SHORT COURSE

DYNAMIC MODELS

Instructor: Dani Gamerman

dani@im.ufrj.br

Instituto de Matemática

UFRJ - Brazil

currently visiting Stats. Department - UConn

Winter 2002

Objectives:

Provide an understanding of the structural approach to time series modelling under a Bayesian perspective. Students will be expected to finish the course with the ability to analyse and criticise results obtained for a time series of interest.

References:

Bayesian Forecasting and Dynamic Models by M. West and P. J. Harrison. Springer. 2nd. Edition. 1997.

Applied Bayesian Forecasting and Time Series Analysis by A. Pole, M. West and P. J. Harrison. Chapman & Hall. 1994.

Main software:

BATS

www.stat.duke.edu/~mw/books_software_data.html

Course outline

Week 1: Introduction: probability, Bayes and 1st order models

Week 2: General models and trend and seasonality

Week 3: Superposition, discount factors, variance learning, filtering and intervention

Week 4: Monitoring, non-normality, variance laws, cycles and hyperparameter estimation (MC and MCMC)

Introduction

time series - collection of observations ordered (in time).

Time may be space, depth, $\cdots$;

Adjancent observations are dependent.

Examples :

Economics - daily prices of stocks

monthly unemployment rate

Medicine - levels of EEG

Epidemiology - weekly cases of measles

monthly cases of AIDS

Metereology - daily temperature

height of tides

Environmental sciences - pollution counts

⋮

Classification

Time series $\{Y(t), t \in T\}$

Y - variable of interest

T - set of indices

Types of time series

1. Discrete : $T = \{t_1, t_2, \cdots, t_n\}$

Example : Monthly exports from 1970 to 1990

$T = \{01/1970, 02/1970, \cdots, 11/1990, 12/1990\}$

Notation : Y_t

2. Continuous : $T = \{t : t_1 < t < t_2\}$

Example : Height of tide in New Haven today

$T = [0, 24]$ if time unit is hour

Notation : $Y(t)$

3. Multivariate : Observations are $\{Y_1(t), \cdots, Y_k(t), t \in T\}$

Example : Monthly sales (Y_{1t}) and advertising expenditure (Y_{2t}).

Y can also be discrete or continuous.

Sometimes, Y is discrete but can be treated as continuous.

Example : Number of notified AIDS cases in the U.S.

In this course, series will be univariate, discrete and observed in equidistant times.

One can identify T with $\{1, 2, \cdots, n\}$

Objectives of a time series analysis

(i) understand the generating mechanism of the series;

(ii) predict future behaviour of the series.

Understanding the generating mechanism of the series enables:

- describing efficiently series behaviour;
- finding reasons for series behaviour; (possibly through auxiliary variables)
- controlling the trajectory of the series.

Predicting the future behaviour of the series enables :

- making short, medium and long term plans;

- taking appropriate decisions.

Objectives (i) and (ii) are related.
It is only possible to forecast well routinely if the model is adequate and vice-versa.

Future usually involves uncertainty $\rightarrow$ predictions are not perfect.

Objective is to reduce **forecast errors** as much as possible.

Modelling, learning and forecasting

Central to the analysis: building a model.

Model - Scheme of description (and explanation) that organizes information (and experiences) inorder to provide learning and forecasting.

good **model** allows **learning** leading to adequate **prediction**.

Must also be economical (parsimony).

Description must be simple and flexible to adapt for the (uncertain) future and ease learning.

Uncertainty $\rightarrow$ model is probabilistic.

Learning is the processing of information by the model.

Prediciton is an hypothesis, conjecture or speculation about the future.

Dynamic nature of a time series requires models to have time adaptation.

Model structure must allow for local changes.

Changes will be modellled stochastically, i.e., using probability.

Basic idea: define models representing structure of series (cycles, trend, seasonal, etc)

Hence the name **structural models**.

Main feature: model parameters vary probabilistically with time

When Bayesian inference is used $\rightarrow$ dynamic models

Example: suppose Y is to be forecast and is known to be influenced by X.

Simplest relation: $Y = X\theta + \epsilon$

θ may change with time

Learning rule - Bayes theorem

- Possible models $M_1, M_2, \cdots, M_p$ with prior probabilities $P(M)$, $M = M_1, M_2, \cdots M_p$;

- Observe Y with description $P(Y \mid M)$ (likelihood of model M);

- After observing $Y = y$ we have $P(M \mid Y = y)$, $M = M_1, \cdots, M_p$ given by

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

Posterior $\propto$ likelihood $\times$ Prior

Important aspects of forecasting systems

Basics: forecasting and estimation

1. INTERVENTION

Analyst may modify model according to information during observation process.

Example:

series: sales of airline tickets

event: 11 Sept. 2001

2. MONITORING

Model performance drops

$\downarrow$

Monitor signals

$\downarrow$

Changes are made

- Retrospective intervention

3. RETROSPECTION (OR FILTERING)

Forecast: what past says about future.

But what future says about past.

Secondary importance: control.

Probability distributions

Let x, y, z be random quantitites

and $f(x,y,z)$ their joint density.

$$f(x|y) = \frac{f(x,y)}{f(y)}$$

$$f(x) = \int_y f(x,y)dy$$

$$f(x,y \mid z) = \frac{f(x,y,z)}{f(z)}$$

$$f(x,y) = \int_z f(x,y,z)dz$$

$$f(x \mid y) = \int_z f(x,z \mid y)dz$$

$$f(x \mid y,z) = \frac{f(y \mid x,z)f(x \mid z)}{f(y \mid z)}$$

Results valid for vectors x, y and z.

Univariate normal distribution

The r.v. x has univariate normal distribution with mean μ and variance σ^2, denoted by $N(\mu, \sigma^2)$, if its density is

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

Standard normal dist.: $\mu = 0$ and $\sigma^2 = 1$.

Multivariate normal distribution

$x = (x_1, x_2, \cdots, x_p)'$ has multivariate normal distribution with mean μ and variance Σ, denoted by $N(\mu, \Sigma)$, if density $f_N(x;\mu,\Sigma) =$

$$(2\pi)^{p/2} \mid \Sigma \mid^{1/2} exp\left\{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right\}$$

(i) $\mid A \mid$ denotes the determinant of A

(ii) Σ diagonal $\rightarrow x_i's$ are indpendent.

(iii) Univariate normal: $p = 1$.

Important properties

Linear transformations:

$$y = Ax + b \sim N(A\mu + b, A\Sigma A')$$

Marginal distributions:

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \, and \, \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

gives $x_i \sim N(\mu_i, \Sigma_{ii})$, $i = 1, 2$.

Conditional distributions:

$$x_1 \mid x_2 \sim N(\mu_{1.2}, \Sigma_{11.2})$$

where $\mu_{1.2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)$ and $\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$.

Analogous results for $x_2 \mid x_1$.

Σ_{22} and Σ_{11} must be full rank.

Reconstruction of joint density:

$$x_1 \mid x_2 \sim N(\mu_1 + B_1(x_2 - \mu_2), B_2)$$

and $x_2 \sim N(\mu_2, \Sigma_{22}) \rightarrow$

$$\begin{aligned} x &= \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \sim N(\mu, \Sigma) \qquad \text{with} \\ \mu &= \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \end{aligned}$$

where

$$\Sigma_{11} = B_2 + B_1 \Sigma_{22} B_1'$$

$$\Sigma_{21}' = \Sigma_{12} = B_1 \Sigma_{22}.$$

Quadratic forms : $(x - \mu)' \Sigma^{-1} (x - \mu) \sim \chi^2_p$.

Gamma distribution

x has Gamma dist. with parameters α and β if its density is

$$p(x|\alpha;\beta) \propto x^{\alpha-1} exp\left\{-\beta x\right\}, x > 0.$$

Notation: $x \sim G(\alpha,\beta)$.

$$E(x) = \frac{\alpha}{\beta} \quad \text{and} \quad V(x) = \frac{\alpha}{\beta^2}$$

Inverse Gamma distribution

x has Gamma dist. with parameters α and β if $x^{-1} \sim G(\alpha,\beta)$.

Its density is

$$p(x|\alpha;\beta) \propto \left(\frac{1}{x}\right)^{\alpha+1} exp\left\{-\frac{\beta}{x}\right\}, x > 0.$$

Notation: $x \sim IG(\alpha,\beta)$.

$$E(x) = \frac{\beta}{\alpha - 1}, \alpha > 1 \text{ and } V(x) = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)}, \alpha > 2$$

Univariate t-Student distribution

x has univariate t-Student distribution with ν degress of freedom, and parameters μ and σ if its density is

$$p(x|\mu,\sigma,\nu) \propto [\nu\sigma^2 + (x-\mu)^2]^{-(\nu+1)/2}, x \in R$$

$$E(x) = \mu, \nu > 1 \quad \text{e} \quad V(x) = \frac{\nu}{\nu-2}\sigma^2, \nu > 2.$$

Notation: $x \sim t_\nu(\mu,\sigma^2)$.

