + \boldsymbol{w}_t \quad (7)$$

donde \mathbf{G}_t es una matriz de valores conocidos y valores \boldsymbol{w}_t es un error aleatorio. La ecuación (7) es la **ecuación del sistema**.

- La **matriz de evolución** \mathbf{G}_t controla la parte determinística de la evolución del sistema y establece la propagación del sistema a lo largo del tiempo.
- La **perturbación** \boldsymbol{w}_t es responsable por la introducción de la incertidumbre debido al paso del tiempo y la consecuente pérdida de información.
- Note que si $\mathbf{G}_t = \mathbf{I}$ y $\boldsymbol{w}_t = 0$, el modelo se reduce al caso estático.

Modelo de regresión dinámico (cont.)

Considere el siguiente ejemplo:

- Se tiene una serie de ventas (Y) explicada por la respectiva serie de precios (X) a través de una relación estable:

Ejemplo

$$Y_t = \mu_t + \beta_t X_t + v_t \quad (8a)$$

$$\mu_t = \mu_{t-1} + \Delta\mu_t \quad (8b)$$

$$\beta_t = \beta_{t-1} + \Delta\beta_t \quad (8c)$$

- Las **ventas** son explicadas por los **precios** en una **regresión dinámica**.
- La mayor (o menor) estabilidad de la relación es controlada por la magnitud de los incrementos $\Delta\mu_t$ y $\Delta\beta_t$.

Modelo lineal dinámico normal: definición

Definición (West and Harrison (1997))

El modelo lineal dinámico normal general (MLD) é dado por

$$\{F, G, V, W\}_t = \{F_t, G_t, V_t, W_t\}$$

para cada tiempo t , donde

- 1 F_t es una matriz $n \times r$ conocida;
- 2 G_t es una matriz $n \times n$ conocida;
- 3 V_t es una matriz de varianza $r \times r$ conocida;
- 4 W_t es una matriz de varianza $n \times n$ conocida.

Modelo lineal dinámico normal (cont.)

Un **modelo lineal dinámico normal** puede ser definido como:

$$\text{Ecuación de la observación} \quad : \quad \mathbf{Y}_t = \mathbf{F}'_t \boldsymbol{\theta}_t + \nu_t, \quad \nu_t \sim N[0, V_t]$$

$$\text{Ecuación del sistema} \quad : \quad \boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \mathbf{w}_t, \quad \mathbf{w}_t \sim N[\mathbf{0}, \mathbf{W}_t]$$

donde

\mathbf{F}_t : Matriz de diseño.

\mathbf{G}_t : Matriz de evolución.

$\boldsymbol{\theta}_t$: Vector de estados.

\mathbf{W}_t : Matriz de varianza de $\boldsymbol{\theta}_t$.

$\boldsymbol{\mu}_t = \mathbf{F}'_t \boldsymbol{\theta}_t$: Respuesta media o nivel;

ν_t : Error de observación.

Modelo lineal dinámico normal (cont.)

Un **modelo lineal dinámico normal** puede ser representado como:

$$Y_t = F_t' \theta_t + \nu_t, \quad \nu_t \sim N[0, V_t] \quad (9a)$$

$$\theta_t = G_t \theta_{t-1} + w_t, \quad w_t \sim N[0, W_t] \quad (9b)$$

o, equivalentemente,

$$(Y_t | \theta_t) \sim N[F_t \theta_t, V_t] \quad (10a)$$

$$(\theta_t | \theta_{t-1}) \sim N[G_t \theta_{t-1}, W_t] \quad (10b)$$

En el ejemplo anterior, $Y_t = \text{venta}$, $X_t = \text{precio}$, $F_t' = (1, X_t)$

$$\theta = \begin{pmatrix} \mu \\ \beta \end{pmatrix}, \quad G_t = I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad y \quad w_t = \begin{pmatrix} \Delta\mu \\ \Delta\beta \end{pmatrix}$$

Modelos de primer y segundo orden

Modelo polinomial de primer orden

- El más simple de los modelos dinámicos es el **modelo de tendencia estable o de primer orden**. Es compuesto de apenas un **nivel que varía según un paseo aleatorio**:

Modelo polinomial de primer orden

$$Y_t = \mu_t + v_t, \quad v_t \sim N(0, V_t) \quad (11a)$$

$$\mu_t = \mu_{t-1} + w_t, \quad w_t \sim N(0, W_t) \quad (11b)$$

- Según este modelo, el **nivel es localmente constante**, mas varía cuando se consideran periodos largos de tiempo.
- Usualmente, la variación de las observaciones en torno de los niveles (medida por V_t) es mayor que las variaciones temporales del nivel a lo largo del tiempo (medidas por W_t), i.e. $V_t > W_t$.
- En la notación del modelo dinámico tenemos: $F_t = 1$ y $G_t = 1$.

Modelo polinomial de segundo orden

- El **modelo polinomial de segundo orden** o modelo de **tendencia lineal** permite un **crecimiento en el nivel**:

Modelo polinomial de segundo orden

$$y_t = \mu_t + v_t \quad (12a)$$

$$\mu_t = \mu_{t-1} + \beta_{t-1} + w_{1t} \quad (12b)$$

$$\beta_t = \beta_{t-1} + w_{2t} \quad (12c)$$

- Ese modelo es obtenido con $F_t = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ y $G_t = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$
- Aquí, el nivel permanece localmente lineal, pero la forma de la recta puede variar con el tiempo.
- Tendencias de mayor orden son poco utilizadas.*

Análisis o inferencia secuencial

Análisis o inferencia secuencial

Sea $D_t = \{Y_1, \dots, Y_t\}$ la información disponible hasta el tiempo t .

$$\dots (\theta_{t-1}|D_{t-1}) \xrightarrow[\text{Ec.Sistema}]{\text{Evolución}} (\theta_t|D_{t-1}) \xrightarrow[\text{Teo.Bayes}]{\text{Actualización}} (\theta_t|D_t) \dots$$

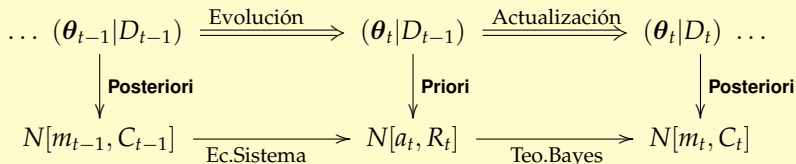
- La **ecuación del sistema** informa como a partir de la **posteriori de ayer** podemos llegar a la **priori de hoy**.
- El **teorema de Bayes** enseña como combinar la **priori de hoy** con la información que acabamos de obtener para llegar a la **posteriori de hoy**. Para días posteriores, el ciclo se repite.

Análisis Secuencial (cont.)

$$(Y_t | \boldsymbol{\theta}_t) \sim N[\mathbf{F}_t \boldsymbol{\theta}_t, \mathbf{V}_t]$$

$$(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1}) \sim N[\mathbf{G}_t \boldsymbol{\theta}_{t-1}, \mathbf{W}_t]$$

$$(\boldsymbol{\theta}_0 | D_0) \sim N[m_0, C_0]$$



$$a_t = G_t m_{t-1},$$

$$m_t = a_t + A_t e_t,$$

$$R_t = G_t C_{t-1} G_t' + W_t,$$

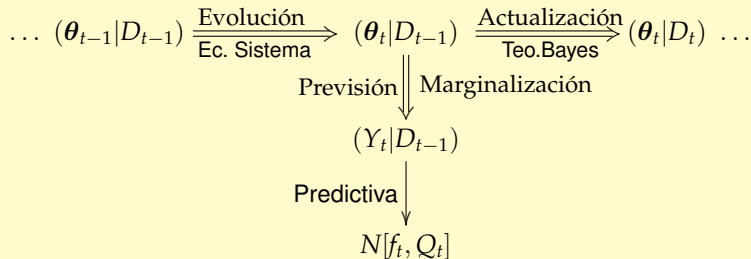
$$A_t = R_t F_t / Q_t,$$

$$e_t = Y_t - f_t,$$

$$C_t = R_t - A_t A_t' Q_t$$

Previsión

Las previsiones a partir de un modelo dinámico son obtenidas por la combinación de información a **priori sobre el estado** con la **ecuación de la observación**.



donde $f_t = F_t' a_t$ y $Q_t = F_t' R_t F_t + V_t$.

Previsión (cont.)

- La combinación de $Y_t = F'_t \theta_t + v_t$ con la priori $p(\theta_t | D_{t-1})$ permite la obtención de la **distribución predictiva**:

$$p(Y_t | D_{t-1}) = \int p(Y_t | \theta_t) p(\theta_t | D_{t-1}) d\theta_t$$

a partir de aquí son realizadas las previsiones.

- Previsión puntual**: se puede tomar la media de esa distribución, $E[Y_t | D_{t-1}]$.
- Intervalo de previsión** de 90 % de probabilidad: basta tomar A de forma que

$$0,90 = Pr(Y_t \in A | D_{t-1}) = \int_A p(Y_t | D_{t-1}) dy_t$$

DLM univariado: varianza V_t conocida

- (a) Ec. de observación: $Y_t = F_t' \theta_t + \nu_t$ $\nu_t \sim N[0, V_t]$
- (b) Ec. del sistema: $\theta_t = G_t \theta_{t-1} + \omega_t$ $\omega_t \sim N[0, W_t]$
- (c) Información: $(\theta_{t-1} | D_{t-1}) \sim N[m_{t-1}, C_{t-1}]$
- (d) Priori: $(\theta_t | D_{t-1}) \sim N[a_t, R_t]$ $a_t = G_t m_{t-1}$
 $R_t = G_t C_{t-1} G_t' + W_t$
- (e) Previsión: $(Y_t | D_{t-1}) \sim N[f_t, Q_t]$ $f_t = F_t' a_t$
 $Q_t = F_t' R_t F_t + V_t$
- (f) Actualización: $(\theta_t | D_t) \sim N[m_t, C_t]$ $m_t = a_t + A_t e_t$
 $C_t = R_t - A_t A_t' Q_t$
 $e_t = Y_t - f_t$
 $A_t = R_t F_t / Q_t$

Distribuciones de Previsión

- La **previsión para varios pasos adelante** es realizada de forma similar.
- Para prever Y_{t+k} en el tiempo $t - 1$, la ecuación del sistema se utiliza sucesivamente hasta escribir θ_{t+k} como función de θ_{t-1} .
- Por ejemplo, combinando

$$\theta_{t-1} = \mathbf{G}_{t+1}\theta_t + \mathbf{w}_{t+1} \quad \text{con} \quad \theta_t = \mathbf{G}_t\theta_{t-1} + \mathbf{w}_t$$

se obtiene:

$$\theta_{t+1} = \mathbf{G}_{t+1}[\mathbf{G}_t\theta_{t-1} + \mathbf{w}_t] + \mathbf{w}_{t+1} = \mathbf{G}_{t+1}\mathbf{G}_t\theta_{t-1} + \mathbf{G}_{t+1}\mathbf{w}_t + \mathbf{w}_{t+1}$$

Luego, se combina con la ecuación de la observación en el tiempo $t + k$ para obtener la distribución predictiva $p(Y_{t+k}|D_{t-1})$.

Distribuciones de Previsión (cont.)

Definición (Función de previsión)

La función de previsión $f_t(k)$ está dada por:

$$f_t(k) = E[\mu_{t+k}|D_t] = E[\mathbf{F}'_{t+k}\boldsymbol{\theta}_{t+k}|D_t], \quad k \geq 0 \quad (13)$$

donde

$$\mu_{t+k} = \mathbf{F}'_{t+k}\boldsymbol{\theta}_{t+k}$$

es la función respuesta media.

Distribuciones de Previsión (cont.)

DLM univariado: varianza V_t conocida

Teorema

Para todo $k \geq 1$, las distribuciones predictivas k pasos adelante están dadas por

$$(\theta_{t+k}|D_t) \sim N[a_t(k), R_t(k)]$$

$$(Y_{t+k}|D_t) \sim N[f_t(k), Q_t(k)].$$

Los momentos son definidos recursivamente por

$$R_t(k) = G_{t+k}R_t(k-1)G'_{t+k} + W_{t+k}$$

$$a_t(k) = G_{t+k}a_t(k-1)$$

$$f_t(k) = F'_{t+k}a_t(k)$$

$$Q_t(k) = F'_{t+k}R_t(k)F_{t+k} + V_{t+k}$$

con valores iniciales dados por $a_t(0) = m_t$ y $R_t(0) = C_t$.

Modelo Polinomial de Primer Orden - Ecuaciones

$$\begin{aligned}
 Y_t &= \mu_t + \nu_t, & \nu_t &\sim N[0, V_t] & \equiv & Y_t | \mu_t \sim N[\mu_t, V_t] \\
 \mu_t &= \mu_{t-1} + \omega_t, & \omega_t &\sim N[0, W_t] & \equiv & \mu_t | \mu_{t-1} \sim N[\mu_{t-1}, W_t]
 \end{aligned}$$

Ecuaciones de actualización para $\{1, 1, V_t, W_t\}$

(*) Información inicial	$(\mu_0 D_0) \sim N[m_0, C_0]$	
(a) Posteriori para μ_{t-1}	$(\mu_{t-1} D_{t-1}) \sim N[m_{t-1}, C_{t-1}]$	
(b) Priori para μ_t	$(\mu_t D_{t-1}) \sim N[m_{t-1}, R_t]$	$R_t = C_{t-1} + W_t$
(c) Previsión para un paso	$(y_t D_{t-1}) \sim N[f_t, Q_t]$	$f_t = m_{t-1}$ $Q_t = R_t + V_t$
(d) Posteriori para μ_t	$(\mu_t D_t) \sim N[m_t, C_t]$	$m_t = m_{t-1} + A_t e_t$
	$C_t = A_t V_t$	
	$A_t = R_t / Q_t$	
	$e_t = Y_t - f_t$	

Modelo de Primer Orden - Ecuaciones (cont.)