Multivariate t-Student distribution

$\mathbf{x} = (x_1, ..., x_p)$ has multivariate t-Student distribution with ν degress of freedom, and parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ if its density is

$$p(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma},\nu) \propto [\nu + (\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})]^{-(\nu+p)/2}$$

Notation: $\mathbf{x} \sim t_\nu(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

$$\begin{aligned} E(\mathbf{x}) &= \boldsymbol{\mu} \quad \text{if} \quad \nu > 1 \\ V(\mathbf{x}) &= \frac{\nu}{\nu - 2}\boldsymbol{\Sigma} \quad \text{if} \quad \nu > 2. \end{aligned}$$

Normal-Gamma distribution

$x|y \sim N(\mu, y^{-1}V)$ and

$y \sim G(\frac{\nu}{2}, \frac{d}{2}) \left(\Leftrightarrow y^{-1} \sim IG(\frac{\nu}{2}, \frac{d}{2})\right)$

$$\Downarrow$$

$$(x, y) \sim NG(\mu, V, \nu, d)$$

$$\Downarrow$$

$x \sim t_\nu(\mu, SV)$ where $S = d/\nu$

Bayesian Inference

Summarized presentation of methodology

Bayes Theorem

Observations y: described by density $f(y|\theta)$

Likelihood: $l(\theta) = f(y|\theta)$

θ: index of f (parameter)

Canonical situation: random sample

$y = (y_1, ..., y_n)$ taken from $f(y|\theta)$.

Example (1). measurements of a physical quantity θ with errors $e_i \sim N(0, \sigma^2)$, σ^2 known.

$y_i = \theta + e_i$, $i = 1, ..., n$ and $f(y|\theta) =$

$$\prod_{i=1}^{n} f_N(y_i; \theta, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} exp \left\{ -\frac{1}{2} \frac{(y_i - \theta)^2}{\sigma^2} \right\}$$

θ is more than simple index

Situation repeats in more general cases

Very likely that researcher has prior information about θ

This may be modelled through density $p(\theta)$

Lots of controversy in the past

Inference process based on distribution of θ after observing $y \rightarrow$ posterior distribution (as opposed to prior)

Obtained through Bayes theorem as

$$p(\theta|y) = \frac{f(y|\theta)p(\theta)}{f(y)} \qquad \text{or}$$
$$\pi(\theta) \propto l(\theta)\, p(\theta)$$

$$f(y) = \int f(y|\theta)p(\theta)d\theta$$

Example (1) (cont.) model may be completed with prior $p(\theta) = N(\mu, \tau^2)$, μ and τ^2 known.

$$l(\theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\frac{(y_i-\theta)^2}{\sigma^2}\right\}$$
$$\propto \exp\left\{-\frac{n}{2\sigma^2}(\bar{y}-\theta)^2\right\}$$

where $\bar{y}$ is the average of the y_i's.

$$\begin{aligned}\pi(\theta) &\propto \exp\left\{-\frac{1}{2}\frac{(\bar{y}-\theta)^2}{\sigma^2/n}\right\}\exp\left\{-\frac{1}{2}\frac{(\theta-\mu)^2}{\tau^2}\right\} \\ &\propto \exp\left\{-\frac{1}{2}\frac{(\theta-\mu_1)^2}{\tau_1^2}\right\}\end{aligned}$$

where

$\tau_1^{-2} = n\sigma^{-2} + \tau^{-2}$ and

$\mu_1 = \tau_1^2(n\sigma^{-2}\bar{y} + \tau^{-2}\mu)$

i.e., $\pi(\theta) = N(\mu_1, \tau_1^2)$.

$\tau^2 \to \infty$: non-informative prior $p(\theta) \propto c$

and

$\pi(\theta) = N(\bar{y}, \sigma^2/n)$.

Example (2) Assume now that the obs. **variance** is also **unknown**. Model becomes

$$
\begin{aligned}
(y \mid \theta, \sigma^2) &\sim N(\theta, \sigma^2) \\
(\theta \mid \sigma^2) &\sim N(a, \sigma^2 R)
\end{aligned}
$$

Parameter now is bidimensional.

Prior must be completed with marginal distribution for σ^2.

Convenience: work with $\phi = \sigma^{-2}$ with prior $\phi \sim G\left(\frac{n}{2}, \frac{d}{2}\right) \Leftrightarrow \sigma^2 \sim IG\left(\frac{n}{2}, \frac{d}{2}\right)$.
So, $(\theta, \phi) \sim NG(a, R, n, d)$.

$$
\begin{aligned}
f(\theta \mid \sigma^2) &= \frac{1}{\sqrt{2\pi\sigma^2 R}} \exp\left\{-\frac{1}{2\sigma^2 R}(\theta - a)^2\right\}, \\
f(\theta \mid \phi) &= \frac{\phi^{1/2}}{\sqrt{2\pi R}} \exp\left\{-\frac{\phi}{2R}(\theta - a)^2\right\}, \\
f(\phi) &\propto \phi^{(n/2)-1} \exp\left\{-\frac{\phi}{2} d\right\}
\end{aligned}
$$

The likelihood is

$$f(y|\theta,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y-\theta)^2\right\}$$
$$f(y|\theta,\phi) = \frac{\phi^{1/2}}{\sqrt{2\pi}} \exp\left\{-\frac{\phi}{2}(y-\theta)^2\right\}$$

Applying Bayes theorem

$$f(\theta,\phi|y) \propto f(y|\theta,\phi)p(\theta,\phi)$$
$$\propto \phi^{[(n+2)/2]-1} exp\left\{-\frac{\phi}{2}\left[\frac{(\theta-m)^2}{C} + \frac{(y-a)^2}{R+1} + d\right]\right\}$$

Since $f(\theta,\phi|y) = f(\theta|\phi,y)f(\phi|y)$,

$$f(\theta|\phi,y) \propto exp\left\{-\frac{\phi}{2}\frac{(\theta-m)^2}{C}\right\}.$$

(i) - $(\theta|\phi,y) \sim N(m, C/\phi) = N(m, \sigma^2 C)$

(ii) - $(\phi|y) \sim G\left(\frac{n+1}{2}, \frac{1}{2}\left[d + \frac{(y-a)^2}{R+1}\right]\right)$

Note: $(\theta|\phi, y)$ and $(\phi|y)$ have same form as $(\theta|\phi)$ and (ϕ) (conjugate distribution).

So, $(\theta, \phi) \sim NG\left(m, C, n+1, d + \frac{(y-a)^2}{R+1}\right)$.

Since $y = \theta + \epsilon$ where $(\epsilon|\sigma^2) \sim N(0, \sigma^2)$,

$$
\begin{aligned}
E(y|\sigma^2) &= E(\theta|\sigma^2) + E(\epsilon|\sigma^2) = a \\
V(y|\sigma^2) &= V(\theta|\sigma^2) + V(\epsilon|\sigma^2) = \sigma^2(R+1)
\end{aligned}
$$

So, $y|\phi \sim N(a, \frac{R+1}{\phi})$

$\Rightarrow (y, \phi) \sim NG(a, (R+1), n, d)$

$\Rightarrow y \sim t_n(a, \frac{d}{n}(R+1))$

Non-informative prior: controversy among Bayesians

$\Rightarrow$ large variance

Prediction (or forecast) of a future observation y after observing x

based on the distribution of $(y|x)$

$$f(y|x) = \int f(y, \theta|x) d\theta = \int f(y|\theta)\, \pi(\theta)\, d\theta$$

if y and x are conditionally independent given θ

eg. random sample

In the multivariate case: $\theta = (\theta_1, ..., \theta_p)$

Marginal posterior density of θ_i is given by

$$\pi(\theta_i) = \int \pi(\theta_1, ..., \theta_p)\, d\theta_{-i}$$

where $\theta_{-i} = (\theta_1, ..., \theta_{i-1}, \theta_{i+1}, ..., \theta_p)$

Important posterior summaries:

a) location: mean, mode and median

b) dispersion: variance, standard deviation, precision and curvature at mode

Regression models

Normal linear regression model

Observations $y = (y_1, ..., y_n)'$ described by

$y_i \sim N(x_{i1}\beta_1 + ... + x_{ip}\beta_p, \sigma^2) \quad , \quad i = 1, ..., n$

$x_{i1}, ..., x_{ip}$ - values of the p explanatory variable for the i-th observation

$\beta_1, ..., \beta_p$ - regression coefficients

Model can be written in matrix form

$$y \sim N(X\beta, \sigma^2 I_n)$$

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix} \quad e \quad \beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}$$

Bayesian model completed with conjugate prior for β and $\phi = \sigma^{-2}$

$$\beta|\phi \sim N(b_0, \phi^{-1} B_0) \quad and \quad \phi \sim G\left(\frac{n_0}{2}, \frac{n_0 S_0}{2}\right)$$

i.e. $(\beta, \phi) \sim NG(b_0, B_0, n_0, n_0 S_0)$

Applying Bayes theorem gives posterior

$$\beta|\sigma^2 \sim N(b_1, \sigma^2 B_1) \quad and \quad \sigma^2 \sim IG\left(\frac{n_1}{2}, \frac{n_1 S_1}{2}\right)$$

i.e. $(\beta, \phi) \sim NG(b_1, B_1, n_1, n_1 S_1)$

Details of regression models

Maximum likelihood estimation

ML estimators of $\boldsymbol{\beta}$ and σ^2 are

$$\begin{aligned}\widehat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ \widehat{\sigma}^2 &= \frac{1}{n}\mathbf{y}'(\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y} \\ &= \frac{1}{n}\sum_{i=1}^{n}[y_i-(x_{i1}\widehat{\beta}_1+\ldots+x_{ip}\widehat{\beta}_p)]^2\end{aligned}$$

but $S^2 = n\widehat{\sigma}^2/(n-p)$ is the unbiased estimator of σ^2.