- Función de previsión:

$$f_t(k) = E[Y_{t+k}|D_t] = E[\mu_t|D_t] = m_t, \quad \forall k > 0$$

pues $E[Y_{t+k}|\mu_t] = E[\mu_{t+k}|\mu_t] = \mu_t$

- e_t : error de previsión un paso adelante $e_t = y_t - f_t = y_t - m_{t-1}$
- $A_t = \frac{R_t}{Q_t} = \frac{R_t}{R_t + V_t} = \frac{C_{t-1} + W_t}{C_{t-1} + W_t + V_t}$
- $A_t : \rho_t^2(y_t, \mu_t)$ o $\hat{\beta}$ en $\mu_t = \alpha + \beta y_t$ $0 \leq A_t \leq 1$
- $m_t = m_{t-1} + \rho_t^2(y_t - m_{t-1}) = A_t Y_t + (1 - A_t)m_{t-1}$
- Previsión k pasos adelante:

$$(Y_{t+k}|D_t) \sim N[m_t, Q_t(k)]$$

$$Q_t(k) = C_t + \sum_{j=1}^k W_{t+j} + V_{t+k}$$

DLM: varianza desconocida $V_t = k_t \phi^{-1}$, k_t conocido

- (a) Ec. Observación: $Y_t = F_t' \theta_t + \nu_t$ $\nu_t \sim N[0, k_t \phi^{-1}]$
- (b) Ec. Sistema: $\theta_t = G_t \theta_{t-1} + \omega_t$ $\omega_t \sim t_{n_{t-1}}[0, W_t]$
- (c) Información: $(\theta_{t-1} | D_{t-1}) \sim t_{n_{t-1}}[m_{t-1}, C_{t-1}]$
 $(\phi_{t-1} | D_{t-1}) \sim G\left[\frac{n_{t-1}}{2}, \frac{d_{t-1}}{2}\right]$
- (d) Priori: $(\theta_t | D_{t-1}) \sim t_{\delta_t n_{t-1}}[a_t, R_t]$ $a_t = G_t m_{t-1}$
 $R_t = G_t C_{t-1} G_t' + W_t$
 $(\phi_t | D_{t-1}) \sim G\left[\frac{\delta_t n_{t-1}}{2}, \delta_t \frac{d_{t-1}}{2}\right]$
- (e) Previsión: $(Y_t | D_{t-1}) \sim t_{\delta_t n_{t-1}}[f_t, Q_t]$ $f_t = F_t' a_t$
 $Q_t = F_t' R_t F_t + k_t S_{t-1}$

DLM: varianza desconocida $V_t = k_t \phi^{-1}$, k_t conocido

(f) Actualización:

$$(\theta_t | D_t) \sim t_{n_t} [m_t, C_t]$$

$$(\phi_t | D_t) \sim G\left[\frac{n_t}{2}, \frac{d_t}{2}\right]$$

$$m_t = a_t + A_t e_t$$

$$C_t = (S_t / S_{t-1}) R_t - A_t A_t' Q_t$$

$$e_t = Y_t - f_t, A_t = R_t F_t / Q_t$$

$$n_t = \delta_t n_{t-1} + 1, d_t = \delta_t d_{t-1} + S_{t-1} e_t^2$$

$$S_t = d_t / n_t$$

(g) Predictivas para $k \geq 1$

$$(\theta_{t+k} | D_t) \sim t_{\delta_t n_t} [a_t(k), R_t(k)]$$

$$(y_{t+k} | D_t) \sim t_{\delta_t n_t} [f_t(k), Q_t(k)]$$

$$a_t(k) = G_{t+k} a_t(k-1)$$

$$R_t(k) = G_{t+k} R_t(k-1) G_{t+k}' + W_{t+k}$$

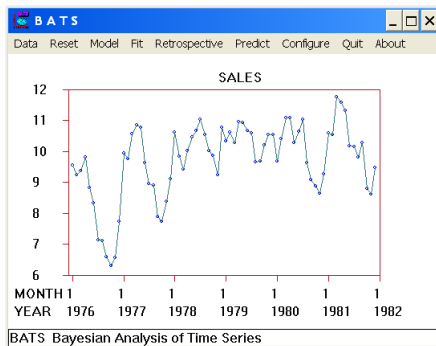
$$f_t(k) = F_{t+k}' a_t(k)$$

$$Q_t(k) = F_{t+k}' R_t(k) F_{t+k} + k_{t+k} S_t$$

$$a_t(0) = m_t, \quad R_t(0) = C_t$$

Ejemplo: Candy

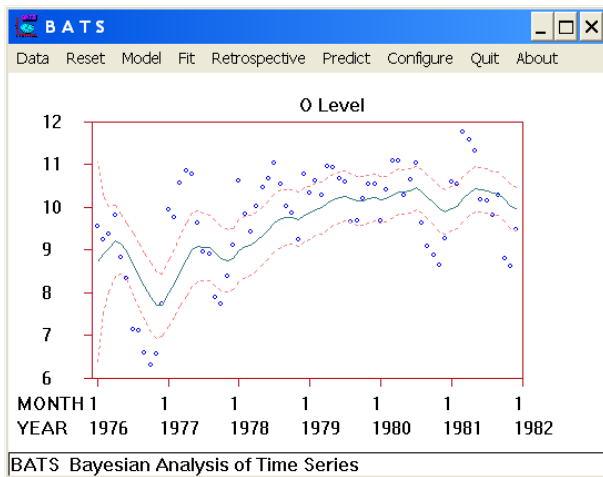
- Los datos de este ejemplo corresponden a las ventas de un caramelo (SALES) del archivo CANDY.DAT, en el software BATS.



- En los siguientes slides se presentan algunos resultados del ajuste de modelos polinomiales de primer y segundo orden a la serie de ventas mensuales de Enero de 1976 a diciembre de 1981.

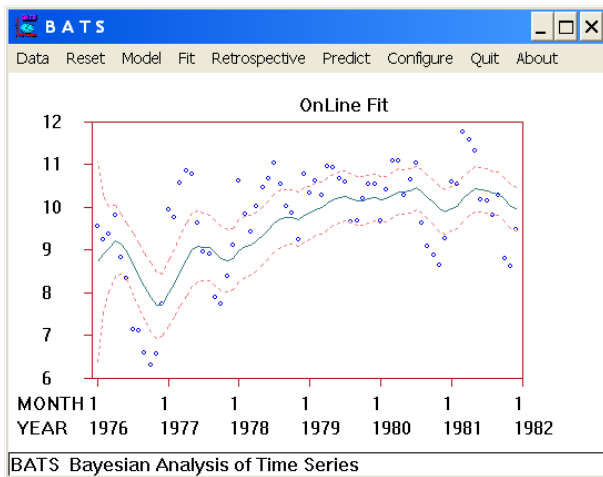
Candy: Modelo polinomial de primer orden

Figura: Estimación del nivel



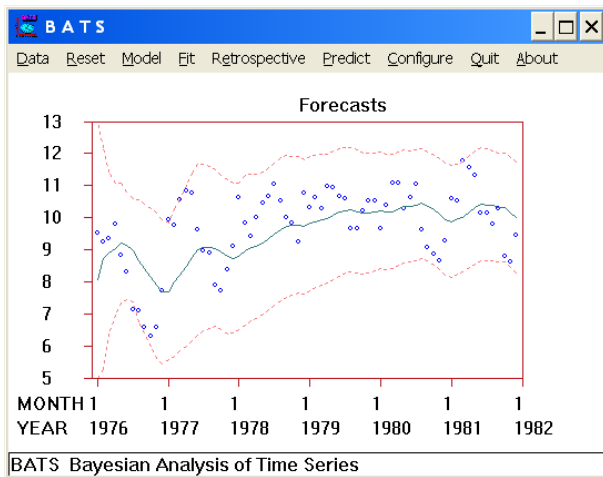
Candy: Modelo de primer orden (cont.)

Figura: Estimación de las ventas



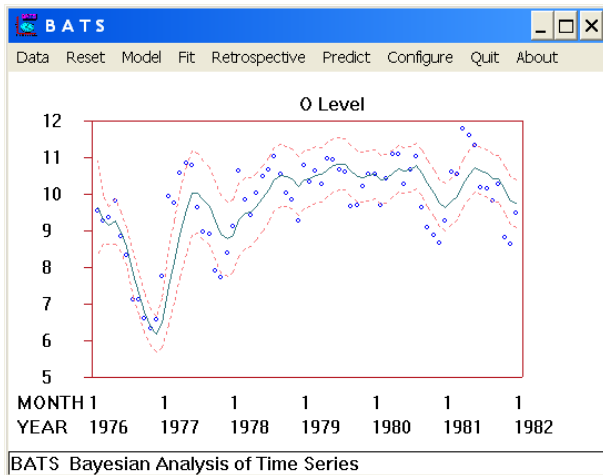
Candy: Modelo de primer orden (cont.)

Figura: Previsión un paso adelante



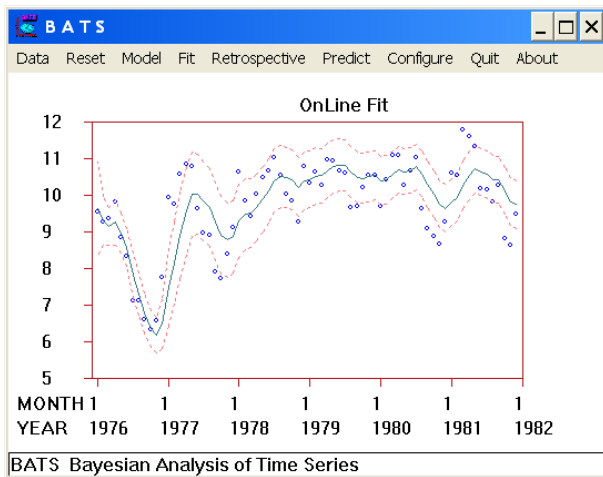
Candy: Modelo de segundo orden (crecimiento)

Figura: Estimación del nivel



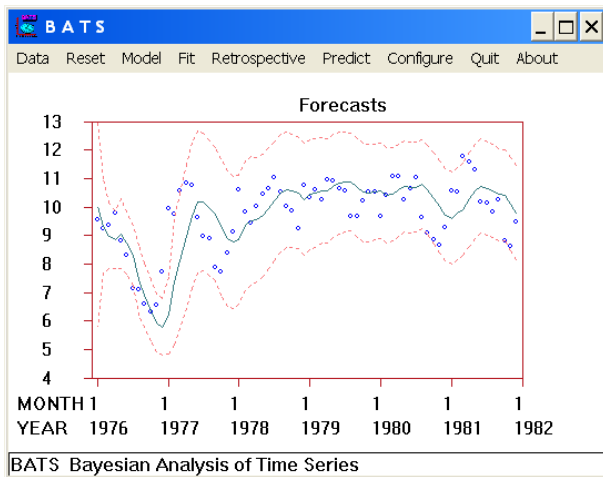
Candy: Modelo de segundo orden (cont.)

Figura: Estimación de las ventas



Candy: Modelo de segundo orden (cont.)

Figura: Previsión un paso adelante



Evolución del sistema

- Aquí discutiremos el procesamiento de la incertidumbre relativa a W .
- La inferencia bayesiana para parámetros desconocidos siempre es la misma: **actualización via teorema de Bayes**.
- El tratamiento dado a la varianza de las observaciones es analítico. Lo mismo no puede ser realizado para la varianza del sistema.

Felizmente, existe una solución con base en **factores de descuento** que produce una alternativa aceptable.

- El **valor de la información disminuye con el tiempo**.
- Esa disminución es controlada por la evolución del sistema, a través del aumento de la incertidumbre del sistema.

Evolución (cont.)

- En el modelo estable se tiene:

$$V[\mu_t|D_{t-1}] = V[\mu_{t-1}|D_{t-1}] + \mathbf{W}_t \quad (14)$$

- Como la **varianza mide la dispersión del sistema**, su inverso, la **precisión mide la información del sistema**. Cuanto más dispersa sea la variable, mayor será su varianza e menor su precisión.
- Pensando en **porcentaje o fracción de información perdida con el paso del tiempo**, se puede definir un **factor de descuento** $\delta \in (0, 1]$, tal que

$$V^{-1}[\mu_t|D_{t-1}] = \delta V^{-1}[\mu_{t-1}|D_{t-1}] \quad (15)$$

(14) y (15) forman una base para especificar \mathbf{W}_t .

Factor de descuento

El factor de descuento δ es el porcentaje de información que pasa de un periodo a otro.

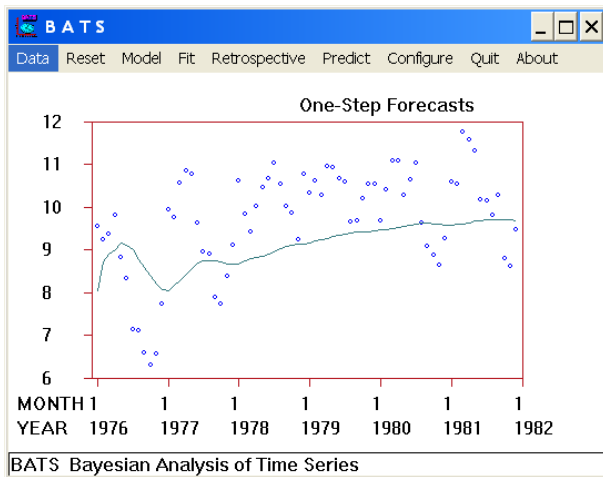
- Valores típicos para sistemas sin variaciones bruscas son $> 90\%$
- La selección de valor adecuado depende de la aplicación, se sugiere que se comparen algunos valores.
- Valores muy próximos (parecidos) no producen diferencias perceptibles.
- Valores muy **bajos** ($< 0,8$) tienden a introducir **mucha incertidumbre** y producen límites de confianza para previsión muy grandes.
- Valores muy **altos** representan un sistema con **cambios muy suaves**.
- En el límite, cuando $\delta = 1$, se tiene el modelo estático donde no hay pérdida de información.
- La misma idea de descuento puede ser extendida para modelos más generales con varios descuentos aplicados a partes diferentes del modelo (superposición de modelos).

Aplicação: Candy

- A seguir se ilustra el uso de factores de descuento con la serie de ventas de un caramelo (SALES) del archivo CANDY.DAT del software BATS.
- Las siguientes figuras presentan las previsiones un paso adelante resultantes del ajuste de un modelo de tendencia constante con los siguientes factores de descuento: 1,0; 0,9 y 0,8.

Candy: Descuento

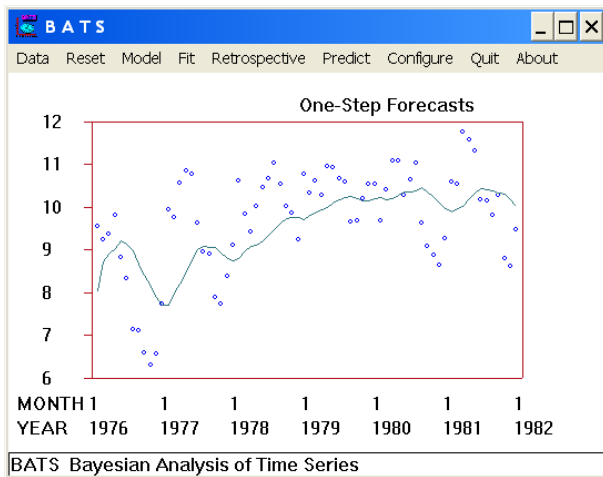
Figura: Factor de descuento $\delta = 1,0$ (estructura estática)



BATS: Trend: Constant, Discount:Trend=1.00,Variance=0.99

Candy: Descuento (cont.)

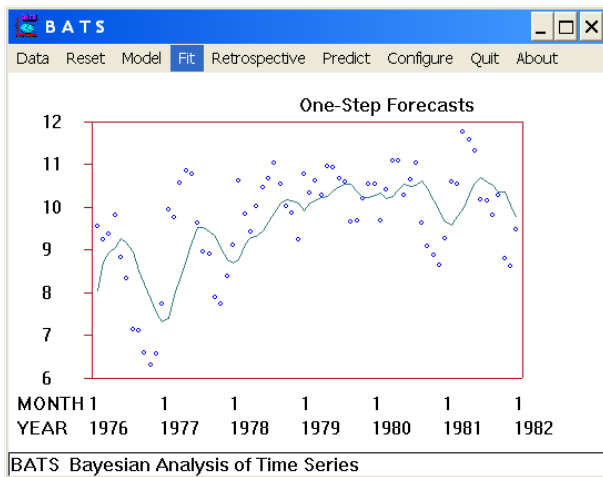
Figura: Factor de descuento $\delta = 0,9$



BATS: Trend: Constant, Discount:Trend=0.90,Variance=0.99

Candy: Descuento (cont.)

Figura: Factor de descuento $\delta = 0,8$



BATS: Trend: Constant, Discount:Trend=0.80,Variance=0.99

Análisis retrospectivo o suavización

Análisis retrospectivo

- En el análisis retrospectivo se usa toda la serie observada para re-evaluar la inferencia realizada durante el procedimiento secuencial. Esta re-evaluación se debe al **uso de observaciones obtenidas después del periodo de interés**.
- Con más información, sabemos más y disponemos de más instrumentos para entender el pasado.
- Esta operación de paso de información para el pasado es conocida como **suavización o análisis retrospectivo**.
- A partir del análisis secuencial se tiene $p(\theta_t|D_t)$. Si se tiene información hasta el tiempo $t + k$, la mejor descripción sobre la incertidumbre asociada a θ_t es a través de $p(\theta_t|D_{t+k})$.