Their sampling distributions are

$$\begin{aligned}\widehat{\boldsymbol{\beta}} &\sim t_{n-p}(\boldsymbol{\beta}, S^2(\mathbf{X}'\mathbf{X})^{-1}) \text{ and} \\ (n-p)\frac{S^2}{\sigma^2} &\sim \chi^2_{n-p} \Leftrightarrow S^2 \sim G\left(\frac{n-p}{2}, \frac{n-p}{2}\sigma^2\right)\end{aligned}$$

Bayesian estimation

Parameters of the NG posterior of $(\boldsymbol{\beta}, \sigma^{-2})$

$$
\begin{aligned}
\mathbf{b}_1 &= \mathbf{B}_1(\mathbf{B}_0^{-1}\mathbf{b}_0 + \mathbf{X}'\mathbf{y}) \\
\mathbf{B}_1^{-1} &= \mathbf{B}_0^{-1} + \mathbf{X}'\mathbf{X} \\
n_1 &= n_0 + n \\
n_1 S_1 &= n_0 S_0 + (n-p)S^2 \\
&+ (\widehat{\boldsymbol{\beta}} - \mathbf{b}_0)'[\mathbf{B}_0 + (\mathbf{X}'\mathbf{X})^{-1}]^{-1}(\widehat{\boldsymbol{\beta}} - \mathbf{b}_0)
\end{aligned}
$$

If prior is non-informative ($n_0 \to 0$ and $\mathbf{B}_0^{-1} \to \mathbf{0}$) then $\mathbf{b}_1 = \widehat{\boldsymbol{\beta}}$, $\mathbf{B}_1 = (\mathbf{X}'\mathbf{X})^{-1}$,

$$n_1 = n - p \text{ and } n_1 S_1 = (n-p)S^2$$

Posterior distributions are

$$
\begin{aligned}
\boldsymbol{\beta} &\sim t_{n-p}(\widehat{\boldsymbol{\beta}}, S^2(\mathbf{X}'\mathbf{X})^{-1}) \text{ and} \\
(n-p)\frac{S^2}{\sigma^2} &\sim \chi^2_{n-p} \Leftrightarrow \sigma^2 \sim IG\left(\frac{n-p}{2}, \frac{n-p}{2}S^2\right)
\end{aligned}
$$

Generalized linear models

Extension of normal models to the exponential family

$$\begin{aligned} f(y_i|\theta_i) &= a(y_i)\, exp\{y_i\theta_i + b(\theta_i)\} \\ E(y_i|\theta_i) &= \mu_i \\ g(\mu_i) &= x_{i1}\beta_1 + ... + x_{ip}\beta_p \quad , \quad i = 1, ..., n \end{aligned}$$

where the link g is differentiable.

Example $y_i|\pi_i \sim bin(n_i, \pi_i)$, $i = 1, ..., n$

probabilities π_i determined by values of x

$$\pi_i = F(\alpha + \beta x_i) \quad , \quad i = 1, ..., n$$

F - any distribution function

Natural prior: $\beta \sim N(b_0, B_0) \rightarrow$ not conjugate

1st. order models

Useful notiation and preliminaries

Inference process is sequential, i.e.,

is redone at each time t (because of y_t)

Process is intialized with info at time $t = 0$ denoted by D_0

Assertions about the future ($t > 0$) are conditional on D_0.

Special interest: predictive distributions of $y_t|D_0$.

At time t: info is concentrated in $D_t \rightarrow$ inference must be based on this set.

Interest: predictive distr of $y_{t+h}|D_t$, $h > 0$.

As time passes the info we have changes.

As we move from t to $t+1$, our info set includes D_t and the new obs. y_{t+1}.

If that is all, $D_{t+1} = \{y_{t+1}, D_t\}$.

Closed system: $D_t = \{D_0, y_1, y_2, \cdots, y_t\}$.

Open system - admits other info entries.

More generally, more info can arrive $\rightarrow$ $D_{t+1} = \{I_{t+1}, D_t\}$

Basic idea: observations fluctuate around a mean.

However, mean **is not static** but subject to (small) time variations.

Model for observations: $(y_t|\mu_t) \sim N(\mu_t, V_t)$, where V_t is **known** for all t.

Time variations μ_t are modelled as random walk, i.e.,

$\mu_t = \mu_{t-1} + \omega_t$ where $\omega_t \sim N(0, W_t)$.

Magnitude of variation depend on evolution errors or disturbances ω_t.

Simplest model but has most features of a dynamic model.

Defining feature: parametric evolution.

Errors ω_t control evolution through their variances W_t.

The larger (smaller) their values, the more erratic (smooth) the trajectory of μ_t.

$E(w_t) = 0$ ensures "local constancy" of μ_t.

If μ is a continuous function of t then 1st order Taylor expansion gives

$$\mu_{t+\Delta} = \mu_t + \textit{higher order terms}.$$

Hence the name 1st order models

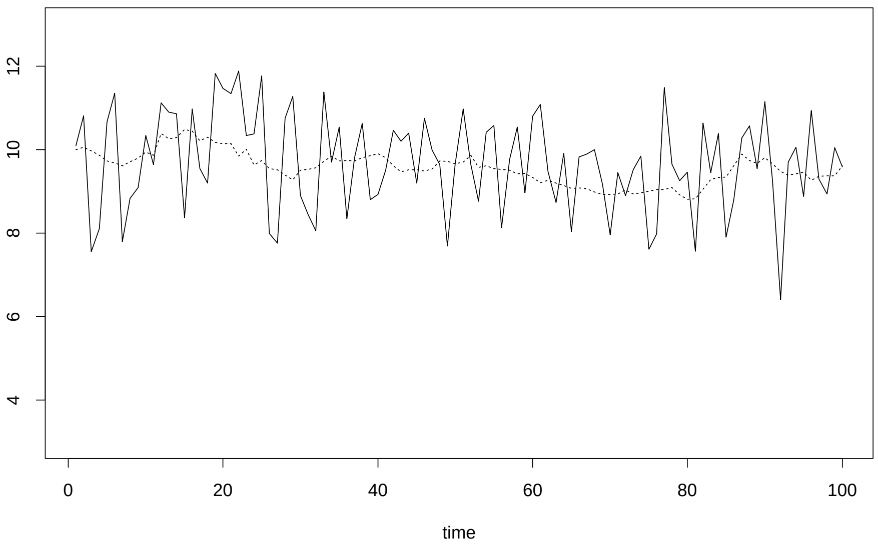
In the model, h.o.t. replaced by errors.

Observational trajectory depends on W_t/V_t:

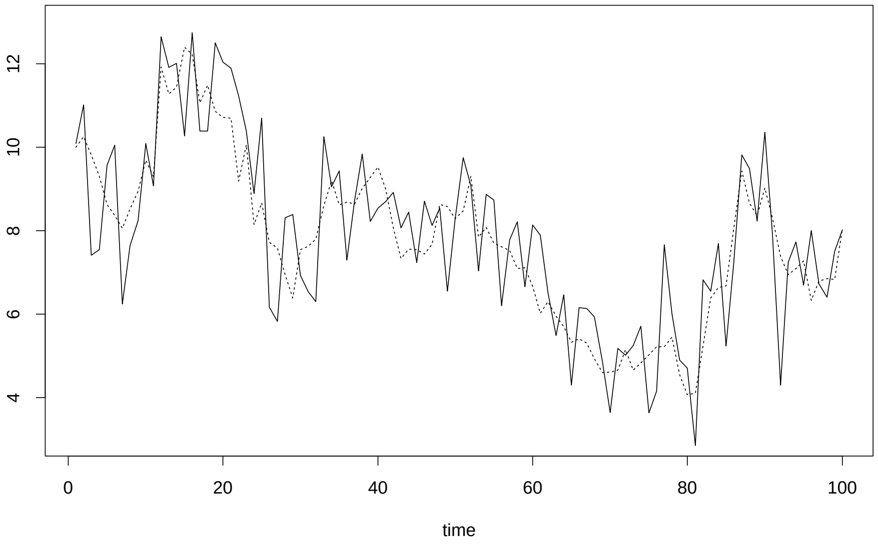
W/V small $\rightarrow$ movements due to obs's.

W/V large $\rightarrow$ movements also due to μ_t.

(a) W/V=0,02



(b) W/V=0,5



Predictive behaviour

Can be studied through the **prediction function**

$f_t(h) = E(y_{t+h}|D_t) = E(\mu_{t+h}|D_t)$.

For 1st order models

$$\begin{aligned} E(\mu_{t+h}|\mu_t) &= E(\mu_{t+h-1} + \omega_{t+h}|\mu_t) = \cdots \\ &= E(\mu_t + \omega_{t+1} + \cdots + \omega_{t+h}|\mu_t) = \mu_t. \end{aligned}$$

So, $f_t(h) = E[E(\mu_{t+h}|\mu_t) \mid D_t] = E[\mu_t|D_t]$,

which does not depend on h

Forecast into the future is constant

(as in simple exponential smoothing).

Model can be formalized through

$$\begin{aligned} \text{Obs. equation: } Y_t &= \mu_t + \nu_t \quad \nu_t \sim N[0, V_t] \\ \text{System equation: } \mu_t &= \mu_{t-1} + \omega_t \quad \omega_t \sim N[0, W_t] \\ \mu_0 \mid D_0 &\sim N[m_0, C_0] \end{aligned}$$

where v_t, ω_t and $\mu_0|D_0$ are independent

Inference system works as follows

$$\mu_{t-1} \mid D_{t-1} \overset{evolution}{\rightarrow} \mu_t \mid D_{t-1} \overset{updating}{\rightarrow} \mu_t \mid D_t$$

posterior prior posterior

$\downarrow$ prediction

$Y_t \mid D_{t-1}$

Evolution done through system equation.

Updating done through **Bayes theorem**.

Prediction done via marginalization

Theorem. For 1st order model, relevant distributions are:

(a) Posterior at $t-1$:

For some mean m_{t-1} and variance C_{t-1},

$$\mu_{t-1} \mid D_{t-1} \sim N[m_{t-1}, C_{t-1}]$$

(b) Prior at t :

$$\mu_t \mid D_{t-1} \sim N[a_t, R_t]$$

where

$$\begin{aligned} a_t &= m_{t-1} \\ R_t &= C_{t-1} + W_t \end{aligned}$$

(c) 1-step-ahead forecast:

$$Y_t \mid D_{t-1} \sim N[f_t, Q_t]$$

where

$$\begin{aligned} f_t &= m_{t-1} \\ Q_t &= R_t + V_t \end{aligned}$$

(d) Posterior at t:

$$\mu_t \mid D_t \sim N[m_t, C_t]$$

where

$$\begin{aligned} m_t &= m_{t-1} + A_t e_t \qquad A_t = R_t/Q_t \qquad e_t = Y_t - f_t \\ C_t &= R_t - A_t^2 Q_t. \end{aligned}$$

Proof: Based on basic density results.

(b) Prior at t:

$$p(\mu_t|D_{t-1}) = \int p(\mu_t, \mu_{t-1}|D_{t-1}) d\mu_{t-1}$$
$$= \int p(\mu_{t-1}|D_{t-1}) p(\mu_t|\mu_{t-1}, D_{t-1}) d\mu_{t-1}$$

(c) 1-step-ahead forecast:

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

(d) Posterior at t:

$$p(\mu_t|D_t) \propto p(Y_t|\mu_t, D_{t-1}) p(\mu_t|D_{t-1})$$

Remarks:

(i) Since $f_t = f_{t-1}(1)$, e_t is the 1-step-ahead **forecast error**

(ii) One can rewrite m_t as $m_{t-1}+$ $A_t(y_t - m_{t-1}) = A_t y_t + (1 - A_t)m_{t-1}$. So, A_t is the adaptive **weight** given to the most recent obs. y_t.

h-steps-ahead forecast distributions are

$$y_{t+h}|D_t \quad \sim \quad N(f_t(h), Q_t(h)) \text{ where}$$

$$f_t(h) = m_t \qquad Q_t(h) = C_t + \sum_{j=1}^{h} W_{t+j} + V_{t+h}$$

Model is constant if $V_t = V$ e $W_t = W$, $\forall t$.

Model is closed if $D_t = \{y_t, D_{t-1}\}$, $\forall t$. (seen)

Properties of closed, constant model

(i) When $t \to \infty$, $A_t \to A$ and $C_t \to AV$

$A = r(\sqrt{1 + 4/r} - 1)/2$ and $r = W/V$.

Limiting behaviour determined by r.

(ii) For large t, $m_t \approx Ay_t + (1 - A)m_{t-1}$.

Vague prior ($C_0^{-1} \approx 0$) $\to A_t$ monotonically decreases with t.

Informative prior ($C_0^{-1} \approx 0$) $\to A_t$ monotonically increases with t

m_t is an estimator of μ_t that coincides for large t with those suggested by simple exponential smoothing (Holt) and general exponential smoothing (Brown). The adaptive weight has the same interpretation.

(iii) Since $m_t = m_{t-1} + A_t e_t$ and $m_{t-1} = y_t - e_t$,

$$y_t = m_{t-1} + e_t \text{ and } y_{t-1} = m_{t-2} + e_{t-1} \Rightarrow$$

$$y_t - y_{t-1} = m_{t-1} - m_{t-2} + e_t - e_{t-1}$$

$$y_t - y_{t-1} = A_{t-1}e_{t-1} + e_t - e_{t-1}$$

$$= e_t - (1 - A_{t-1})e_{t-1}$$

For large t and $\theta = 1 - A$, $y_t - y_{t-1} \approx e_t - \theta e_{t-1}$

Since at t, $e_t | D_{t-1} \sim N(0, Q_t) \approx N(0, Q)$, for large t.

y_t admits the ARIMA(0,1,1) representation assintotically

(iv) Since $R_t = C_{t-1} + W$, as t gets large $R = C + W$

Also, $V^{-1} + R_t^{-1} = C_t^{-1}$ and in the limit $V^{-1} + R^{-1} = C^{-1}$.

$V = C/A \Rightarrow AC^{-1} + R^{-1} = C^{-1} \Rightarrow R^{-1} = C^{-1}(1 - A) \Rightarrow R = C/(1 - A)$.

This leads to $W = \frac{A}{1-A}C$.

So, W implies an inflation of $100\frac{A}{1-A}\%$ of the system variance.

As limit is quickly reached, one may establish a constant rate as a rule.

Making $\delta = 1 - A$ as the constant **discount factor** $\rightarrow R_t = C_{t-1}/\delta$ or $R_t^{-1} = \delta C_{t-1}^{-1}$.

Taking inverse variance as precision, only $100\delta\%$ of info is kept through time. The corresponding values of W_t would be given by

$$R_t = C_{t-1} + W_t \Rightarrow C_{t-1}/\delta = C_{t-1} + W_t \Rightarrow$$
$$W_t = C_{t-1}(\delta^{-1} - 1)$$

This concept will be extended for general models.

Dynamic Linear Models (DLM)

1st order model and regression model are special cases of a more general structure.

Model is again described by two equations

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

$$\text{System equation: } \boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t \quad \boldsymbol{\omega}_t \sim N[0, \mathbf{W}_t]$$

Initial info: $(\boldsymbol{\theta}_0 \mid D_0) \sim N[\mathbf{m}_0, \mathbf{C}_0]$

- y_t (scalar) observation;
- $\boldsymbol{\theta}_t$ $n \times 1$ vector of parameters;
- $\mathbf{F}_t$ $n \times 1$ vector of constants ;
- $\mathbf{G}_t$ $n \times n$ transition matrix;
- V_t observational variance;
- $\mathbf{W}_t$ $n \times n$ covariance matrix;

Examples : (i) 1st order model:

$F_t = 1$, $G_t = 1$, $\theta_t = \mu_t$

(ii) Dynamic regression model:

$\mathbf{F}_t' = (1, x_{t,2}, \cdots, x_{n,t})$, $\mathbf{G} = \mathbf{I}_n$, $\boldsymbol{\theta}_t' = (\beta_{t,1} \cdots, \beta_{t,n})$

where $\mathbf{I}_n$ is the identity matrix of order n.

The inference cycle remains as before

$$\boldsymbol{\theta}_{t-1} \mid D_{t-1} \overset{evolution}{\rightarrow} \boldsymbol{\theta}_t \mid D_{t-1} \overset{updating}{\rightarrow} \boldsymbol{\theta}_t \mid D_t$$

posterior prior posterior

$\downarrow$ prediction

$Y_t \mid D_{t-1}$

Theorem. For general DLM model, relevant distributions are:

(a) Posterior at $t-1$:

For some mean $\mathbf{m_{t-1}}$ and variance matrix $\mathbf{C}_{t-1}$,

$$\boldsymbol{\theta}_{t-1} \mid D_{t-1} \sim N[\mathbf{m}_{t-1}, \mathbf{C}_{t-1}]$$

(b) Prior at t:

$$\boldsymbol{\theta}_t \mid D_{t-1} \sim N[\mathbf{a}_t, \mathbf{R}_t], \text{ where}$$

$$\begin{aligned} \mathbf{a}_t &= \mathbf{G}_t \mathbf{m}_{t-1} \\ \mathbf{R}_t &= \mathbf{G}_t \mathbf{C}_{t-1} \mathbf{G}_t' + \mathbf{W}_t \end{aligned}$$

(c) 1-step-ahead forecast

$$Y_t \mid D_{t-1} \sim N[f_t, \mathbf{Q}_t], \text{ where}$$

$$\begin{aligned} f_t &= \mathbf{F}_t' \mathbf{a}_t \\ \mathbf{Q}_t &= \mathbf{F}_t' \mathbf{R}_t \mathbf{F}_t + V_t \end{aligned}$$

(d) Posterior at t:

$$\boldsymbol{\theta}_t \mid D_t \sim N[\mathbf{m}_t, \mathbf{C}_t], \text{ where}$$

$$\begin{aligned} \mathbf{m}_t &= \mathbf{a}_t + \mathbf{A}_t e_t \qquad \mathbf{A}_t = \mathbf{R}_t \mathbf{F}_t \mathbf{Q}_t^{-1} \qquad e_t = Y_t - f_t \\ \mathbf{C}_t &= \mathbf{R}_t - \mathbf{A}_t \mathbf{A}_t' Q_t. \end{aligned}$$

Proof: Textbook

Similar results obtained for r-vector $\mathbf{Y}_t$

Predictive distributions

Definition The forecast function $f_t(k)$ is

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

where

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

is the **mean response function**.

Theorem. For every t and $k \geq 1$, the k-steps-ahead distributions of $\boldsymbol{\theta}_{t+k}$ and Y_{t+k} given D_t are

(a) System distr.: $(\boldsymbol{\theta}_{t+k} \mid D_t) \sim N[\mathbf{a}_t(k), \mathbf{R}_t(k)]$,

(b) Forecast distr.: $(Y_{t+k} \mid D_t) \sim N[f_t(k), Q_t(k)]$, Moments are recursively defined as

$$\begin{aligned} f_t(k) &= \mathbf{F}_t' \mathbf{a}_t(k) \quad and \\ Q_t(k) &= \mathbf{F}_t' \mathbf{R}_t(k) \mathbf{F}_t + V_{t+k} \quad , \quad where \\ \mathbf{a}_t(k) &= \mathbf{G}_{t+k} \mathbf{a}_t(k-1) \quad and \\ \mathbf{R}_t(k) &= \mathbf{G}_{t+k} \mathbf{R}_t(k-1) \mathbf{G}_{t+k}' + W_{t+k} \quad , \end{aligned}$$

with initial values $\mathbf{a}_t(0) = \mathbf{m}_t$, $\mathbf{R}_t(0) = \mathbf{C}_t$.