Distribuciones suavizadas para $1 \leq k \leq t$

Modelo con varianza conocida

$$\begin{aligned}
 (\theta_{t-k}|D_t) &\sim N[a_t(-k), R_t(-k)] & a_t(-k) &= m_{t-k} - B_{t-k}[a_{t-k+1} - a_t(-k+1)] \\
 & & R_t(-k) &= C_{t-k} - B_{t-k}[R_{t-k+1} - R_t(-k+1)]B'_{t-k} \\
 & & B_t &= C_t G'_{t+1} R_{t+1}^{-1}
 \end{aligned}$$

Modelo con varianza desconocida

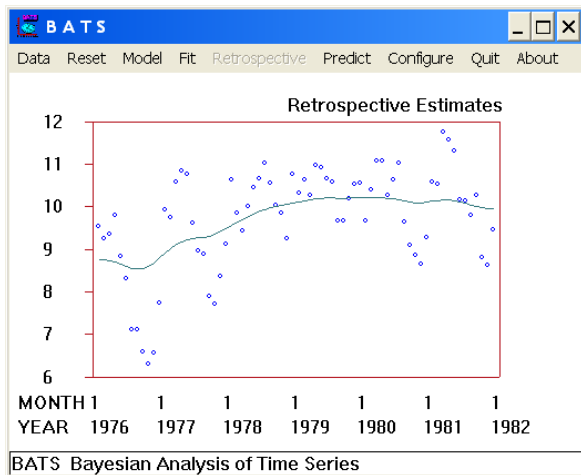
$$\begin{aligned}
 (\theta_{t-k}|D_t) &\sim t_{n_t(-k)}[a_t(-k), R_t(-k)] & a_t(-k) &= m_{t-k} - B_{t-k}[a_{t-k+1} - a_t(-k+1)] \\
 (\phi_{t-k}|D_t) &\sim G[n_t(-k)/2, d_t(-k)/2] & R_t(-k) &= C_{t-k} - B_{t-k}[R_{t-k+1} - R_t(-k+1)]B'_{t-k} \\
 & & B_t &= C_t G'_{t+1} R_{t+1}^{-1} \\
 & & n_t(-k) &= n_{t-k} + \delta_{t-k+1}(n_t(-k+1) - \delta_{t-k+1}n_{t-k}) \\
 & & S_t^{-1} &= S_{t-k}^{-1} + \delta_{t-k+1}(S_t^{-1}(-k+1) - S_{t-k}^{-1}) \\
 & & d_t(-k) &= n_t(-k)S_t(-k)
 \end{aligned}$$

Aplicación: Candy

- A seguir el resultado del análisis retrospectivo realizado a la serie de ventas de un caramelo (SALES) del archivo CANDY.DAT del software BATS.
- Las siguientes figuras presentan las estimaciones suavizadas para las ventas y para el nivel obtenidas a partir del ajuste de un modelo polinomial de primer orden.

Candy: Análisis retrospectivo

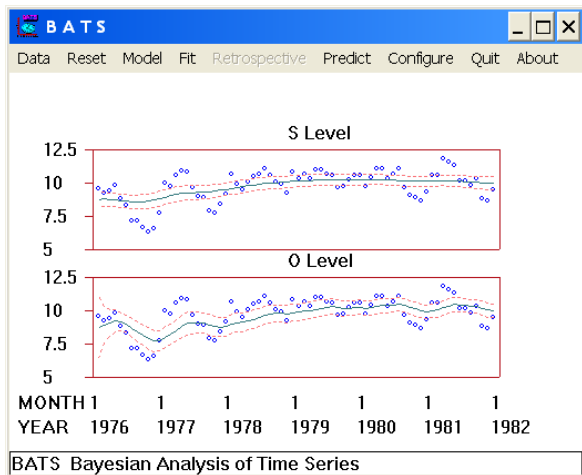
Figura: Estimaciones retrospectivas



BATS: Trend: Constant, Discount:Trend=0.90,Variance=0.99

Candy: Análisis retrospectivo (cont.)

Figura: Estimativas on-line y suavizada del nivel



Se observa que con la suavización se tienen trayectorias más suaves y límites de incertidumbre más

Parte III

Modelos Lineales Dinámicos Normales: Aspectos Prácticos

Modelos de regresión dinámicos

Modelo de regresión dinámico

- Si la serie de ventas, (Y_t) es explicada por la serie de precios (X_t) se tiene:

$$\begin{aligned} Y_t &= \mu_t + \beta_t X_t + v_t \\ \mu_t &= \mu_{t-1} + w_{1t} \\ \beta_t &= \beta_{t-1} + w_{2t} \end{aligned}$$

- Nada impide que otras variables sean incluidas como regresoras. En el caso de una serie Y_t con variables explicativas X_{1t}, \dots, X_{pt} , se tiene

$$\begin{aligned} Y_t &= \beta_{0t} + \beta_{1t} X_{1t} + \dots + \beta_{pt} X_{pt} + v_t \\ \beta_{it} &= \beta_{i,t-1} + w_{it}; \quad i = 0, 1, \dots, p \end{aligned}$$

conocido como **modelo de regresión dinámica**

- En la notación de modelos dinámico: $F'_t = (1, X_{1t}, \dots, X_{pt})$, $G_t = I_{p+1}$, la matriz identidad de orden $p + 1$ y $w'_t = (w_{01}, w_{1t}, \dots, w_{pt})$.

Modelo de regresión dinámico (cont.)

Modelo de regresión dinámico

$$\text{Ec. observación :} \quad Y_t = \mathbf{F}'_t \boldsymbol{\theta}_t + \nu_t, \quad \nu_t \sim N[0, V_t]$$

$$\text{Ec. sistema :} \quad \boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1} + \omega_t, \quad \omega_t \sim N[\mathbf{0}, \mathbf{W}_t]$$

donde

$\mathbf{F}_t = (X_{t1}, \dots, X_{tn})'$: vector de regresoras

X_1, \dots, X_n : variables independientes

X_{ti} : valor de la i -ésima variable X en t

$\boldsymbol{\theta}_t$: $n \times 1$ vector de parámetros de la regresión

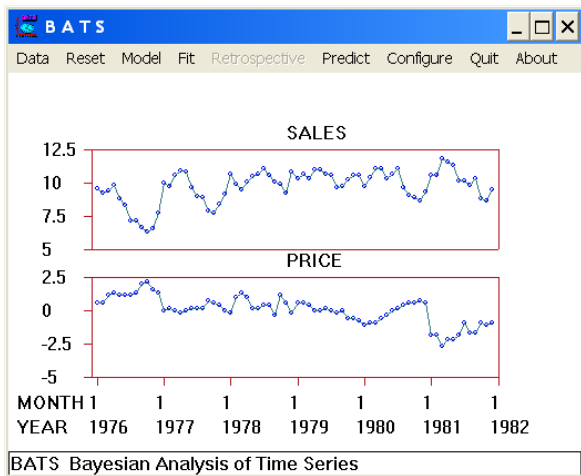
ω_t : matriz de la varianza de $\boldsymbol{\theta}_t$.

Aplicación: Candy

- Los datos de este ejemplo corresponden a las ventas y precios mensuales de un caramelo, del archivo CANDY.DAT del software BATS.
- Se espera que la serie de ventas (SALES) este relacionada con la serie de precios (PRICE).
- En las figuras que siguen se presentan algunos de los resultados del ajuste de un modelo de regresión dinámica con tendencia estable, utilizando el precio como variable explicativa.

Candy: Regresión dinámica

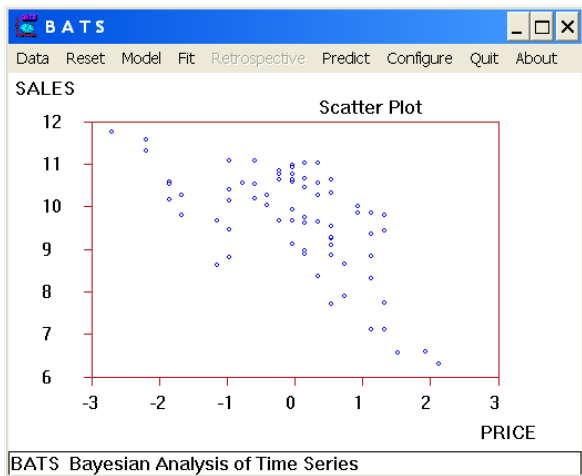
Figura: MPlot de Ventas y Precios



Se observa un aparente movimiento común de las 2 series.

Candy: Regresión dinámica (cont.)

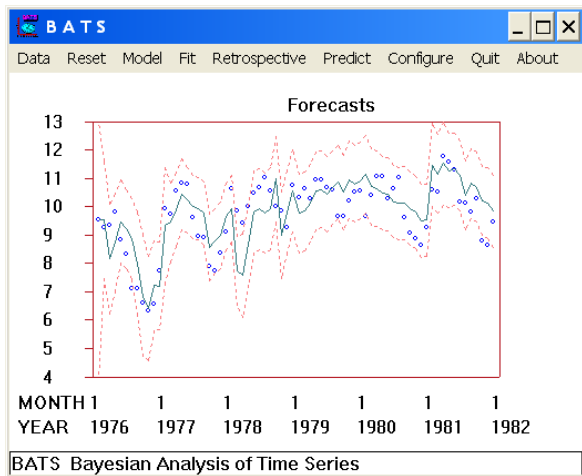
Figura: X-Y Plot de Ventas y Precios



El gráfico ayuda a evidenciar la relación que existe entre ambas series. (Correlación=-0.63)

Candy: Regresión dinámica (cont.)

Figura: Previsión

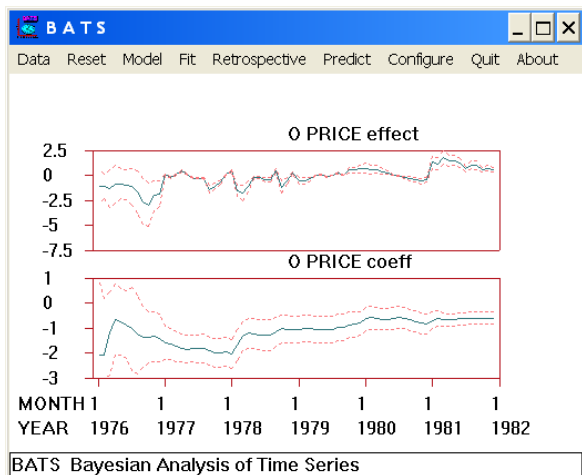


Las previsiones resultan considerablemente mejores en comparación con los modelos sin la variable precio.

BATS:
Trend: CONSTANT
Discount: Trend=0.90,
Regression=0.98,
Variance=0.99

Candy: Regresión dinámica (cont.)

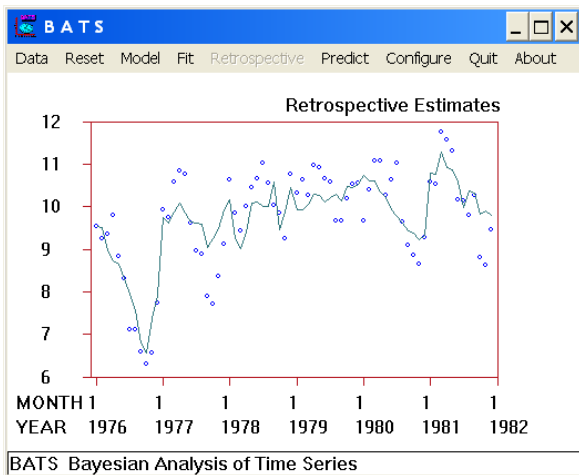
Figura: Coeficiente y efecto de los precios



Este gráfico muestra las trayectorias de $E[\beta_t|D_t]$ y de $E[\beta_t|D_t] \times X_t$.

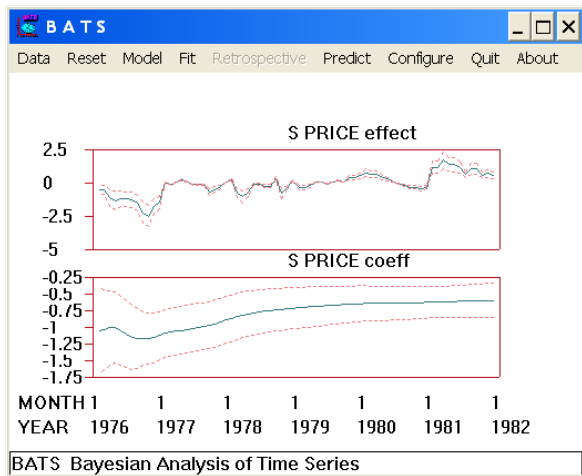
Candy: Regresión dinámica (cont.)

Figura: Estimaciones retrospectivas para ventas



Candy: Regresión dinámica (cont.)

Figura: Coeficiente y efecto de precio suavizados



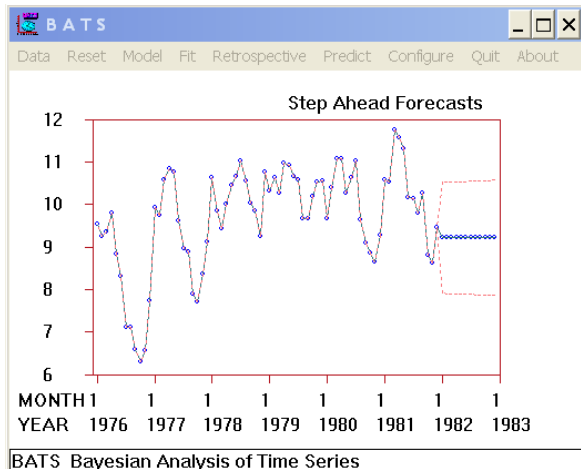
Este gráfico muestra las trayectorias de $E[\beta_t|D_t]$ y de $E[\beta_t|D_t] \times X_t$ después de la suavización. Se puede observar que el coeficiente de la regresión varía de -1 a $-0,6$. Ese movimiento es permitido por el MLD a través del δ .

Previsión con regresores

- Para hacer previsión con modelos de regresión dinámicos es necesario tener el valor de los regresores a lo largo del periodo de previsión.
- Normalmente, esos valores también son desconocidos y el tratamiento a ser utilizado es mucho más complicado.
- Una alternativa es hacer previsión bajo varios escenarios posibles.

Candy: Previsión con regresión dinámica

Figura: Previsión para el periodo 1982/1 - 1982/12



En este caso, todos los valores de precio para el periodo de previsión son cero, por lo tanto, las previsiones tienen la forma constante del modelo de primer orden.

Modelos con estacionalidad

Estacionalidad: Forma libre

- Modelos estacionales **requieren un componente periódico** en el modelo.
- La representación más simple es a través de **factores o indicadores de cada periodo en el ciclo**. Para datos trimestrales, son usados cuatro indicadores.
- Una pequeña alteración implica el uso de efectos indicando la variación estacional en torno de un nivel. En este caso, los efectos tienen como restricción: sumar cero.
- Factores trimestrales de 100, 140, 80 y 120 equivalen a un nivel de 110 y efectos trimestrales de -10 , 30, -30 y 10.
- La ultima formulación es mas atrayente pues permite la separación entre estacionalidad y tendencia.
- La restricción debe ser mantenida en todas las afirmaciones probabilísticas, mas es fácilmente incorporable al método de inferencia utilizado.

Estacionalidad: Forma libre (cont.)

- Para datos trimestrales, los modelos dinámicos utilizan cuatro indicadores. El paso del tiempo hace que ellos pasen por una rotación. Así,

$$\theta_{t-1} = \begin{pmatrix} trim4 \\ trim1 \\ trim2 \\ trim3 \end{pmatrix}$$

- Esa rotación puede ser efectuada por la matriz de evolución

$$G = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

- El modelo es completado con una ecuación de observaciones que considera apenas el primer componente del vector paramétrico, es decir, $F'_t = (1, 0, 0, 0)$. La extensión para un ciclo de p periodos es análoga.

Estacionalidad: Armónicos

- Otra forma de modelar padrones cíclicos puede ser realizada usando **funciones trigonométricas**.
- La función $\cos(\omega(t - 1))$ es periódica con periodo $\frac{2\pi}{\omega}$. Si $\omega = \pi/6$, el periodo es 12 y el máximo ocurre para $t = 1$.
- Datos mensuales con ciclo anual pueden ser modelados via

$$Y_t = a_t \cos\left(\frac{\pi(t - 1)}{6}\right) + v_t$$

donde a_t es un parámetro que controla la amplitud y el máximo ocurre em *enero*. Observe la reducción drástica en la dimensión del vector paramétrico de 11 para 1.

Estacionalidad: Armónicos (cont.)

- Rezagos en el punto máximo del ciclo pueden ser acomodadas con un parámetro extra:

$$Y_t = a_t \cos\left(\frac{\pi(t-1)}{6}\right) + b_t \sin\left(\frac{\pi(t-1)}{6}\right) + v_t$$

La formulación dinámica de esta función armónica utiliza 2 parámetros,

$$F_t = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, G_t = \begin{pmatrix} \cos \omega & \sin \omega \\ -\sin \omega & \cos \omega \end{pmatrix}$$

donde $\omega = \pi/6$.

Esse modelo describe un ciclo segundo una función *coseno*.

Estacionalidad: Armónicos (cont.)

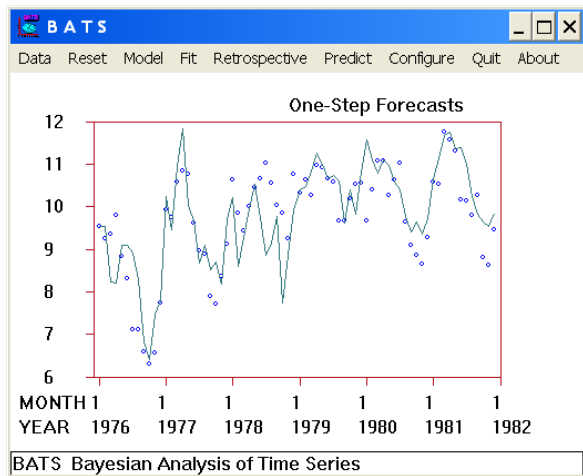
- Comportamientos cíclicos más complicados pueden ser modelados con la inclusión de formas armónicas de mayor frecuencia. La función $\cos(2\omega(t - 1))$ es similar, sin embargo completa 2 ciclos durante un periodo de tiempo $p = 2\pi/\omega$ o $\omega = 2\pi/p$.
- El resultado fundamental aquí informa que cualquier comportamiento cíclico de periodo p puede ser reproducido con la suma de, máximo, $p/2$ armónicos de periodos $p/j, j = 1, \dots, [p/2]$.
- La ventaja de este resultado reside en que podemos hacer **economía en el número de parámetros utilizados** y consecuentemente aumentar nuestra capacidad de aprendizaje sobre el sistema y mejorar nuestras previsiones.

Aplicación: Candy

- La serie de ventas del archivo CANDY.DAT del BATS exhibe un comportamiento cíclico que hasta ahora no ha sido tratado.
- En las figuras que siguen se presentan los resultados del análisis de la serie CANDY.DAT considerando un modelo de tendencia constante y una regresora (precio), incluyendo la componente estacional representada de forma libre y con armónicos.

Candy: Estacionalidad con forma libre

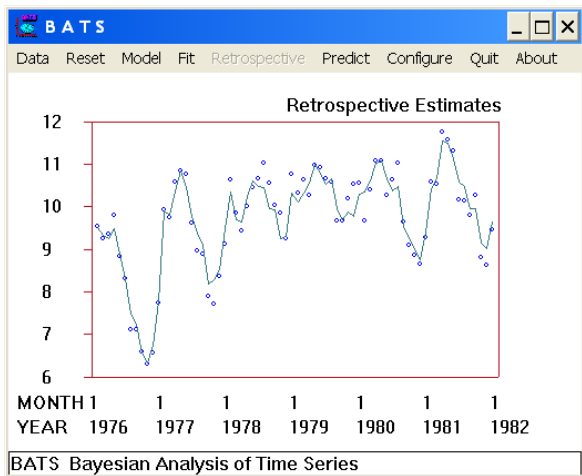
Figura: Previsión un paso adelante



BATS:
Trend: CONSTANT
Regression: PRICE
Seasonal: FREE-FORM

Candy: Estacionalidad con forma libre (cont.)

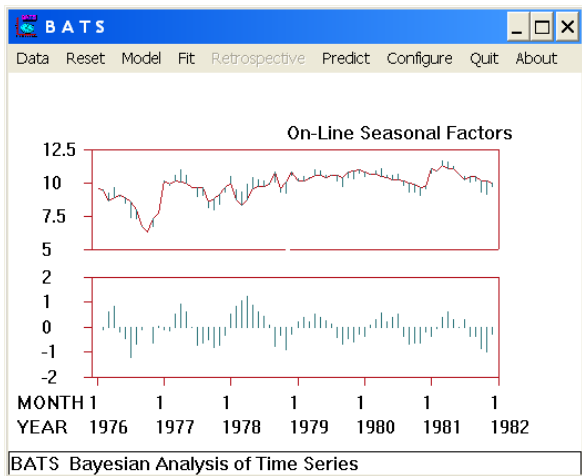
Figura: Estimación retrospectiva



BATS:
Trend: CONSTANT
Regression: PRICE
Seasonal: FREE-FORM

Candy: Estacionalidad con forma libre (cont.)

Figura: Estimación on-line de la estacionalidad

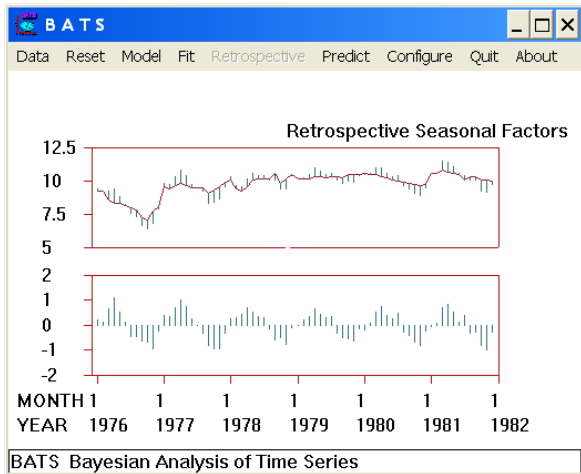


Este gráfico muestra las estimaciones **on-line** de los efectos estacionales en dos formas: *solos* y *sobrepuestos* al nivel estimado.

BATS:
Trend: CONSTANT
Regression: PRICE
Seasonal: FREE-FORM

Candy: Estacionalidad con forma libre (cont.)

Figura: Estimación suavizada de la estacionalidad

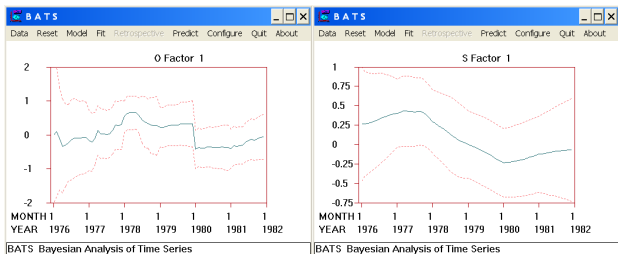


Este gráfico muestra las estimaciones **suavizadas** de los efectos estacionales en dos formas: *solos* y *sobrepuestos* al nivel estimado.

BATS:
Trend: CONSTANT
Regression: PRICE
Seasonal: FREE-FORM

Candy: Estacionalidad con forma libre (cont.)

Figura: Estimaciones on-line y suavizadas del efecto de enero

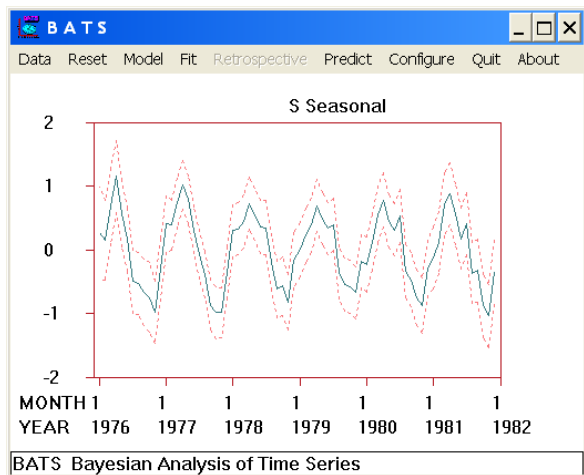


Este gráfico muestra claramente la variación del efecto de enero a lo largo del tiempo, resaltando la importancia de la modelación dinámica.

BATS:
 Trend:CONSTANT
 Regression:PRICE
 Seasonal:FREE-FORM

Candy: Estacionalidad con forma libre (cont.)

Figura: Estimación de la estacionalidad con límites de incertidumbre: forma libre



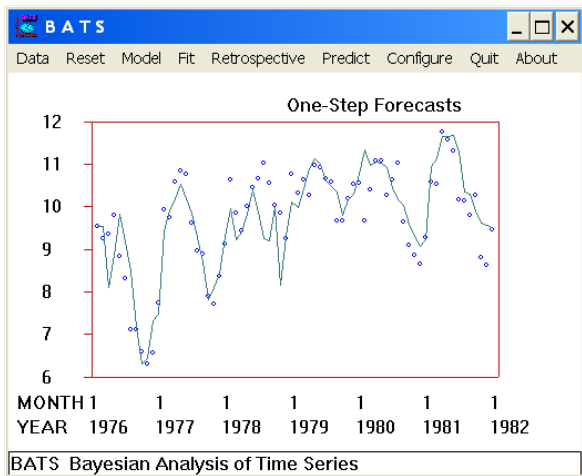
Esta figura muestra un patrón estacional bastante similar al de una función seno variando en el tiempo. Sugiere que la flexibilidad proporcionada por la modelación en forma libre puede no ser necesaria. Se pueden contemplar la posibilidad de reducción del tamaño del modelo a través del uso de formas armónicas para el componente estacional.

Candy: Modelación con armónicos

- Como los datos son mensuales, tenemos que el periodo es de tamaño 12, $p = 12$, habiendo por lo tanto, hasta 6 armónicos. El primero o fundamental, de periodo $12 = 12/1$; el segundo, de periodo $6 = 12/2 \dots$ hasta el último, de periodo $2 = 12/6$.
- La disminución del vector paramétrico es importante pues, aunque no altere las previsiones puntuales, disminuye la incertidumbre y facilita el monitoramiento de la performance del modelo. Además, un modelo mas parsimonioso acelera el tiempo de procesamiento.
- Esta forma de modelar la estacionalidad permite que el periodo de ciclo estacional sea diferente del periodo natural de la serie y que el modelo seleccionado contenga apenas el armónico fundamental, es decir, una única función senoidal.

Candy: Estacionalidad con 1^o armónico

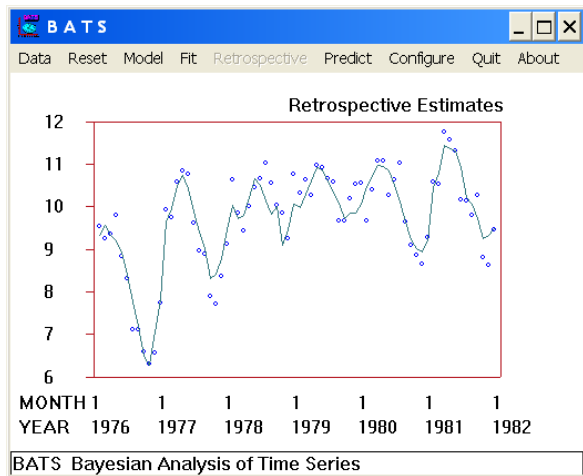
Figura: Previsión un paso adelante



BATS:
Trend: CONSTANT
Regression: PRICE
Seasonal: RESTRICTED-
HARMONICS

Candy: Estacionalidad con 1^o armónico (cont.)

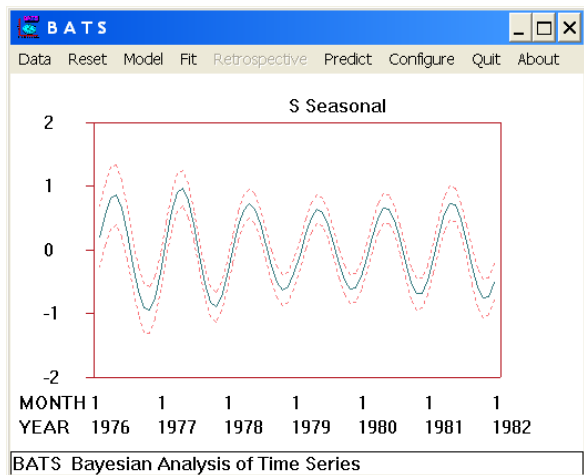
Figura: Estimaciones retrospectivas



BATS:
 Trend: CONSTANT
 Regression: PRICE
 Seasonal: RESTRICTED-
 HARMONICS

Candy: Estacionalidad con 1^o armónico (cont.)

Figura: Estimación de la estacionalidad con límites de incertidumbre: 1^o armónico

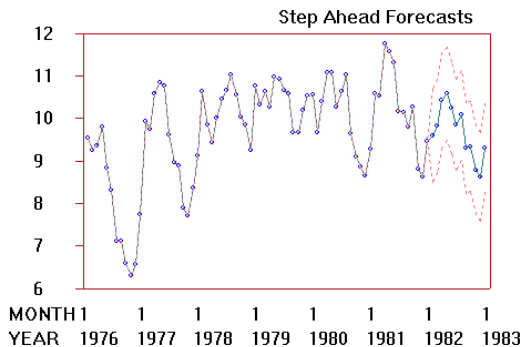


Las estimaciones son bastante parecidas con la modelación en forma libre.

BATS:
Trend: CONSTANT
Regression: PRICE
Seasonal: RESTRICTED-HARMONICS

Candy: Previsión incluyendo estacionalidad

Figura: Previsión con límites: Forma libre

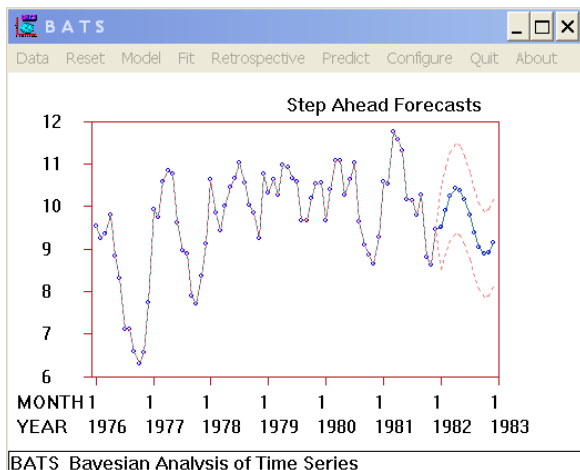


El horizonte de previsión es 1982/1 a 1982/12.
Aquí la regresora es igual a cero.

BATS Bayesian Analysis of Time Series

Candy: Previsión incluyendo estacionalidad (cont.)

Figura: Previsión con límites: 1^o armónico



De la comparación con la figura anterior, se concluye que las previsiones son similares mas tienen mayor incertidumbre en la modelación en forma libre.

Especificación de la Priori

- Hasta ahora, todos los análisis fueron realizados utilizando las distribuciones a priori del referencia del BATS. Nada impide el uso de prioris informativas.
- Para el componente estacional, sólo es necesaria la especificación de la incertidumbre respecto de los factores estacionales.
- En el caso de la modelación en forma libre, BATS se encarga de garantizar que la restricción, que la suma sea cero, sea respetada.
- En el caso de la modelación via armónicos, BATS trata de ajustar esa incertidumbre especificada por el usuario de la mejor forma posible aun cuando el modelo haya sido especificado apenas con algunos de los armónicos.

Superposición de modelos

Superposición de modelos

- Muchas series de tiempo exhiben un comportamiento bastante complejo. Es útil tener una forma general para estructurar y acomodar los varios componentes que intervienen en un modelo dinámico.
- Después de identificar las características mas importantes, debemos realizar la formulación del modelos.
- Considere, por ejemplo, la serie de accidentes: La tendencia global parece ser de una **variación suave del nivel**. La variación en torno del nivel tiene un **comportamiento cíclico**.

Entonces, fueron identificados dos componentes del modelo: un componente para la **tendencia y otro para la **estacionalidad**.**

Superposición de modelos (cont.)

- La estructura de los modelos dinámicos es apropiada pues permite que los componentes sean modelados separadamente y después integrados en un modelo.
- En el caso de dos componentes: tendencia y estacionalidad, estructuramos la ecuación de las observaciones con dos términos.

$$Y_t = Y_{Nt} + Y_{St} + v_t$$

cada uno de los términos es descrito a través de un modelo dinámico

$$\begin{aligned} Y_{Nt} &= F'_{Nt} \theta_{Nt} \\ \theta_{Nt} &= G_{Nt} \theta_{Nt-1} + w_{Nt} \end{aligned}$$

$$\begin{aligned} Y_{St} &= F'_{St} \theta_{St} \\ \theta_{St} &= G_{St} \theta_{St-1} + w_{St} \end{aligned}$$

Superposición de modelos (cont.)