Corollary. If $\mathbf{G}_t = \mathbf{G}$ is constant, for all t

$$\begin{aligned} \mathbf{a}_t(k) &= \mathbf{G}^k \mathbf{m}_t \\ f_t(k) &= \mathbf{F}_{t+k}' \mathbf{G}^k \mathbf{m}_t \end{aligned}$$

If, in addition, $\mathbf{F}_t = \mathbf{F}$ for all t,

$$f_t(k) = \mathbf{F}' \mathbf{G}^k \mathbf{m}_t.$$

Importance of the result:

forecasts are governed by powers of $\mathbf{G}$.

2nd. order (linear growth) models

Extension over 1st order models:

allow Taylor expansion up to 2nd order

$\mu_{t+\Delta t} = \mu_t + \Delta t \, \mu'_t +$ *h.o.t.* Δt.

In addition to the level μ, the growth β can also be defined.

The difference equation becomes

$\mu_{t+1} = \mu_t + \beta_{t+1} + \omega_{1,t+1}$

with $\omega_{1,t+1} \sim N(0, W_{1,t+1})$

$\omega_{1,t+1}$ is the disturbance replacing h.o.t.

Evolution of the growth is modelled as

$\beta_{t+1} = \beta_t + \omega_{2,t+1}$

with $\omega_{2,t+1} \sim N(0, W_{2,t+1})$ independently of $\omega_{1,t+1}$.

Growth also subject to local fluctuations.

System equation has now 2 components.

Complete model is

$$\begin{aligned} \text{Obs. equation: } y_t &= \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} \mu_t \\ \beta_t \end{pmatrix} + v_t \\ \text{System equation: } \begin{pmatrix} \mu_t \\ \beta_t \end{pmatrix} &= \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_{t-1} \\ \beta_{t-1} \end{pmatrix} + \boldsymbol{\omega}_t \\ \text{Initial info: } \begin{pmatrix} \mu_0 \\ \beta_0 \end{pmatrix} &\sim N\left(\begin{pmatrix} m_0 \\ b_0 \end{pmatrix}, \mathbf{C}_0 \right) \end{aligned}$$

where

$$\boldsymbol{\omega}_t \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} W_1 + W_2 & W_2 \\ W_2 & W_2 \end{pmatrix} \right)$$

Alternatively, we could have $V(\boldsymbol{\omega}_t) = diag(W_{1,t}, W_{2,t})$

Model is useful for series with "locally linear" trends.

"growth" β may also be negative (despite name).

Predictive behaviour analysed through

$$E\left[y_{t+h}|\begin{pmatrix}\mu_t\\ \beta_t\end{pmatrix}\right] = E\left[\mu_{t+h}|\begin{pmatrix}\mu_t\\ \beta_t\end{pmatrix}\right] + 0$$

But

$$\begin{aligned}\mu_{t+h} &= \mu_{t+h-1} + \beta_{t+h-1} + \textit{sum of errors}\\ &= \mu_{t+h-2} + 2\beta_{t+h-2} + \textit{sum of errors}\\ &= \mu_t + h\beta_t + \textit{sum of errors}\end{aligned}$$

Since all errors are 0 mean,

$$E\left[y_{t+h}|\begin{pmatrix}\mu_t\\ \beta_t\end{pmatrix}\right] = \mu_t + h\beta_t$$

Forecast function is given by

$$f_t(h) = E[y_{t+h}|D_t] = E[\mu_t + h\beta_t|D_t] = m_t + hb_t$$

which is a linear function of h, as expected.

Inference process is as before.

Using normal theory and $D_t = \{y_t, D_{t-1}\}$ we have that

$$\begin{pmatrix} \mu_t \\ \beta_t \end{pmatrix} |D_t \sim N(\mathbf{m}_t, \mathbf{C}_t) \text{ where}$$

$\mathbf{m}_t = \mathbf{a}_t + \mathbf{S}_t Q_t^{-1} e_t$ and $\mathbf{C}_t = \mathbf{R}_t - \mathbf{S}_t Q_t^{-1} \mathbf{S}_t'$

where $e_t = y_t - f_t$ and $\mathbf{S}_t = \mathbf{R}_t \mathbf{F}_t$.

$$\mathbf{m}_t = \begin{pmatrix} m_t \\ b_t \end{pmatrix} = \begin{pmatrix} m_{t-1} + b_{t-1} + A_{t,1} e_t \\ b_{t-1} + A_{t,2} e_t \end{pmatrix}$$

Since $y_t = f_t + e_t$ and $f_t = m_{t-1} + b_{t-1}$, we can rewrite

$$\begin{aligned} y_t &= m_{t-1} + b_{t-1} + e_t \\ m_t &= m_{t-1} + b_{t-1} + A_{t,1} e_t \\ b_t &= b_{t-1} + A_{t,2} e_t \end{aligned}$$

Taking diferences twice in the first equation and using next ones

$$y_t - 2y_{t-1} + y_{t-2} = e_t + \beta_{t,1} e_{t-1} + \beta_{t,2} e_{t-2}$$

where $\beta_{t,1} = -(2 - A_{t,1} - A_{t,2})$ e $\beta_{t,2} = 1 - A_{t,1}$.

As in 1st order models, limiting results are obtained for the constant model ($V_t = V$ e $W_t = W$)

When $t \longrightarrow \infty$, $\mathbf{A}_t \longrightarrow \mathbf{A} = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix}$

In the limit, model approaches constant form

$$y_t - 2y_{t-1} + y_{t-2} = e_t + \beta_1 e_{t-1} + \beta_2 e_{t-2}$$

which is an ARIMA(0,2,2) form with errors given by forecast errors.

Also, parameter estimates are updated as

$$\begin{aligned} m_t &\approx m_{t-1} + b_{t-1} + A_1 e_t \\ &= A_1 y_t + (1 - A_1)(m_{t-1} + b_{t-1}) \\ b_t &\approx b_{t-1} + A_2 e_t \end{aligned}$$

Holt's biparametric exponential smoothing updates estimates as

$$\begin{aligned} m_t &= \alpha y_t + (1-\alpha)(m_{t-1} + b_{t-1}) \\ b_t &= \gamma(m_t - m_{t-1}) + (1-\gamma)b_{t-1}. \end{aligned}$$

So, $m_t = m_{t-1} + b_{t-1} + \alpha e_t$ and $b_t = b_{t-1} + \gamma(m_t - m_{t-1} - b_{t-1}) = b_{t-1} + \alpha\gamma e_t$.
Taking $A_1 = \alpha$ and $A_2 = \alpha\gamma$ gives Holt's equations as limits for constant model.

General exponential smoothing in a line estimates μ and β through minimization of

$$S(\mu, \beta) = \sum_{t=1}^{N} \delta^{N-t}[y_t - \mu - \beta(N-t)]^2$$

Minimizers for μ and β updated through

$$\begin{aligned} m_t &= m_{t-1} + b_{t-1} + (1-\delta^2)e_t \\ b_t &= b_{t-1} + (1-\delta)^2 e_t \end{aligned}$$

One can define

$$\mathbf{G} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{P}_t = \mathbf{G}\mathbf{C}_{t-1}\mathbf{G}'$$

So,

$$V\left[\begin{pmatrix} \mu_t \\ \beta_t \end{pmatrix} | D_{t-1}\right] = \mathbf{R}_t = \mathbf{P}_t + V(\boldsymbol{\omega}_t)$$

Main role of the system disturbances $\boldsymbol{\omega}_t$ is to "increase" system variance since they are generally 0 mean.

Idea of discount factors can again be used here to help specification of $V(\boldsymbol{\omega}_t)$.

Polynomial trend models

1st and 2nd order models are special cases.

n-order polynomial trend model defined as

$$
\begin{aligned}
y_t &= \theta_{t,1} + v_t \\
\theta_{t,j} &= \theta_{t-1,j} + \theta_{t,j+1} + \delta\theta_{t,j} \quad (j = 1, \cdots, n-1) \\
\theta_{t,n} &= \theta_{t-1,n} + \delta\theta_{t,n}
\end{aligned}
$$

$\theta_{t,j}$ changes with respect to its predecessor $\theta_{t-1,j}$ by $\theta_{t-1,j+1}$.

Examples:

(i) quadratic trend DLM:

$$\begin{aligned} y_t &= \mu_t + v_t \\ \mu_t &= \mu_{t-1} + \beta_t + \delta\mu_t \\ \beta_t &= \beta_{t-1} + \gamma_t + \delta\beta_t \\ \gamma_t &= \gamma_{t-1} + \delta\gamma_t \end{aligned}$$

$\boldsymbol{\theta}_t = (\mu_t, \beta_t, \gamma_t)'$, μ_t represents level, β_t growth and γ_t change in growth.

Forecast function is

$$f_t(k) = m_t + kb_t + k(k+1)g_t/2,$$

where $E(\boldsymbol{\theta}_t|D_t) = \mathbf{m_t} = (m_t, b_t, g_t)'$.

(ii) 2nd order DLM: $\boldsymbol{\theta}_t = (\mu_t, \beta_t)'$

(iii) 1st order DLM: $\boldsymbol{\theta}_t = \mu_t$

Seasonal component

Cyclical behaviour is common to many time series.

Here we consider only the description.

Possible explanation may be given by covariates.

In what follows, let $g(t)$ be any real function defined for non-negative integers.