Al integrar esos términos, se tiene la siguiente **ecuación de observación**

$$Y_t = \mathbf{F}'\boldsymbol{\theta}_t + v_t$$

$$\text{donde } \mathbf{F} = \begin{pmatrix} F_{Nt} \\ F_{St} \end{pmatrix} \text{ y } \boldsymbol{\theta} = \begin{pmatrix} \theta_{Nt} \\ \theta_{St} \end{pmatrix}$$

De forma similar, la **ecuación del sistema (integrado)** es

$$\boldsymbol{\theta}_t = \mathbf{G}\boldsymbol{\theta}_{t-1} + \mathbf{w}_t; \quad \mathbf{w}_t \sim N(0, \mathbf{W}_t)$$

$$\text{donde } \mathbf{G}_t = \begin{pmatrix} G_{Nt} & 0 \\ 0 & G_{St} \end{pmatrix} \text{ y } \mathbf{W}_t = \begin{pmatrix} W_{Nt} & 0 \\ 0 & W_{St} \end{pmatrix}$$

Superposición de modelos (cont.)

- Modelos con más componentes son construidos de la misma forma: cada término contribuye para la ecuación de observación y con un block de parámetros para la ecuación del sistema.
- La especificación de la varianza del sistema a través del método de los factores de descuento sigue el mismo procedimiento, es decir, es realizada en forma independiente para cada componente. En este caso, conjuntos de parámetros son formados teniendo en cuenta que tienen el mismo comportamiento en relación a la variación temporal.
- Ejemplo: en el modelo de ventas explicadas por el precio tenemos dos parámetros, μ_t y β_t (coeficiente de *precio*), que evolucionan según los descuentos δ_N y δ_P tales que

$$V^{-1}[\mu_t|D_{t-1}] = \delta_N V^{-1}[\mu_{t-1}|D_{t-1}]$$

$$V^{-1}[\beta_t|D_{t-1}] = \delta_S V^{-1}[\beta_{t-1}|D_{t-1}]$$

Monitoramiento

Monitoramiento

- Después de realizar la previsión y observar el valor correspondiente, podemos evaluar la precisión de un modelo.
- La idea básica es compararlo con algunas alternativas. Esta comparación puede servir para señalar acontecimientos inesperados.
- Como la previsión es con base en una distribución de probabilidad, cuanto más en la cola está la observación, más extrema e inesperada (para el modelo) es la observación. Esto puede ocurrir debido a una serie de motivos:
 - Un cambio pasajero y ocasional en la estructura de los datos.
 - Un cambio persistente y estructural;
 - Una deterioración de la performance del modelo.

Lo importante es que el sistema tenga capacidad de “sonar la alarma”

Monitoramiento (cont.)

Un esquema secuencial de monitoramiento funciona de la siguiente manera:

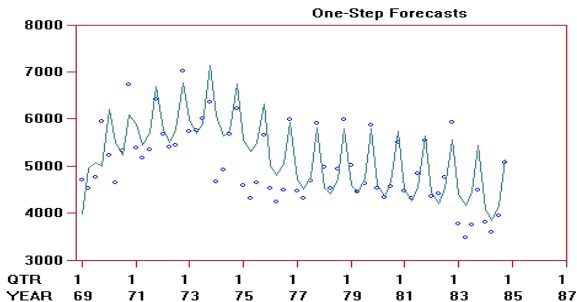
- 1 Cuando la observación cae mucho en la cola, “**suenan la alarma**”.
- 2 En ese momento, el predictor tendrá de **reflexionar** sobre la adecuación de su modelo y, si es necesario, modificarlo.
- 3 Para auxiliarlo, es útil tener una colección de **alternativas** y posibles **direcciones de acción**.

Aplicación: Accidentes graves

- Los datos de este ejemplo corresponden al número de accidentes de carretera graves (NUMBER) del archivo QACCIDS.DAT del software BATS.
- En la serie de accidentes se pueden notar 3 intervalos de tiempo distintos dentro de los cuales el comportamiento de la serie es estable. Una alternativa es analizar cada intervalo separadamente, mas estaríamos perdiendo información, por ejemplo, sobre la componente estacional. Eso porque el comportamiento estacional es estable a lo largo de toda la serie.
- En las figuras que siguen se presentan algunos resultados del ajuste de un modelo de tendencia lineal con componente estacional de forma libre utilizando la opción de monitoramiento del BATS.

Accidentes graves: monitoramiento

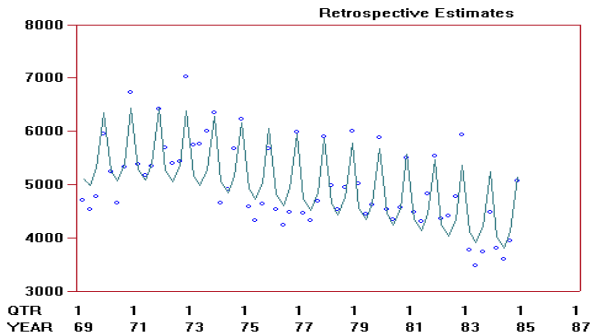
Figura: Previsiones sin monitoramiento



Resultado del ajuste de un modelo con 2 componentes: **una tendencia lineal y una forma libre para la estacionalidad** y especificando **descuentos constantes de 0,98 para la tendencia y la estacionalidad**. Se puede ver que **el modelo aprende rápido sobre el comportamiento estacional mas se comporta muy mal en el inicio de los intervalos donde ocurren los cambios**. El motivo de la demora a ajustarse a la mudanza es la especificación de **descuentos altos**.

Accidentes graves: monitoramiento (cont.)

Figura: Estimaciones suavizadas

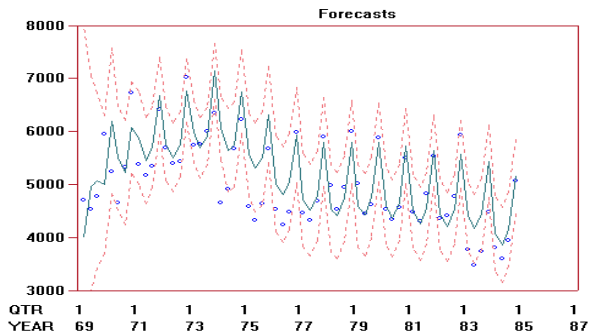


Resultado del análisis retrospectivo. Se puede ver que un comportamiento estacional bastante estable sobre una tendencia lineal que varía suavemente. La estructura del modelo imposibilita mudanzas bruscas en cualquier de sus componentes.

BATS:
 Trend: LINEAR
 Seasonal: FREE-FORM
 Discount: CONSTANT
 Fit: REFERENCE.

Accidentes graves: monitoramiento (cont.)

Figura: Previsiones con monitoramiento



Para mejorar la performance del modelo, **"encendemos" el monitor.**

Por lo menos, esperamos que señale los tiempos donde existe deterioración do modelo.

BATS:
Trend: LINEAR
Seasonal: FREE-FORM
Discount: CONSTANT
Interrupt: Level Decrease
Monitor

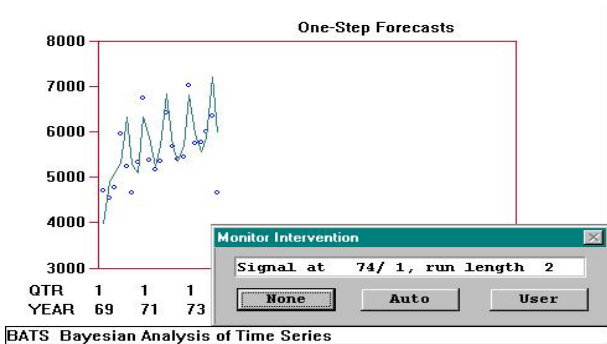
Fit: REFERENCE.

Aplicación: Accidentes graves (cont.)

- La figura anterior muestra que el análisis es el mismo!
- Los errores de 1974 no fueron lo suficientemente grandes para hacer que el monitor suene. En 1974 el sistema aun estaba poco cierto debido a la inicialización de referencia.
- Observe que los límites de 90 % de incertidumbre no están muy lejos de las observaciones de 1974 y que los límites aumentaron en seguida a ellas.
- Para confirmar el punto anterior, vamos a repetir el análisis con **menor incertidumbre a priori**.

Accidentes graves: monitoramiento (cont.)

Figura: Previsiones con monitoramiento

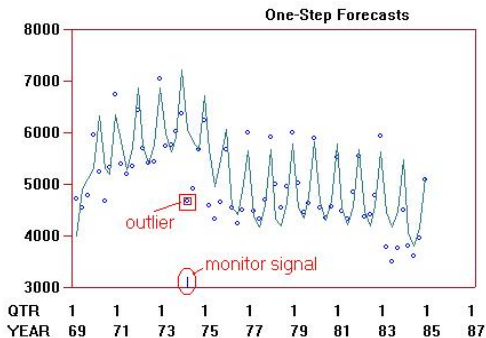


El monitor ahora señala el **1^o trimestre de 1974**. La **reducción de la incertidumbre inicial** fue suficiente para tornar esa observación extrema lo suficiente para **"sonar la alarma"**.

En este momento tenemos tres opciones: No hacer nada (vimos que no es apropiado); dejar que el BATS **actúe automáticamente** o hacer algo **manualmente**.

Accidentes graves: monitoramiento (cont.)

Figura: Previsiones con monitoramiento con menos incertidumbre a priori



Si dejamos al BATS solucionar el problema tenemos que el análisis continua tratando esa observación como aberrante y no la incorpora al análisis. La adaptación al nuevo nivel es mucho más rápida.

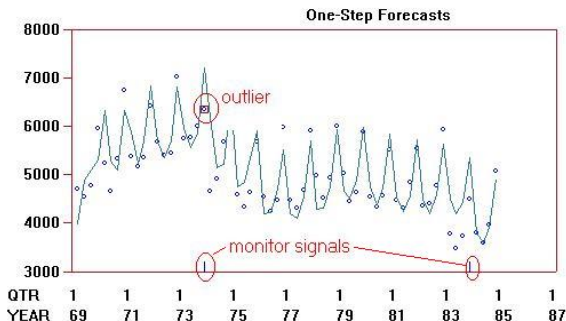
Monitor del BATS

El **monitor del BATS** funciona con base en las siguientes reglas:

- Si la señalización se debe **apenas a la observación más reciente**, el BATS la ignora y aumenta la incertidumbre al respecto de los **parámetros**
- Se la señalización se debe **a las últimas k observaciones, existe una indicación de mudanza estructural en la serie y apenas ocurre un aumento en la incertidumbre al respecto de los parámetros.**
- El aumento de la incertidumbre es realizada a través de la disminución momentánea de los factores de descuento para **0,1 para la tendencia y componente estacional, 0,8 para regresoras y 0,9 para la varianza de las observaciones**. Esos valores y la sensibilidad del monitor también pueden ser alterados.

Cambiando las especificaciones del monitor de disminución en el nivel (desvío estándar para $-2,5$ y límite del factor de Bayes para $0,3$) tenemos:

Figura: Previsiones con monitor más sensible



Ahora, el monitor señala en 1973/4 y 1983/4, sólo que en el último caso con base en una serie de 4 observaciones. El monitor no señaló en 1974/1 pues el modelo estaba preparado para mudanzas con incertidumbres aumentadas e ninguna observación de 1983 fue ignorada.

Intervención

Intervención

- Normalmente, un sistema de previsión funciona sin cambios durante su análisis. Sin embargo, existen momentos que es necesario hacer cambios en su funcionamiento.
- Anteriormente, discutimos la posibilidad del conocimiento de eventos excepcionales interviniendo en la serie estudiada, usando como ejemplo la quiebra de un competidor. Aunque este tipo de información no sea parte de la serie histórica, es fundamental para el éxito del modelo, que sea incorporada al análisis.
- En la notación utilizada, **en lugar de evolucionar de $p(\theta_{t-1}|D_{t-1})$ para $p(\theta_t|D_{t-1})$** debemos hacerlo para $p(\theta_t|D_{t-1}, I_t)$ donde I_t consiste en la información relevante. Note que esta mudanza es esencialmente subjetiva.

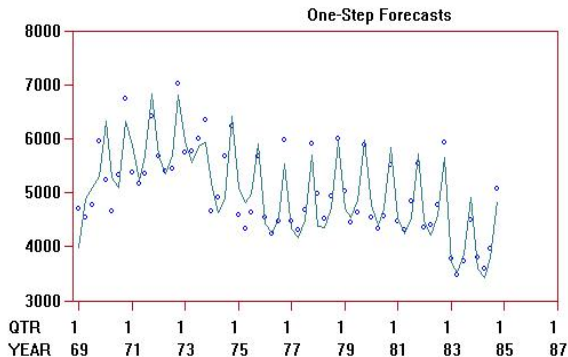
Intervención (cont.)

- Hasta ahora sólo hemos intervenido en el modelo después de la señalización del monitor. Sin embargo, esta restricción no es necesaria. En la práctica, muchas veces tenemos información que nos sugiere posibles puntos de mudanzas en la serie.
- En el caso de la serie de accidentes tenemos algunas informaciones importantes:
 - En 1974/1 la crisis del petróleo forzó un aumento significativo del precio de la gasolina, y
 - En 1983/1 pasó a ser obligatorio el uso del cinturón de seguridad en los carros.

Aplicación: Accidentes graves

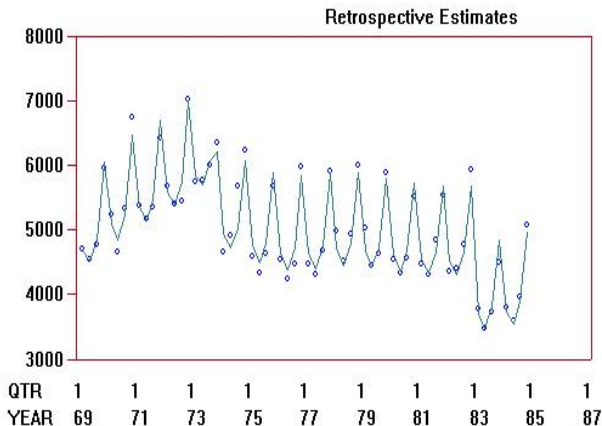
Figura: Previsiones con intervención anticipada en 1973/4 y 1983/1

En la serie de accidentes indicamos al BATS que vamos a intervenir en 1973/4 y 1983/1. El programa se detiene en esos dos puntos antes de incorporar las observaciones. Las prioridades del nivel y del crecimiento fueron cambiadas.



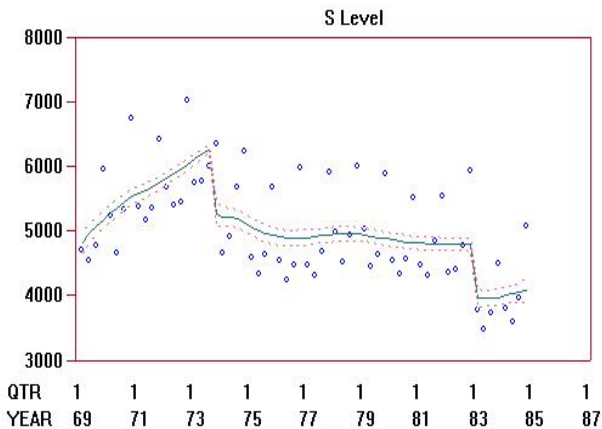
Accidentes graves: intervención

Figura: Análisis retrospectivo con intervención anticipada



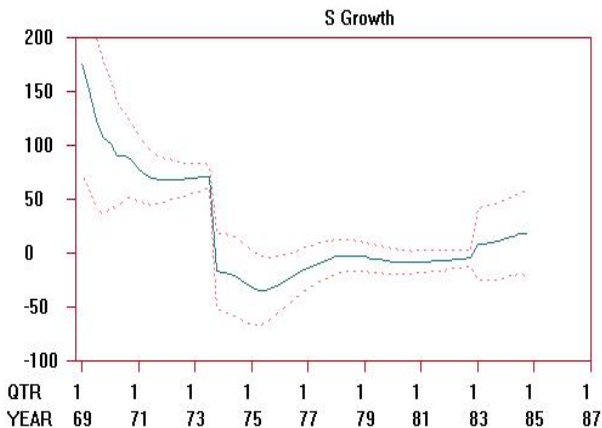
Accidentes graves: intervención (cont.)

Figura: Nivel suavizado con intervención anticipada



Accidentes graves: intervención (cont.)

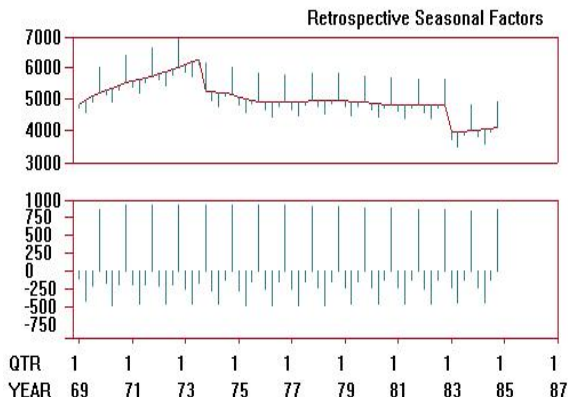
Figura: Crecimiento suavizado con intervención anticipada



Accidentes graves: intervención (cont.)

Figura: Padrón Estacional del análisis con intervención anticipada

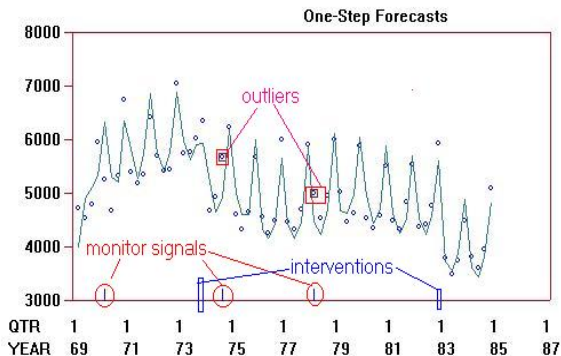
El comportamiento estacional es bastante estable a lo largo de toda la serie.



Accidentes: Todas as intervenciones

Normalmente **podemos monitorear e intervenir en un mismo análisis**

Figura: Previsiones con todas las intervenciones



Accidentes: Todas as intervenciones (cont.)

En la figura anterior:

- El monitor señala en **1970/1**, mas la incertidumbre aun es grande y por lo tanto es mejor no intervenir.
- El análisis para el periodo pre-especificado **1973/4** donde se hacen los cambios en la priori.
- El monitor señala en **1974/3** y **1978/1**, en ambos se permite la intervención automática.
- Finalmente el análisis para en **1983/1** donde se repiten las mudanzas en la priori.

Parte IV

Modelos Lineales Dinámicos Normales: Inferencia con Métodos de Simulación Estocástica

Métodos de simulación estocástica

Monte Carlo (MC)

- Métodos Monte Carlo (MC): métodos de inferencia con base en simulación.
- MC posibilita la implementación eficiente de métodos de simulación con modelos complejos como los modelos dinámicos.
- Necesario si se quiere estimar W o elementos desconocidos de F y G .
- En los MLD no se consigue generar directamente de la posteriori
- Se puede usar MC para generar valores de la distribución a posteriori de los parámetros de los modelos dinámicos. En particular, **MCMC**.

Monte Carlo via cadenas de Markov (MCMC)

Definición (MCMC)

- Algoritmos utilizados para generar muestras de distribuciones complejas (Gamerman & Lopes, 2006).
- Consisten en la construcción de cadenas de Markov, que bajo ciertas condiciones, tengan como **distribución estacionaria** a la **distribución de interés**.
- En inferencia bayesiana: utilizados para **generar muestras de la distribución a posteriori**. Una vez que las cadenas alcanzan la convergencia, se considera que los valores generados son muestras de la distribución de interés. Estadísticas como media, mediana, deciles, etc, calculadas a partir de esas muestras constituyen las estimaciones.
- Software: **WinBUGS** (<http://www.mrc-bsu.cam.ac.uk/bugs/>).
Ox (<http://www.doornik.com/download.html>).

Monte Carlo via cadenas de Markov (MCMC) (cont.)

- Un algoritmo Monte Carlo via cadenas de Markov (MCMC) para simular de una distribución π es cualquier método que produzca una cadena de Markov homogénea, ergódica e irreducible cuya distribución estacionaria sea π . (Una cadena es ergódica si es aperiódica y recurrente positiva)
- MCMC: muestras dependientes son generadas de una cadena de Markov cuya distribución de equilibrio es la distribución de interés. Aviso: (1) muestras de MCMC generalmente son muy correlacionadas; (2) estimaciones de muestras correlacionadas tienden a tener varianza mayores de que muestras independientes.
- Tópicos importantes: ¿Cuál debe ser el tamaño del *burn-in*? ¿Las cadenas están visitando todo el espacio paramétrico? ¿Cuántas iteraciones realizar?
- Ejemplos de MCMC: [Gibbs Sampling](#) (Gelfand & Smith, 1990). [Metropolis-Hastings](#) (Metropolis et al. (1953) y Hastings (1970)).

Gibbs sampling

- Algoritmo presentado en Gelfand and Smith (1990). Consiste en:

Gibbs sampling

- 1 Dar valores iniciales a los parámetros desconocidos.
 - 2 Muestrear valores de las **densidades condicionales completas** de cada parámetro, respetando un orden de muestreo y condicionando el muestreo en el último valor generado de los otros parámetros.
 - 3 Repetir el paso 2 N veces.
- Entonces, para obtener una muestra de $p(\theta_1, \theta_2)$ paseamos aleatoriamente por el espacio paramétrico de acuerdo con la siguiente regla de transición:
 - Cada iteración produce un movimiento en una dirección solamente, simulando de $p(\theta_1|\theta_2)$ y $p(\theta_2|\theta_1)$.
 - Después del periodo de *burn-in* (cuando se pierde la influencia del punto inicial, alcanzando la convergencia) comenzamos a tener muestras de $p(\theta_1, \theta_2)$

Metropolis-Hastings

- Este algoritmo fue originalmente propuesto por Metropolis et al. (1953) y modificado por Hastings (1970). Se utiliza cuando las densidades condicionales completas no tienen forma cerrada conocida.
- Suponga que $\pi(x)$ es la densidad de interés y x es el valor actual de la cadena de Markov. El algoritmo de Metropolis-Hastings consiste en:

Metropolis-Hastings

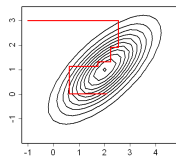
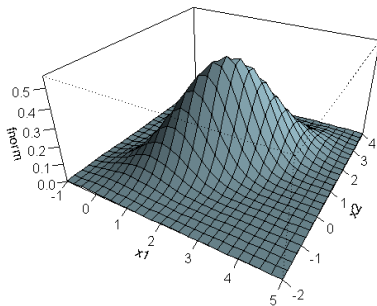
- 1 Generar un valor candidato, x^* , a partir de una densidad propuesta, o núcleo de transición, $q(x^*|x)$
 - 2 Aceptar el valor generado con probabilidad $\min\left\{1, \frac{\pi(x^*)q(x|x^*)}{\pi(x)q(x^*|x)}\right\}$.
- En la práctica, un paso de Metropolis-Hastings es insertado dentro del Gibbs sampling para generar las muestras. La densidad $q(\cdot)$ debe ser relativamente simple de muestrear. La tasa de convergencia depende de la proximidad entre $\pi(\cdot)$ y $q(\cdot)$.

Metropolis-Hastings (cont.)

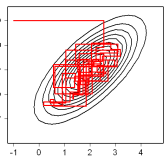
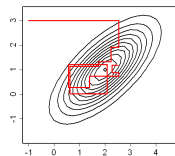
Sobre las reglas de transición propuestas

- Principales diferencias:
 - Selección de las reglas de transición propuestas,
 - Forma de acelerar la convergencia
- Actualización del vector de estados:
 - θ_t a cada paso (*Single move*),
 - $\Theta = (\theta_1, \dots, \theta_n)$ a cada paso (*Multi move*),
 - (θ_r, θ_s) a cada paso (*Block move*).

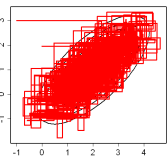
Ejemplo de MCMC: Normal bivariada



(a) 10 iteraciones (b) 20 iteraciones



(c) 100 iteraciones



(d) 1000 iteraciones

Figura: Gibbs Sampling

Ejemplo de MCMC: Normal bivariada

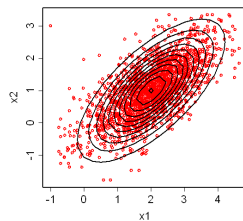
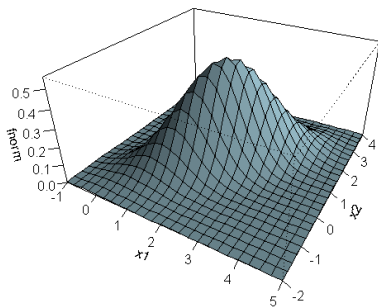
$\mathbf{x} = (x_1, x_2) \sim N(\mathbf{m}, \mathbf{S})$ donde $\mathbf{m} = (m_1, m_2)'$ y $\begin{pmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{pmatrix}$. Se sabe que

$$x_1|x_2 \sim N(m_1 + (s_{12}/s_{22}) \times (x_2 - m_2); s_{11} - s_{12}s_{21}/s_{22})$$

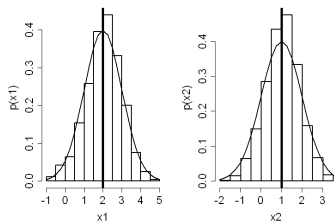
$$x_2|x_1 \sim N(m_2 + (s_{21}/s_{11}) \times (x_1 - m_1); s_{22} - s_{21}s_{12}/s_{11})$$

```
# Gibbs Sampler
M <-1000
x <-c(-1, 3)
xs <-x
  for (i in 1:M){
x[1]<-rnorm( 1,m[1]+S[1,2]/S[2,2]*(x[2]-m[2]),
           sqrt(S[1,1]-S[1,2]^2/S[2,2]) )
x[2]<-rnorm( 1,m[2]+S[2,1]/S[1,1]*(x[1]-m[1]),
           sqrt(S[2,2]-S[2,1]^2/S[1,2]) )
xs <-rbind(xs,x)
}
```

Ejemplo de MCMC (cont.)



(a) Muestra conjunta



(b) Muestra marginal

Inferencia en MLD con MCMC

MLD: Distribuciones condicionales completas

- Las distribuciones condicionales completas se calculan a partir de la distribución a posteriori conjunta de todas las cantidades desconocidas de un modelo.
- Sean $\theta = (\theta_1, \dots, \theta_n)$ y $\mathbf{Y}^t = (Y_1, \dots, Y_t)$
- En el caso de los MLD univariado, los parámetros son: θ , V y \mathbf{W} (asumiendo que $V_t = V$ y $\mathbf{W}_t = \mathbf{W}, \forall t$)

Distribución conjunta a posteriori de θ , V y \mathbf{W}

$$\begin{aligned} p(\theta, V, \mathbf{W} | \mathbf{Y}^n) &\propto p(\mathbf{Y}^n | \theta, V) p(\theta | \mathbf{W}) p(\theta_1) p(V) p(\mathbf{W}) \\ &\propto \prod_{t=1}^n p(Y_t | \theta_t, V) \prod_{t=2}^n p(\theta_t | \theta_{t-1}, \mathbf{W}) p(\theta_1) p(V) p(\mathbf{W}) \end{aligned}$$

MLD: Distribuciones condicionales completas (cont.)

Distribución condicional completa de θ_t

$$\begin{aligned}\pi_t(\boldsymbol{\theta}_t) &\propto p(Y_t|\boldsymbol{\theta}_t)p(\theta_{t+1}|\boldsymbol{\theta}_t)p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}) \\ &\propto N[\mathbf{F}'_t\boldsymbol{\theta}_t, V]N[\mathbf{G}'_{t+1}\boldsymbol{\theta}_{t+1}, \mathbf{W}_{t+1}]N[\mathbf{G}_t\boldsymbol{\theta}_{t-1}, \mathbf{W}_t] \\ &= N[\mathbf{b}_t, \mathbf{B}_t]\end{aligned}$$

donde

- Para $t = 2, \dots, n - 1$:

$$\begin{aligned}\mathbf{b}_t &= \mathbf{B}_t(V^{-1}\mathbf{F}_tY_t + \mathbf{G}'_{t+1}\mathbf{W}_{t+1}^{-1}\boldsymbol{\theta}_{t+1} + \mathbf{W}_t^{-1}\mathbf{G}_t\boldsymbol{\theta}_{t-1}) \\ \mathbf{B}_t &= (V^{-1}\mathbf{F}_t\mathbf{F}'_t + \mathbf{G}'_{t+1}\mathbf{W}_{t+1}^{-1}\mathbf{G}_{t+1} + \mathbf{W}_t^{-1})^{-1}\end{aligned}$$

- Para $t = 1, n$:

$$\begin{aligned}\mathbf{b}_1 &= \mathbf{B}_1(V^{-1}\mathbf{F}_1Y_1 + \mathbf{G}'_2\mathbf{W}_2^{-1}\boldsymbol{\theta}_2 + \mathbf{R}^{-1}a) \\ \mathbf{B}_1 &= (V^{-1}\mathbf{F}_1\mathbf{F}'_1 + \mathbf{G}'_2\mathbf{W}_2^{-1}\mathbf{G}_2 + \mathbf{R}^{-1})^{-1} \\ \mathbf{b}_n &= \mathbf{B}_n(V^{-1}\mathbf{F}_nY_n + \mathbf{W}_n^{-1}\mathbf{G}_n\boldsymbol{\theta}_{n-1}) \\ \mathbf{B}_n &= (V^{-1}\mathbf{F}_n\mathbf{F}'_n + \mathbf{W}_n^{-1})^{-1}\end{aligned}$$

MLD: Distribuciones condicionales completas (cont.)

Si: $V_t = V = \phi^{-1}$, $W_t = W = \Phi^{-1}$, $\phi \sim \text{Gamma}(n_V/2, n_V S_V/2)$ y $\Phi \sim \text{Gamma}(n_W/2, n_W S_W/2)$.

Condicional completa de $\phi = V^{-1}$

$$\pi(\phi | \boldsymbol{\theta}, \Phi) \propto \prod_{t=1}^n N[\mathbf{F}'_t \boldsymbol{\theta}_t, \phi^{-1}] G(n_V/2, n_V S_V/2) \propto G(n_V^*/2, n_V^* S_V^*/2)$$

Condicional completa de $\Phi = W^{-1}$

$$\pi(\Phi | \boldsymbol{\theta}, \Phi) \propto \prod_{t=2}^n N[\mathbf{G}_t \boldsymbol{\theta}_{t-1}, \phi^{t-1}] G(n_W/2, n_W S_W/2) \propto G(n_W^*/2, n_W^* S_W^*/2)$$

donde:

$$n_V^* = n_V + n \qquad n_V^* S_V^* = n_V S_V + \sum_t^n (Y_t - \mathbf{F}'_t \boldsymbol{\theta}_t)^2$$

$$n_W^* = n_W + n - 1 \qquad n_W^* S_W^* = n_W S_W + \sum_{t=2}^n (\boldsymbol{\theta}_t - \mathbf{G}_t \boldsymbol{\theta}_{t-1})(\boldsymbol{\theta}_t - \mathbf{G}_t \boldsymbol{\theta}_{t-1})'$$

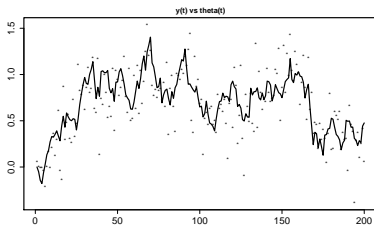
MLD: Ejemplo en el WinBUGS

Programa en WinBUGS

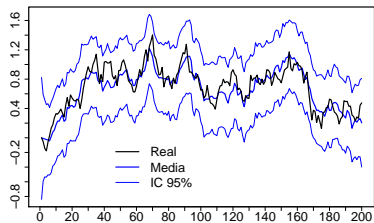
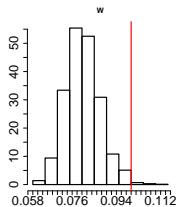
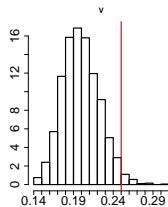
```
model "modelo lineal normal dinámico";
{
  for(t in 2:ndados)
  {
    th[t] ~ dnorm(th[t-1],tau.th)
    y[t] ~ dnorm(th[t],tau.y)
    y.hat[t] ~ dnorm(th[t],tau.y)
  }

  # Prioris
  tau.y ~ dgamma(0.001,0.001)
  tau.th ~ dgamma(0.001,0.001)
  th[1] ~ dnorm(0,0.01)
  var.y <- 1/tau.y
  var.th <- 1/tau.th
}
```

MLD: Ejemplo en el WinBUGS (cont.)