Definitions:

1) $g(t)$ is cyclical or periodic if, $\exists p \geq 1$ such that $g(t + np) = g(t)$, $\forall t \geq 0$ and $n \geq 0$.

2) The smallest p is the period of $g(.)$.

3) $g(.)$ exhibits a complete cycle in any interval containing p consecutive points, such as $[t, t+p-1]$, $\forall t \geq 0$.

4) The seasonal factors of $g(.)$ are the p values of any complete cycle

$$\theta_j = g(j) \quad j = 0, 1, 2, \cdots, p-1$$

For $t > 0$, $g(t) = g(j)$ where j is the remainder of the division of t by p. $(j = p \mid t)$

5) The vector of seasonal factors at t is

$$\boldsymbol{\theta}_t = (\theta_j, \theta_{j+1}, \cdots, \theta_{p-1}, \theta_0, \cdots, \theta_{j-1})'$$

where $j = p \mid t$.

For any n and $k = np$: $\boldsymbol{\theta}_k = (\theta_0, \cdots, \theta_{p-1})'$.

Characterization of seasonal factors

Two routes: indicators or trigonometric functions.

If model only has seasonal component mean response function μ_t = 1st component of $\boldsymbol{\theta}_t$

$\mu_t = \mathbf{E}_p' \boldsymbol{\theta}_t$ where $\mathbf{E}_p' = (1, \underbrace{0, 0, \cdots, 0}_{\text{(p-1) terms}})$.

If time t corresponds to j-th period, $t+1$ corresponds $(j+1)$-th. Permuting $\boldsymbol{\theta}_t$ by one position gives $\boldsymbol{\theta}_{t+1}' = (\theta_{j+1}, \cdots, \theta_p, \theta_1, \cdots, \theta_j)$ and $\mu_{t+1} = \mathbf{E}_p' \boldsymbol{\theta}_{t+1}$.

Passage from $\boldsymbol{\theta}_t$ to $\boldsymbol{\theta}_{t+1}$ is done through matrix

$$\mathbf{P}_p = \begin{pmatrix} \mathbf{0} & \mathbf{I}_{p-1} \\ 1 & \mathbf{0}' \end{pmatrix}.$$

So, $\boldsymbol{\theta}_{t+1} = \mathbf{P}_p \boldsymbol{\theta}_t$.

$\mathbf{P}_p$ satisfies $\mathbf{P}_p^{k+np} = \mathbf{P}_p^k$ and, in particular, $\mathbf{P}_p^{np} = \mathbf{I}_p$.

Formulation leads to the dynamic model

$$\begin{array}{rcll} \text{Obs. equation: } y_t & = & \mathbf{E}_p' \boldsymbol{\theta}_t + \nu_t & \nu_t \sim N[0, V_t] \\ \text{System equation: } \boldsymbol{\theta}_t & = & \mathbf{P}_p \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t & \boldsymbol{\omega}_t \sim N[0, \mathbf{W}_t] \\ \text{Initial info: } (\boldsymbol{\theta}_0 | D_0) & \sim & N(\mathbf{m_0}, \mathbf{C}_0) & \end{array}$$

Remarks:

(i) If $\mathbf{W}_t = c\mathbf{I}_p$, errors $\boldsymbol{\omega}_t$ have independent components and problem reduces to the analysis of p 1st order models, if prior variance $\mathbf{C}_0 = diag(c_1, \cdots, c_p)$.

No passage of information through levels.

(ii) The h-steps-ahead forecast function is

$$
\begin{aligned}
f_t(h) &= E(y_{t+h}|D_t) \\
&= E(\mu_{t+h}|D_t) \\
&= E(\mathbf{E}_p'\boldsymbol{\theta}_{t+h}|D_t) \\
&= \mathbf{E}_p'\mathbf{P}_p^h E(\boldsymbol{\theta}_t|D_t) \\
&= \mathbf{E}_p'\mathbf{P}_p^h \mathbf{m}_t \\
&= \mathbf{E}_p'(m_{t,h+1}, \cdots, m_{t,p}, m_{t,1}, \cdots, m_{t,h})' \\
&= m_{t,h+1}
\end{aligned}
$$

Usually, seasonality is modelled by an overall level μ_t and seasonal variations around it. In this case, seasonal factors satisfy $\mathbf{1}'\boldsymbol{\theta}_t = \sum_{j=1}^{p} \theta_{t,j} = 0$.

The mean response at time t is given by

$$\begin{aligned} \mu_t + \theta_{t,j} &= \mu_t + \mathbf{E}_p'\boldsymbol{\theta}_t \\ &= (1, \mathbf{E}_p') \begin{pmatrix} \mu_t \\ \boldsymbol{\theta}_t \end{pmatrix} = \mathbf{F}_t'\boldsymbol{\beta}_t. \end{aligned}$$

This strategy imposes restrictions over $\boldsymbol{\theta}_t$:

i) Initial prior: Since $\boldsymbol{\theta}_0 | D_0 \sim N(\mathbf{m}_0, \mathbf{C}_0)$ and $\mathbf{1}'\boldsymbol{\theta}_0 = 0$,

$$\begin{aligned} 0 &= \mathbf{1}'\mathbf{m}_0 \\ 0 &= \mathbf{1}'\mathbf{C}_0\mathbf{1} \quad \Rightarrow \quad \mathbf{C}_0\mathbf{1} = 0 \end{aligned}$$

If initial spec. does not satisfy conditions above, it can be imposed.

Staring from an incorrect $\boldsymbol{\theta}_0|D_0 \sim N(\mathbf{m}_0^*, \mathbf{C}_0^*)$, gives

$(\boldsymbol{\theta}_0|D_0, \mathbf{1}'\boldsymbol{\theta}_0 = 0) \sim N(\mathbf{m}_0, \mathbf{C}_0)$, with

$$
\begin{aligned}
\mathbf{m}_0 &= \mathbf{m}_0^* - \mathbf{A}(\mathbf{1}'\mathbf{m}_0^*)/(\mathbf{1}'\mathbf{C}_0^*\mathbf{1}) \\
\mathbf{C}_0 &= \mathbf{C}_0^* - \mathbf{A}\mathbf{A}'/(\mathbf{1}'\mathbf{C}_0^*\mathbf{1}),
\end{aligned}
$$

onde $\mathbf{A} = Cov(\boldsymbol{\theta}_0, \mathbf{1}'\boldsymbol{\theta}_0) = \mathbf{C}_0\mathbf{1}$

ii) Restrictions are preserved if $\mathbf{1}'\boldsymbol{\omega}_t = 0$.

Seasonal model can be redefined as

$$\begin{aligned}
\text{Obs. equation: } y_t &= \mu_t + \theta_{t,j} + \nu_t \quad \nu_t \sim N(0, V_t) \\
&= \mu_t + \mathbf{E}_p' \boldsymbol{\theta}_t + \nu_t \\
&= \mathbf{F}' \boldsymbol{\beta}_t + \nu_t \\
\text{System equation: } \mu_t &= \mu_{t-1} + \omega_{0,t},\ \omega_{0,t} \sim N(0, W_{0,t}) \\
\boldsymbol{\theta}_t &= \mathbf{P}_p \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t\, \boldsymbol{\omega}_t \sim N(0, \mathbf{W}_t) \\
\text{Initial info: } (\mu_0 | D_0) &\sim N(m_{00}, \mathbf{C}_{00}) \\
(\boldsymbol{\theta}_0 | D_0) &\sim N(\mathbf{m}_0, \mathbf{C}_0)
\end{aligned}$$

Additional restriction: $\mathbf{1}' \boldsymbol{\theta}_t = 0$, $\forall t$.

Assume again that

$E[\boldsymbol{\beta}_t|D_t] = (m_{t,0}, m_{t,1}, \cdots, m_{t,p})'$

where $\sum_{j=1}^{p} m_{t,j} = 0$.

The forecast function is

$$
\begin{aligned}
f_t(h) &= E(y_{t+h}|D_t) \\
&= E(\mu_{t+h} + \mathbf{E}_p' \boldsymbol{\theta}_{t+h}|D_t) \\
&= E(\mu_{t+h}|D_t) + \mathbf{E}_p' E(\boldsymbol{\theta}_{t+h}|D_t) \\
&= E\left[\mu_t + \sum_{j=1}^{h} \omega_{0,t+j}|D_t\right] \\
&+ \mathbf{E}_p' E\left[\mathbf{P}_p^h \boldsymbol{\theta}_t + \sum_{j=1}^{h} \mathbf{P}_p^{h-j} \boldsymbol{\omega}_{t+j}|D_t\right] \\
&= E(\mu_t|D_t) + \mathbf{E}_p' E(\mathbf{P}_p^h \boldsymbol{\theta}_t|D_t) \\
&= m_{t,0} + m_{t,h+1}
\end{aligned}
$$

Trigonometric representation

Seasonal factors θ_j, $j = 1, \cdots, p$ for a cycle of period p can also be represented by

$$\theta_j = \begin{cases} a_0 + \sum_{r=1}^{q-1} [a_r \cos(\omega r t) + b_r \sin(\omega r t)], & \text{p odd,} \\ a_0 + \sum_{r=1}^{q-1} [a_r \cos(\omega r t) + b_r \sin(\omega r t)] + a_q cos(\pi t), \end{cases}$$

for p even,

where time t corresponds to the j-th time in a cycle, $q = [p + 1/2]$ and $\omega = 2\pi/p$.

Each component of sum is an harmonic.

It can be rewritten as A_r cos $(\omega r t + \phi_r)$, where

i) A_r - amplitude of harmonic of order r;

ii) ϕ_r - phase of harmonic of order r, Advantage: parameter economy by exclusion of irrelevant harmonics (A_r small)

Example: In many monthly series in economics and meteorology, 1st harmonic is enough.

To fit harmonics into dynamic model consider a single harmonic ($r = 1$).

$$
\begin{aligned}
\theta_j &= a_0 + a \cos(\omega t) + b \sin(\omega t) \\
&= \begin{pmatrix} 1 & \cos(\omega t) & \sin(\omega t) \end{pmatrix} \begin{pmatrix} a_0 \\ a \\ b \end{pmatrix}
\end{aligned}
$$

Define $\beta_{1,t} = a \cos(\omega t) + b \sin(\omega t)$ and

$\gamma_{1,t} = -a \sin(\omega t) + b \cos(\omega t)$.

The relation

$$\begin{pmatrix} a_0 \\ \beta_1 \\ \gamma_1 \end{pmatrix} \to \begin{pmatrix} a_0 \\ a \\ b \end{pmatrix} \text{ is 1-to-1.}$$

$$\theta_j = a_0 + \beta_{1,t} = \begin{pmatrix} 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} a_0 \\ \beta_1 \\ \gamma_1 \end{pmatrix}_t = \begin{pmatrix} 1 & 1 & 0 \end{pmatrix} \boldsymbol{\beta}_t$$

Likewise,

$$\begin{aligned} \theta_{j+1} &= a_0 + a \cos \omega(t+1) + b \sin \omega(t+1) \\ &= a_0 + \beta_{1,t+1} \end{aligned}$$

Define

$$\mathbf{G}_1 = \begin{pmatrix} \cos \omega & \sin \omega \\ -\sin \omega & \cos \omega \end{pmatrix} \text{ and } \mathbf{G} = diag(1, \mathbf{G}_1)$$

So,

$$
\begin{aligned}
\mathbf{G}\boldsymbol{\beta}_t &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\,\omega & \sin\,\omega \\ 0 & -\sin\,\omega & \cos\,\omega \end{pmatrix} \begin{pmatrix} a_0 \\ \beta_{1,t} \\ \gamma_{1,t} \end{pmatrix} \\
&= \begin{pmatrix} a_0 \\ \beta_{1,t}\,\cos\,\omega + \gamma_{1,t}\,\sin\,\omega \\ -\beta_{1,t}\,\sin\,\omega + \gamma_{1,t}\,\cos\,\omega \end{pmatrix}
\end{aligned}
$$

But $\beta_{1,t}\,\cos\,\omega + \gamma_{1,t}\,\sin\,\omega$

$$
\begin{aligned}
&= [a\,\cos\,\omega t + b\,\sin\,\omega t]\,\cos\,\omega \\
&+ [(-a\,\sin\,\omega t) + b\,\cos\,\omega t]\,\sin\,\omega \\
&= \beta_{1,t+1}.
\end{aligned}
$$

Analogously, $-\beta_{1,t}\,\sin\,\omega + \gamma_{1,t}\,\cos\,\omega = \gamma_{1,t+1}$.

Then, $\mathbf{G}\boldsymbol{\beta}_t = \begin{pmatrix} a_0 \\ \beta_{1,t+1} \\ \gamma_{1,t+1} \end{pmatrix} = \boldsymbol{\beta}_{t+1}$ and
$\theta_{j+1} = \begin{pmatrix} 1 & 1 & 0 \end{pmatrix} \boldsymbol{\beta}_{t+1}$

This reasoning can be extended to all harmonics since

$$\theta_j = a_0 + \sum_{r=1}^{q-1} [a \ \cos\ (\omega r t) + b_r \ \sin\ (\omega r t)].$$

Set
$$\begin{aligned} \beta_{r,t} &= a_r \ \cos(\omega r t) + b_r \ \sin(\omega r t)\ ,\ r = 1, \cdots, q-1 \\ \gamma_{r,t} &= -a_r \ \sin(\omega r t) + b_r \ \cos(\omega r t) \end{aligned}$$

$$\begin{aligned} \boldsymbol{\beta}'_t &= (\ a_0 \quad \beta_{1,t} \quad \gamma_{1,t} \quad \beta_{2,t} \quad \gamma_{2,t} \quad \cdots \quad \beta_{q-1,t} \quad \gamma_{q-1,t}\ ) \\ \mathbf{F}'_t &= (\ 1 \quad 1 \quad 0 \quad 1 \quad 0 \quad \cdots \quad 1 \quad 0\ ) \end{aligned}$$

Then, $\theta_j = \mathbf{F}'_t \boldsymbol{\beta}_t$ and $\boldsymbol{\beta}_{t+1} = \mathbf{G}\boldsymbol{\beta}_t$ where

$$\begin{aligned} \mathbf{G} &= diag(1, \mathbf{G}_1, \cdots, \mathbf{G}_{q-1}), \\ \mathbf{G}_r &= \begin{pmatrix} \cos\ \omega r & \sin\ \omega r \\ -\ \sin\ \omega r & \cos\ \omega r \end{pmatrix} \quad r = 1, \cdots, q-1 \end{aligned}$$

Dynamic nature may be incorporated through system errors.

Model can be written as

$$\begin{aligned} \text{Obs. equation: } y_t &= \mathbf{F}_t' \boldsymbol{\beta}_t + \nu_t \\ \text{System equation: } \boldsymbol{\beta}_{t+1} &= \mathbf{G}\boldsymbol{\beta}_t + \boldsymbol{\omega}_{t+1} \\ \text{Initial info: } (\boldsymbol{\beta}_0 | D_0) &\sim N(\mathbf{m_0}, \mathbf{C}_0) \end{aligned}$$

where

$$\mathbf{G} = \begin{cases} diag(1, \mathbf{G}_1, \cdots, \mathbf{G}_{q-1}) \text{ , p odd;} \\ diag(1, \mathbf{G}_1, \cdots, \mathbf{G}_{q-1}, -1) \text{ , p even.} \end{cases}$$

$$\mathbf{F} = \begin{cases} \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & \cdots & 1 & 0 \end{pmatrix}, \text{ p odd;} \\ \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & \cdots & 1 & 0 & 1 \end{pmatrix}, \text{ p even.} \end{cases}$$

Selection of harmonics

Complete description involves $q-1$ complete harmonics.

They may not all be needed.

Sometimes with monthly data, only the 1st (anual) and 4th (quarterly) harmonics are needed.

If we do not know which ones to include, look at its amplitude.

To do that for the r-th harmonic, observe $\beta_{r,t}$ and $\gamma_{r,t}$ and obtain $\sqrt{\beta_{r,t}^2 + \gamma_{r,t}^2} = \sqrt{a_r^2 + b_r^2} = A_r$, its amplitude. Estimates of $\beta_{r,t}$ and $\gamma_{r,t}$ provide evidence about *significance* of A_r.

Model Superposition

The linear structure of the models allow many components to be combined into a single model

Consider models 1 and 2 given by

$$
\begin{aligned}
y_{1,t} &= \mathbf{F}'_{1,t}\boldsymbol{\theta}_{1,t} + v_{1,t}, \quad v_{1,t} \sim N(0, V_{1,t}) \\
\boldsymbol{\theta}_{1,t} &= \mathbf{G}_{1,t}\boldsymbol{\theta}_{1,t-1} + \boldsymbol{\omega}_{1,t} \quad \boldsymbol{\omega}_{1,t} \sim N(\mathbf{0}, \mathbf{W}_{1,t})
\end{aligned}
$$

and

$$
\begin{aligned}
y_{2,t} &= \mathbf{F}'_{2,t}\boldsymbol{\theta}_{2,t} + v_{2,t}, \quad v_{2,t} \sim N(0, V_{2,t}) \\
\boldsymbol{\theta}_{2,t} &= \mathbf{G}_{2,t}\boldsymbol{\theta}_{2,t-1} + \boldsymbol{\omega}_{2,t} \quad \boldsymbol{\omega}_{2,t} \sim N(\mathbf{0}, \mathbf{W}_{2,t})
\end{aligned}
$$

If y_t is the sum of the 2 series then

$$
\begin{aligned}
y_t &= y_{1,t} + y_{2,t} \\
&= \mathbf{F}'_{1,t}\boldsymbol{\theta}_{1,t} + \mathbf{F}'_{2,t}\boldsymbol{\theta}_{2,t} + v_{1,t} + v_{2,t}.
\end{aligned}
$$

Defining

$$\mathbf{F}'_t = (\mathbf{F}'_{1,t}, \mathbf{F}'_{2,t}),\ \boldsymbol{\theta}_t = \begin{pmatrix} \boldsymbol{\theta}_{1,t} \\ \boldsymbol{\theta}_{2,t} \end{pmatrix},$$

$v_t = v_{1,t} + v_{2,t}$ and $V_t = V_{1,t} + V_{2,t}$ gives

$y_t = \mathbf{F}'_t \boldsymbol{\theta}_t + v_t$, $v_t \sim N(0, V_t)$ where $v_{1,t}$ and $v_{2,t}$ were assumed independent.

If $\mathbf{G}_t = diag(\mathbf{G}_{1,t}, \mathbf{G}_{2,t})$ gives

$$\begin{aligned} \boldsymbol{\theta}_t = \begin{pmatrix} \boldsymbol{\theta}_{1,t} \\ \boldsymbol{\theta}_{2,t} \end{pmatrix} &= \begin{pmatrix} \mathbf{G}_{1,t} & 0 \\ 0 & \mathbf{G}_{2,t} \end{pmatrix} \begin{pmatrix} \boldsymbol{\theta}_{1,t-1} \\ \boldsymbol{\theta}_{2,t-1} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\omega}_{1,t} \\ \boldsymbol{\omega}_{2,t} \end{pmatrix} \\ &= \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t \end{aligned}$$

where $\boldsymbol{\omega}_t = \begin{pmatrix} \boldsymbol{\omega}_{1,t} \\ \boldsymbol{\omega}_{2,t} \end{pmatrix} \sim N(\mathbf{0}, \mathbf{W}_t)$ with $\mathbf{W}_t = diag(\mathbf{W}_{1,t}, \mathbf{W}_{2,t})$.