(c) Datos

(d) θ_t 

(e) Varianzas

MLD: FFBS

- Algoritmo propuesto independientemente por (Frühwirth-Schnater, 1994) y Carter and Kohn (1994)
- Para **muestrear de $p(\boldsymbol{\theta} | \mathbf{Y}^n)$ en block (*multimove*)** considere la siguiente representación condicional de una distribución normal multivariada:

$$p(\boldsymbol{\theta} | \mathbf{Y}^n) = p(\boldsymbol{\theta}_n | \mathbf{Y}^n) \prod_{t=1}^{n-1} p(\theta_t | \theta_{t+1}, \dots, \theta_n, \mathbf{Y}^n). \quad (16)$$

- Usando el teorema de Bayes:

$$\underbrace{p(\boldsymbol{\theta} | \mathbf{Y}^n)}_{\text{posteriori}} \propto p(\boldsymbol{\theta}_n | \mathbf{Y}^n) \prod_{t=1}^{n-1} \underbrace{p(\theta_t | \theta_{t+1}, \mathbf{Y}^t)}_{\text{retrospectiva}} \propto p(\boldsymbol{\theta}_n | \mathbf{Y}^n) \prod_{t=1}^{n-1} p(\theta_{t+1} | \theta_t) \underbrace{p(\theta_t | \mathbf{Y}^t)}_{\text{on-line}}.$$

Esta relación motivó el nombre de

FFBS = “**forward-filtering-backward-sampling**”

MLD: FFBS(cont.)

FFBS

- 1 Usar el filtro de Kalman para determinar la media y varianza de las “on-line” $p(\theta_t | \mathbf{Y}^t)$ de $t = 1$ a $t = N$.
 - 2 Generar un valor de θ_n , de $p(\theta_n | \mathbf{Y}^n)$
 - 3 Calcular la media y la varianza de $p(\theta_{n-1} | \theta_n, \mathbf{Y}^n)$ y generar un valor de θ_{n-1} .
 - 4 Calcular recursivamente $p(\theta_t | \theta_{t+1}, \mathbf{Y}^n)$ y generar θ_t para $t = n - 1$ a $t = 1$.
- Cada θ_t es generado condicionado en \mathbf{Y}^n
 - El FFBS es más eficiente que la generación individual (estado a estado) porque lleva en cuenta la dependencia temporal existente entre las observaciones.

Datos artificiales I: FFBS

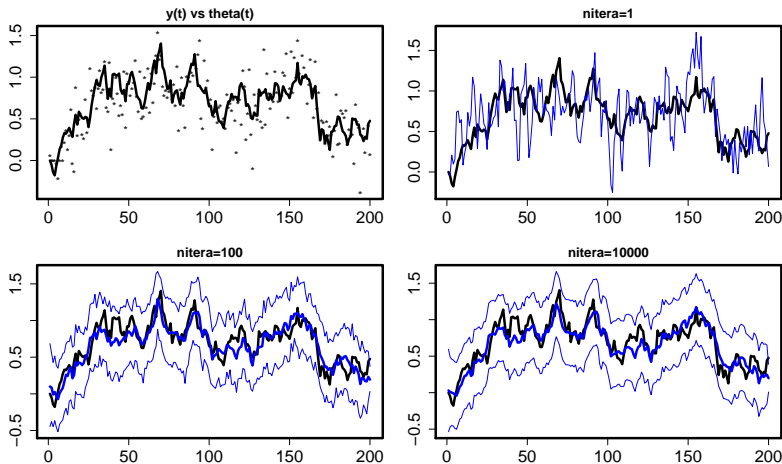


Figura: FFBS: comparación de resultados según número de iteraciones

MLD: Ejercicio de FFBS (cont.)

- Considere el siguiente modelo:

$$Y_t = \mu + E_t + \nu_t, \quad \nu_t \sim N(0, V_t) \quad (17a)$$

$$E_t = \rho E_{t-1} + \gamma X_t + w_t \quad w_t \sim N(0, W_t) \quad (17b)$$

- El modelo en (17) puede ser re-escrito como:

$$Y_t - \mu = F\theta_t + \nu_t \quad \Rightarrow \quad Y_t^* = 1E_t + \nu_t, \quad \nu_t \sim N(0, V_t)$$

$$\theta_t = G_t\theta_{t-1} + w_t^* \quad \Rightarrow \quad E_t = \rho E_{t-1} + w_t^*, \quad w_t^* \sim N(\gamma X_t, W_t)$$

- Realizaremos los cuentas del FFBS, suponiendo que ρ es conocido y que se tiene la siguiente información:

$$E_{t-1} | D_{t-1} \sim N(m_{t-1}, C_{t-1})$$

$$E_t = \rho E_{t-1} + w_t^*, \quad w_t^* \sim N(\gamma X_t, W_t)$$

MLD: FFBS (Frühwirth-Schnater, 1994) (cont.)

Evolución:

$$E(E_t|D_{t-1}) = E[E(E_t|E_{t-1}, D_{t-1})] = E(\rho E_{t-1} + \gamma X_t) = \rho m_{t-1} + \gamma X_t.$$

$$\begin{aligned} \text{Var}(E_t|D_{t-1}) &= \text{Var}[E(E_t|E_{t-1}, D_{t-1})] + E[\text{Var}(E_t|E_{t-1}, D_{t-1})] \\ &= \text{Var}(\rho E_{t-1} + \gamma X_t) + W = \rho^2 C_{t-1} + W. \end{aligned}$$

$$\therefore E_t|D_{t-1} \sim N(\rho m_{t-1} + \gamma X_t, \rho^2 C_{t-1} + W)$$

Actualización:

$$\begin{aligned} p(E_t|D_t) &\propto p(Y_t|E_t)p(E_t|D_{t-1}) \\ &\propto N(E_t, V)N(\rho m_{t-1} + \gamma X_t, \rho^2 C_{t-1} + W) \end{aligned}$$

$$\therefore E_t|D_t \sim N(m_t, C_t)$$

onde:

$$m_t = C_t [V^{-1}Y_t + (\rho^2 C_{t-1} + W)^{-1}(\rho m_{t-1} + \gamma X_t)]$$

$$C_t = (V^{-1} + (\rho^2 C_{t-1} + W)^{-1})^{-1}$$

MLD: FFBS (Frühwirth-Schnater, 1994) (cont.)

Suavización:

$$E_{t+1} = \rho E_t + w_{t+1}^*, \quad w_{t+1}^* \sim N(\gamma X_{t+1}, W)$$

$$E_{t+1}|E_t \sim N[\rho E_t + \gamma X_{t+1}, W]$$

$$\left(\begin{array}{c} E_{t+1} \\ E_t \end{array} \middle| D_t \right) \sim N \left[\begin{pmatrix} \rho m_t + \gamma X_{t+1} \\ m_t \end{pmatrix}, \begin{pmatrix} \rho^2 C_t + W & \rho C_t \\ \rho C_t & C_t \end{pmatrix} \right]$$

$$\therefore E_t | E_{t+1}, D_t \sim N(m_t^s, C_t^s)$$

donde:

$$m_t^s = m_t + \rho C_t (\rho^2 C_t + W)^{-1} (E_{t+1} - \rho m_t - \gamma X_{t+1})$$

$$C_t^s = C_t - \rho^2 C_t^2 (\rho^2 C_t + W)^{-1}$$

MLD: Errores del sistema (Gamerman, 1998)

- Considere el siguiente MLD normal:

$$y_t = \beta_t + v_t, \quad v_t \sim N(0, V), \quad (18a)$$

$$\beta_t = G\beta_{t-1} + w_t, \quad w_t \sim N(0, W), \quad t = 1, \dots, n. \quad (18b)$$

$$\beta_1 \sim N(a_1, b_1). \quad (18c)$$

- A partir de la ecuación del sistema:

$$\beta_1 = w_1 = w_1$$

$$\beta_2 = G\beta_1 + w_2 = G^1 w_1 + w_2$$

$$\beta_3 = G\beta_2 + w_3 = G^2 w_1 + G w_2 + w_3$$

$$\vdots$$

$$\beta_t = G\beta_{t-1} + w_t = G^t w_1 + G^{t-1} w_2 + G^{t-2} w_3 + \dots + G^1 w_{t-1} + w_t$$

$$\therefore \beta_1 = w_1, \quad \beta_t = \sum_{j=1}^t G^{t-j} w_j, \quad t = 2, \dots, n \quad (19)$$

MLD: Errores del sistema (cont.)

- El MLD en (18) puede ser escrito como:

$$y_t = \sum_{j=1}^t G^{t-j} w_j + v_t, \quad v_t \sim N(0, V), \quad w_t \sim N(0, W), \quad t = 2, \dots, N, \quad w_0 \sim N(a_0, b_0).$$

entonces,

$$y_1 = G^{1-1} w_1 + v_1$$

$$y_2 = G^{2-1} w_1 + G^{2-2} w_2 + v_2$$

$$\vdots$$

$$y_j = G^{j-1} w_1 + G^{j-2} w_2 + \dots + G^1 w_{j-1} + w_j + v_j$$

Luego, para un t dado,

$$y_j = \sum_{k=1}^{t-1} G^{j-k} w_k + G^{j-t} w_t + \sum_{k=t+1}^j G^{j-k} w_k \quad j = 2, \dots, n \quad (20)$$

MLD: Errores del sistema (cont.)

- La verosimilitud es

$$\begin{aligned} p(y_j | w_0, \dots, w_j, V) &\propto \exp\left\{-\frac{1}{2V} \left(y_j - \mathbf{G}^{j-t} \mathbf{w}_t - \sum_{k=0}^{t-1} \mathbf{G}^{j-k} \mathbf{w}_k - \sum_{k=t+1}^j \mathbf{G}^{j-k} \mathbf{w}_k\right)^2\right\} \\ &\propto \exp\left\{-\frac{1}{2V} (y_j - H_{tj} \mathbf{w}_t - h_{tj})^2\right\} \end{aligned}$$

donde

$$\begin{aligned} H_{tj} &= \mathbf{G}^{j-t} \\ h_{tj} &= \sum_{k=1}^{t-1} \mathbf{G}^{j-k} \mathbf{w}_k + \sum_{k=t+1}^j \mathbf{G}^{j-k} \mathbf{w}_k \end{aligned}$$

MLD: Errores del sistema (cont.)

La distribución condicional completa de w_t , $t = 1, \dots, n$ es

$$\pi(w_t) \propto \prod_{j=t}^n p(y_j | w_0, \dots, w_j, V) p(w_t | W) = N(b_t, B_t) \quad (21)$$

donde

$$b_t = B_t \sum_{j=t}^n V^{-1} H_{tj} (y_j - h_{tj})$$
$$B_t = (W^{-1} + \sum_{j=t}^n V^{-1} H_{tj}^2)^{-1}$$

- Nota 1: $E[y_j]$ depende de w_0, \dots, w_j .
- Nota 2: w_t está presente en $E[y_j]$ para $j = t, \dots, n$.

Datos artificiales II: errores (Gamerman, 1998)

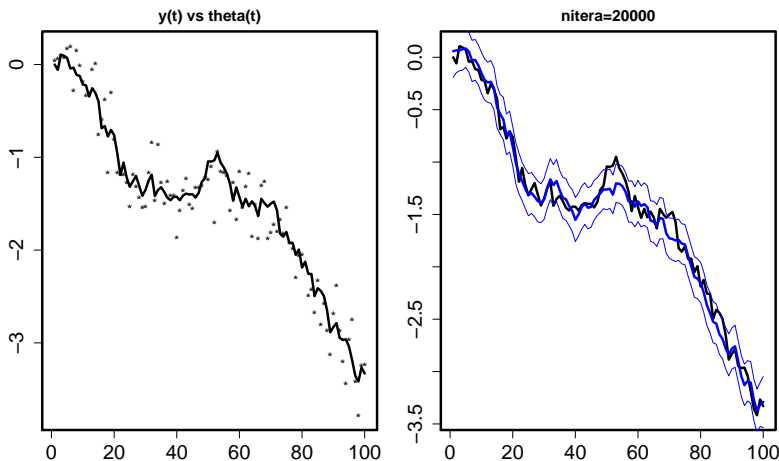


Figura: Algoritmo: muestreando los errores del sistema

Parte V

Modelos Dinámicos No-Lineales No-Normales

Modelos lineales dinámicos generalizados (MLDG)

Modelos lineales dinámicos generalizados (MLDG)

- Los modelos lineales dinámicos generalizados son una extensión de los modelos lineales dinámicos normales para la familia exponencial.
- Tienen como base los modelos lineales generalizados.
- Se supone que la distribución de las observaciones es miembro de la familia exponencial, donde η_t es el **parámetro natural** y $\phi_t = V_t^{-1}$ es la **precisión**.

$$p(Y_t|\eta_t) = \exp\left\{\phi_t[Y_t\eta_t - a(\eta_t)]\right\}b(Y_t, V_t)$$

- En el caso dinámico, η_t tiene una relación lineal con el vector de estados θ_t .
- La ecuación del sistema es igual al caso normal.

Modelos Lineales Dinámicos Generalizados(cont.)

Definición

Un modelo lineal dinámico generalizado es especificado por:

$$\text{Ecuación de observación : } p(Y_t | \eta_t) = \exp\left\{ \phi_t [Y_t \eta_t - a(\eta_t)] \right\} b(Y_t, V_t)$$

$$\text{Función de la media : } \mu_t = E(Y_t | \eta_t) = a'(\eta_t)$$

$$\text{Función de ligación : } g(\mu_t) = \mathbf{F}_t' \boldsymbol{\theta}_t$$

$$\text{Equación del sistema : } \boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t$$

$$\text{Error del sistema : } \boldsymbol{\omega}_t \sim (\mathbf{0}, \mathbf{W}_t)$$

$$\text{Información a priori : } \boldsymbol{\theta}_1 \sim (\mathbf{a}_1, \mathbf{R}_1)$$

★ No caso da normal, $\eta_t = \mu_t$

Ejemplo I

Modelo de Regresión Log-lineal Dinámico

Se $Y_t \sim \text{Poisson}(\lambda_t)$, entonces:

$$p(Y_t|\lambda_t) \propto \exp\{Y_t \log(\lambda_t) - \lambda_t\}$$

$$\phi = 1$$

$$\eta_t = \log \lambda_t$$

$$a(\eta_t) = \exp(\eta_t) = \lambda_t$$

$$\eta_t = \theta_{1t} + \theta_{2t}X_t = (1, X_t)\boldsymbol{\theta}_t$$

$$\mathbf{F}'_t = (1, X_t)$$

$$\boldsymbol{\theta}_t = (\theta_{1t}, \theta_{2t})' = \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t$$

$$\boldsymbol{\omega}_t \sim N(\mathbf{0}, \mathbf{W}), \quad \boldsymbol{\omega}_1 \sim N(\mathbf{0}, \mathbf{R})$$

Ejemplo II

Modelo de Regresión Logístico Dinámico

Se $Y_t \sim \text{Binomial}(n_t, \pi_t)$, entonces:

$$p(Y_t|n_t, \pi_t) \propto \exp\left\{Y_t \text{logit}(\pi_t) + n_t \log(1 - \pi_t)\right\}$$

$$\phi = 1$$

$$\eta_t = \text{logit}(\pi_t)$$

$$a(\eta_t) = -n_t \log(1 - \pi_t)$$

$$\eta_t = \theta_{1t} + \theta_{2t}X_t = (\mathbf{1}, X_t)\boldsymbol{\theta}_t$$

$$\mathbf{F}'_t = (\mathbf{1}, X_t)$$

$$\boldsymbol{\theta}_t = (\theta_{1t}, \theta_{2t})' = \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t$$

$$\boldsymbol{\omega}_t \sim N(\mathbf{0}, \mathbf{W}), \quad \boldsymbol{\omega}_1 \sim N(\mathbf{0}, \mathbf{R})$$

Inferencia en MLDG: Linear Bayes

Recordando: $p(y_t | \eta_t) \quad \eta_t = \mathbf{F}'_t \boldsymbol{\theta}_t;$
 $\boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim (\mathbf{0}, \mathbf{W}_t)$

- Sea $(\boldsymbol{\theta}_{t-1} | D_{t-1}) \sim (\mathbf{m}_{t-1}, \mathbf{C}_{t-1})$. Entonces $(\boldsymbol{\theta}_t | D_{t-1}) \sim (\mathbf{a}_t, \mathbf{R}_t)$ donde $\mathbf{a}_t = \mathbf{G}_t \mathbf{m}_{t-1}$ y $\mathbf{R}_t = \mathbf{G}_t \mathbf{C}_{t-1} \mathbf{G}'_t + \mathbf{W}_t$
- **Priori para η_t :** $(\eta_t | D_{t-1}) \sim (f_t, q_t)$.
 Si la priori de η_t es conjugada, se tiene que priori y posteriori pertenecen a la misma familia, entonces $(\eta_t | D_t) \sim (f_t^*, q_t^*)$.
- **Estructura Condicional de $(\boldsymbol{\theta}_t | \eta_t, D_{t-1})$:**

$$\left(\begin{pmatrix} \eta_t \\ \boldsymbol{\theta}_t \end{pmatrix} \middle| D_{t-1} \right) \sim \left(\begin{pmatrix} f_t \\ \mathbf{a}_t \end{pmatrix}, \begin{pmatrix} q_t & \mathbf{F}'_t \mathbf{R}_t \\ \mathbf{R}_t \mathbf{F}_t & \mathbf{R}_t \end{pmatrix} \right).$$

Linear Bayes $\Rightarrow \quad \widehat{E}(\boldsymbol{\theta}_t | \eta_t, D_{t-1}) = \mathbf{a}_t + \mathbf{R}_t \mathbf{F}_t (\eta_t - f_t) / q_t$
 $\widehat{Var}(\boldsymbol{\theta}_t | \eta_t, D_{t-1}) = \mathbf{R}_t - \mathbf{R}_t \mathbf{F}_t \mathbf{F}'_t \mathbf{R}_t / q_t$

Inferencia en MLDG: Linear Bayes (cont.)

- Actualización para θ_t .

$$p(\theta_t | D_t) = \int \underbrace{p(\theta_t | \eta_t, D_{t-1})}_{\text{no es conocida}} \underbrace{p(\eta_t | D_t)}_{\text{conjugación}} d\eta_t$$

Entonces tenemos que $\theta_t \sim (\mathbf{m}_t, \mathbf{C}_t)$, donde:

$$\begin{aligned} \mathbf{m}_t &= E[\theta_t | D_t] \\ &= E[\widehat{E}\{\theta_t | \eta_t, D_{t-1}\} | D_t] \\ &= \mathbf{a}_t + \mathbf{R}_t \mathbf{F}_t (f_t^* - f_t) / q_t \end{aligned}$$

$$\begin{aligned} \mathbf{C}_t &= V[\theta_t | D_t] \\ &= V[\widehat{E}\{\theta_t | \eta_t, D_{t-1}\} | D_t] + E[\widehat{V}\{\theta_t | \eta_t, D_{t-1}\} | D_t] \\ &= \mathbf{R}_t - \mathbf{R}_t \mathbf{F}_t \mathbf{F}_t' \mathbf{R}_t (1 - q_t^* / q_t) / q_t \end{aligned}$$

Aplicación: Número de crímenes con Linear Bayes

- **Objetivo:** Análisis temporal del número mensual de crímenes violentos, ocurridos entre enero de 1998 y agosto de 2001 en una región de Belo Horizonte (MG, Brasil).
- **Modelo:**

$$(Y_t | \lambda_t) \sim \text{Poisson}(\lambda_t), t = 1, \dots, 45; \quad (22a)$$

$$\log \lambda_t = \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} \mu_t \\ \beta_t \end{pmatrix} = \mu_t \quad (22b)$$

$$\begin{pmatrix} \mu_t \\ \beta_t \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_{t-1} \\ \beta_{t-1} \end{pmatrix} + \begin{pmatrix} w_{1t} \\ w_{2t} \end{pmatrix} \quad (22c)$$

$$\mu_0 \sim N(m_{10}, C_{10})$$

$$\beta_0 \sim N(m_{20}, C_{20})$$

- $w_t \sim (0, W) \rightarrow W$ especificado via descuentos ($\delta = 0,98$)

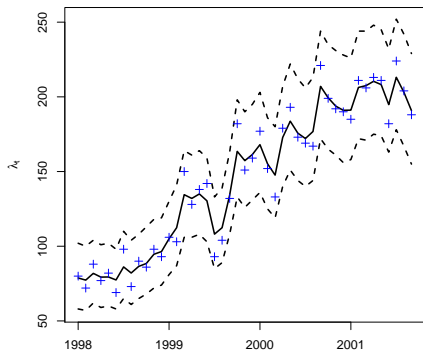
Programa en Ox

```

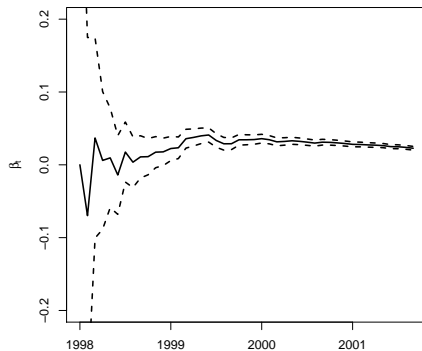
# include <oxstd.h> main() {
// Datos y Variables
  decl data = loadmat("cia6.dat"); decl nobs = (rows(data)+1);
  decl Ft = <1; 0>; decl Gt = <1, 1; 0, 1>; decl delta = 0.99;
  ...
// Información inicial
  mt[0] = <4.30; 0.00>; Ct[0] = <0.10, 0.00; 0.00, 0.01>;
// Linear Bayes
  for(i=1; i<(nobs); i++){
    at[i] = Gt*mt[i-1];
    Rt[i] = (Gt*Ct[i-1]*Gt')*(1/delta);
    ft[i] = Ft'*at[i];
    qt[i] = Ft'*Rt[i]*Ft;
    rt[i] = 1/qt[i];
    st[i] = 1/(qt[i]*exp(ft[i]));
    rs[i] = rt[i] + y[i];
    ss[i] = st[i] + 1;
    fs[i] = log(rs[i]) - log(ss[i]);
    qs[i] = 1/(rs[i]);
    mt[i] = at[i] + Rt[i]*Ft*(fs[i]-ft[i])*(1/qt[i]);
    Ct[i] = Rt[i] - (Rt[i]*Ft*Ft'*Rt[i])*(1-(qs[i]/qt[i]))*(1/qt[i]);
  }... }

```

Número de crímenes con Linear Bayes (cont.)



(a) $\lambda_t|D_t$



(b) $\beta_t|D_t$

Figura: IC 95 % para el nivel y parámetro de crecimiento de (22) estimado con Linear Bayes

Aplicación: Número de crímenes con MCMC

- **Objetivo:** Análisis temporal del número mensual de crímenes violentos, ocurridos entre enero de 1998 y agosto de 2001 en una región de Belo Horizonte (MG, Brasil).
- **Modelo:**

$$Y_t | \lambda_t \sim \text{Poisson}(\lambda_t) \quad (23a)$$

$$\log \lambda_t = \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} \mu_t \\ \beta_t \end{pmatrix} = \mu_t \quad (23b)$$

$$\begin{pmatrix} \mu_t \\ \beta_t \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_{t-1} \\ \beta_{t-1} \end{pmatrix} + \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}; \quad \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} \sim \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} W_1 & 0 \\ 0 & W_2 \end{pmatrix} \right]$$

$$\mu_0 \sim N(m_{10}, C_{10})$$

$$\beta_0 \sim N(m_{20}, C_{20})$$

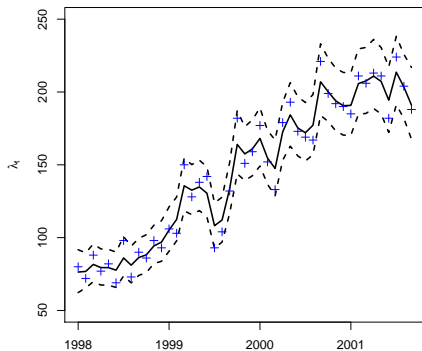
$$W_1^{-1} \sim \text{Gamma}(0,01, 0,01)$$

$$W_2^{-1} \sim \text{Gamma}(0,01, 0,01)$$

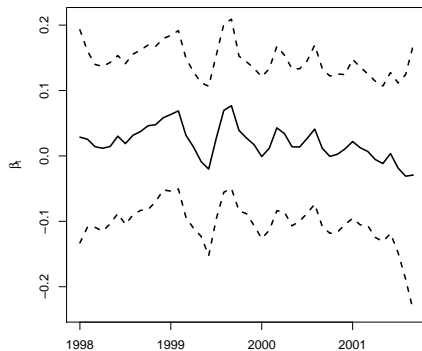
Programa en WinBUGS

```
model "modelo dinámico poisson de 2-orden";
{
# Ecuaciones del modelo
  for(t in 2:46)
  {
    beta[t]      ~ dnorm(beta[t-1],iw2);
    mean.mu[t]   <- mu[t-1]+beta[t-1];
    mu[t]        ~ dnorm(mean.mu[t],iw1);
    log(lambda[t]) <- mu[t]
    y[t]         ~ dpois(lambda[t])
    y.rep[t]     ~ dpois(lambda[t]) I(0,10000)
  }
# Información inicial
  mu[1] ~ dnorm(a[1], iR10)
  beta[1] ~ dnorm(a[2], iR20)
# Distribución a priori
  iw1 ~ dgamma(0.01,0.01)
  iw2 ~ dgamma(0.01,0.01)
  w1 <- 1/iw1;   w2 <- 1/iw2;
# Hyper-parámetros
  a[1] <- 4.5;   a[2] <- 0.2;
  iR10 <- 2.0;  iR20 <- 1;
}
```


Número de crímenes con MCMC (cont.)



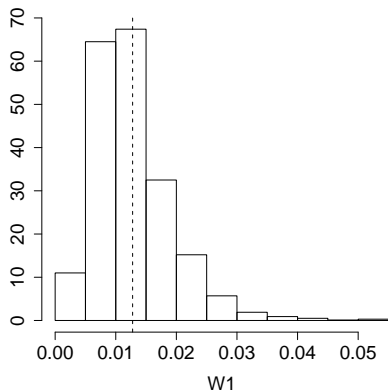
(a) $\lambda_t | \mathbf{y}$



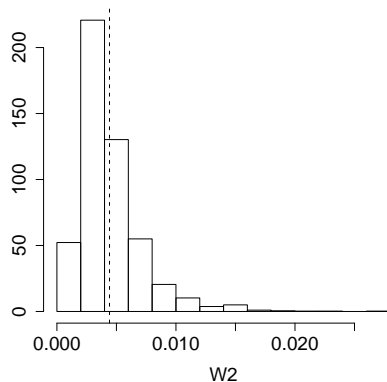
(b) $\beta_t | \mathbf{y}$

Figura: IC 95 % para el nivel y parámetro de crecimiento de (23) estimado con MCMC

Número de crímenes con MCMC (cont.)



(a) $W_1|\mathbf{y}$



(b) $W_2|\mathbf{y}$

Figura: Muestras de las distribuciones a posteriori de W_1 y W_2 de (23) obtenidas con MCMC (A línea punteada indica la media a posteriori)

Modelos no lineales dinámicos generalizados (MNLDG)

Modelos no lineales dinámicos generalizados (MNLDG)

Sea Y_t una serie temporal observada de $t = 1$ a $t = T$.

Definición (MNLDG)

El modelo no lineal dinámico generalizado está dado por:

$$\text{Ecuación de observación :} \quad Y_t = f_t(\boldsymbol{\theta}_t) + v_t \quad (24a)$$

$$\text{Ecuación del sistema :} \quad \boldsymbol{\theta}_t = g_t(\boldsymbol{\theta}_{t-1}) + w_t \quad (24b)$$

donde $f_t(\cdot)$ e $g_t(\cdot)$ son funciones conocidas.

Ejemplo: [Modelo de Función de Transferencia](#)

$$Y_t = \mu + E_t + v_t, \quad v_t \sim N(0, V) \quad (25a)$$

$$E_t = \rho E_{t-1} + \gamma X_t \quad (25b)$$

Aplicación: datos de lluvia y caudal

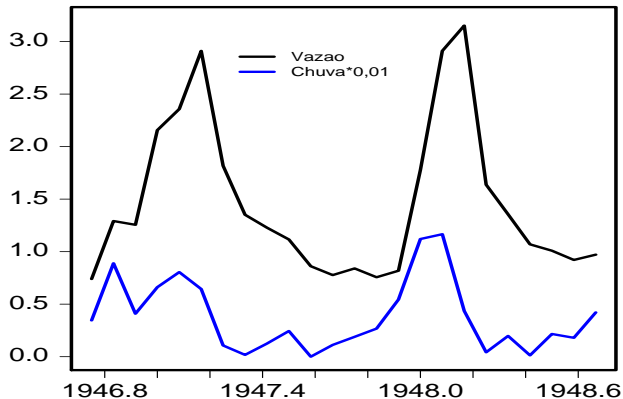


Figura: Caudal y Precipitación de octubre de 1946 a setiembre de 1948 en la cuenca del Riberão Pinheirinho - SP

Aplicación: datos de lluvia y caudal (cont.)

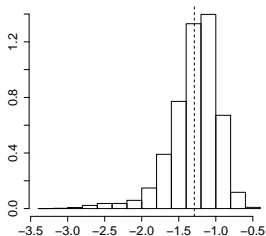
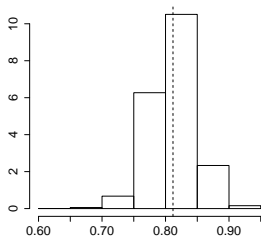
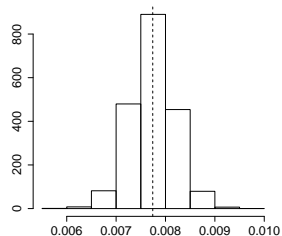
(a) $\mu|y$ (b) $\rho|y$ (c) $\gamma|y$

Figura: Muestras de la distribución a posteriori de μ , ρ y γ de (23) obtenidas con MCMC (La línea punteada indica la media a posteriori)

Aplicación: datos de lluvia y caudal (cont.)

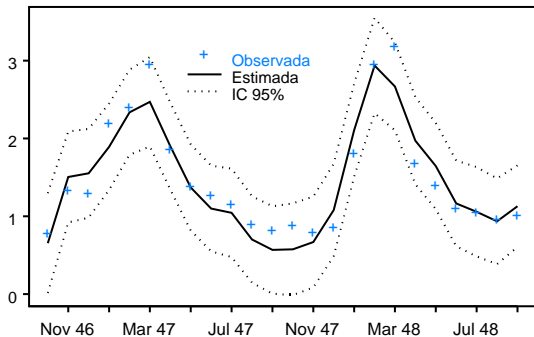


Figura: Caudal observado e IC 95 % para el caudal estimado con el modelo (25)

Algumas referencias

- Carter, C., & Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika*, *81*, 541–553.
- Frühwirth-Schnater, S. (1994). Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, *15*(2), 183–202.
- Gamerman, D. (1998). Markov chain Monte Carlo for dynamic generalised linear models. *Biometrika*, *85*(1), 215–227.
- Gamerman, D., & Lopes, H. F. (2006). *Markov chain Monte Carlo: Stochastic simulation for Bayesian inference*. New York: Chapman & Hall / CRC.
- Gelfand, A. E., & Smith, A. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, *85*(410), 398–409.
- Hastings, W. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, *57*, 97–109.
- Metropolis, N., Rosenbluth, M., Teller, A., & Teller, E. (1953). Equations of state calculations by fast computing machines. *Journal of chemical physics*, *21*, 1087–1092.
- West, M., & Harrison, J. (1997). *Bayesian forecasting and dynamic models* (2nd ed.). New York: Springer-Verlag.

romy@dme.ufrj.br
romy@fiocruz.br
dme.ufrj.br/romy