Most common example: superposition of polynomial trend + seasonal component

Assuming linear growth and trigonometric modeling with m harmonics gives

$$\mathbf{F}'_{1,t} = \left(\begin{array}{cc} 1 & 0 \end{array} \right),\ \mathbf{F}'_{2,t} = \underbrace{\left(\begin{array}{cccccccc} 1 & 0 & 1 & 0 & \cdots & 1 & 0 \end{array} \right)}_{2m},$$

$$\mathbf{F}'_t = \underbrace{\left(\begin{array}{cccccccc} 1 & 0 & 1 & 0 & \cdots & 1 & 0 \end{array} \right)}_{2(m+1)}.$$

$$\mathbf{G}_{1,t} = \left(\begin{array}{cc} 1 & 1 \\ 0 & 1 \end{array} \right),\ \mathbf{G}_{2,t} = diag(\mathbf{G}^{(1)}, \cdots, \mathbf{G}^{(m)}),$$

$$\mathbf{G}_t = diag(\mathbf{G}_{1,t}, \mathbf{G}_{2,t})$$

with $\mathbf{G}^{(j)} = \left(\begin{array}{cc} c_j & s_j \\ -s_j & c_j \end{array} \right)$,

$c_j = \cos\ 2\pi j/p$, $s_j = \text{sen}\ 2\pi j/p$, $j = 1, ..., m$.

Theorem. Consider $h \geq 2$ time series Y_{it} genearated by DLM's M_i : $\{\mathbf{F_i}, \mathbf{G_i}, \mathbf{W_i}, \mathbf{V_i}\}_t$, with state vectors $\boldsymbol{\theta}_{it}$ of dimension n_i, for $i = 1, 2, \ldots, h$. Denote observation and evolution errors in M_i by v_{it} and $\boldsymbol{\omega}_{it}$, respectively. Assume, $\forall i \neq j$, $(1 \leq i, j \leq h)$, v_{it} and $\boldsymbol{\omega}_{it}$ are independent. Then,

$$Y_t = \sum_{i=1}^{h} Y_{it}$$

follows a DLM $\{\mathbf{F}, \mathbf{G}, V, \mathbf{W}\}_t$ with state vector $\boldsymbol{\theta}'_t = (\boldsymbol{\theta}'_{1t}, \ldots, \boldsymbol{\theta}'_{ht})$, of dimension $n = n_1 + \ldots + n_h$ and quadruple

$$\begin{aligned}
\mathbf{F}'_t &= (\mathbf{F}'_{1t}, \ldots, \mathbf{F}'_{ht}) \\
\mathbf{G}_t &= diag[\mathbf{G}_{1t}, \ldots, \mathbf{G}_{ht}] \\
V_t &= \sum_{i=1}^{h} V_{it} \\
\mathbf{W}_t &= diag[\mathbf{W}_{1t}, \ldots, \mathbf{W}_{ht}]
\end{aligned}$$

Proof:

$Y_t = \mathbf{F}_t'\boldsymbol{\theta}_t + v_t$ where $v_t = \sum_{i=1}^h v_{it}$.

v_t is normal with zero mean.

Independence assumptions lead to variance V_t, as needed.

For state vector $\boldsymbol{\theta}_t = (\boldsymbol{\theta}'_{1t}, \ldots, \boldsymbol{\theta}'_{ht})$,
$\boldsymbol{\theta}_t = \mathbf{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t$ where $\boldsymbol{\omega}'_t = (\boldsymbol{\omega}'_{1,t}, \ldots, \boldsymbol{\omega}'_{h,t})$.

Again, by independence,
$\boldsymbol{\omega}_t \sim N[\mathbf{0}, \mathbf{W}_t]$ that is independent of v_t,
thus defining the DLM.

Discount Factors

Seen for 1st order model: specification of matrix $\mathbf{W}_t$ (variance of system errors $\boldsymbol{\omega}_t$) was made indirectly with use of discounts.

More specifically, if system equation is
$\boldsymbol{\theta}_t = \mathbf{G}\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t$, $\boldsymbol{\omega}_t \sim N(\mathbf{0}, \mathbf{W}_t)$ then,
for $V(\boldsymbol{\theta}_{t-1}|D_{t-1}) = \mathbf{C}_{t-1}$,
$\mathbf{R}_t = V(\boldsymbol{\theta}_t|D_{t-1}) = \mathbf{P}_t + \mathbf{W}_t$ where
$\mathbf{P}_t = V(\mathbf{G}_t\boldsymbol{\theta}_{t-1}|D_{t-1}) = \mathbf{G}_t\mathbf{C}_{t-1}\mathbf{G}_t'$.
So, $\mathbf{W}_t = \mathbf{R}_t - \mathbf{P}_t$.

Defining δ so that $\mathbf{R}_t = \mathbf{P}_t/\delta$ we can interpret δ as the percentage of information passing from $t-1$ to t. In this case,

$$\mathbf{W}_t = \mathbf{R}_t - \mathbf{P}_t = \mathbf{P}_t/\delta - \mathbf{P}_t = \mathbf{P}_t(\delta^{-1} - 1).$$

The above reasoning can be extended in the case of model superposition
(eg.: trend + seasonality)

In the general case of k models, define

$$\mathbf{P}_{i,t} = V(\mathbf{G}_{it}\boldsymbol{\theta}_{i,t-1}|D_{t-1}), i = 1, 2, \cdots, k$$

and use discounts δ_i for each block
such that $\mathbf{R}_{i,t} = \mathbf{P}_{i,t}/\delta_i$ and
$\mathbf{W}_{i,t} = \mathbf{P}_{i,t}(\delta_i^{-1} - 1)$, $i = 1, 2, \cdots, k$.

The full model obtained by the superposition of k components has

$$\mathbf{W}_t = diag(\mathbf{W}_{1,t}, \cdots, \mathbf{W}_{k,t}),$$

where $\mathbf{W}_{i,t}$ is given as above.

Practical discount strategies

Discount: suitable for 1-step evolution.

Multiplicative tool used for additive factor.

$\mathbf{C}_t \to \mathbf{R}_{t+1}$ via discount δ (leading to $\mathbf{W}_{t+1}$).

Observing y_{t+1}, $\mathbf{C}_{t+1} \to \mathbf{R}_{t+2}$ using discount δ (and leading to $\mathbf{W}_{t+2} \neq \mathbf{W}_{t+1}$).

If it is used k-steps-ahead $\to$

exponential decay of information $\to$

inconsistent with arithmetic decay

If we want to predict y_{t+2},

$V(\boldsymbol{\omega}_{t+2}|D_t)$ is needed and is set as $\mathbf{W}_{t+1}$.

Similar reasoning is valid for k-steps-ahead.

All variances $\mathbf{W}_t$ are specified 1-step-ahead.

Unknown variances

So far, all models assumed known observational variances V_t.

In practice, this is rarely true and variances must be estimated.

Following same model structure, variances are estimated sequentially and should have thir own evolution.

For now, assume $V_t = V_{t-1} = V = \phi^{-1}$, $\forall t$.

Let us start with the simplest case:
1st. order model

Model is slightly modified to

Obs. equation: $y_t = \mu_t + v_t$, $v_t \sim N(0, V)$

System equation: $\mu_t = \mu_{t-1} + \omega_t$, $\omega_t \sim N(0, VW_t^*)$

Initial info: $\mu_0 | V, D_0 \sim N(m_0, VC_0^*)$ and

$$V \sim IG(\frac{n_0}{2}, \frac{d_0}{2})$$

The inclusion of V in all variances gives analytic tractability and interpretation ease.

Independence between errors v_t and ω_t are still valid but conditional on V.

Inferential process consists again in

$$\mu_{t-1}, V \mid D_{t-1} \overset{evolution}{\rightarrow} \mu_t, V \mid D_{t-1} \overset{updating}{\rightarrow} \mu_t, V \mid D_t$$

posterior prior posterior

$\downarrow$ prediction

$$Y_t \mid D_{t-1}$$

In steps above we can work with the joint distr. of (μ, V) or with the conditional of $\mu|V$ and the marginal of V.

Note that $E[V^{-1}|D_0] = \frac{n_0/2}{d_0/2} = \frac{n_0}{d_0} = S_0^{-1}$ where S_0 is an initial estimate of V.
Value of n_0 informs the precision of this estimate since $CV[V^{-1}|D_0] = \sqrt{2}\, n_0^{-1/2}$.

Theorem. With model above, the following distributions are obtained

(a) Conditional on V:

$$\begin{aligned}
(\mu_{t-1} \mid D_{t-1}, V) &\sim N[m_{t-1}, VC^*_{t-1}], \\
(\mu_t \mid D_{t-1}, V) &\sim N[a_t, VR^*_t], \\
(Y_t \mid D_{t-1}, V) &\sim N[f_t, VQ^*_t], \\
(\mu_t \mid D_t, V) &\sim N[m_t, VC^*_t],
\end{aligned}$$

with

$$\begin{aligned}
a_t &= m_{t-1}, & R^*_t &= C^*_{t-1} + W^*_t \\
f_t &= a_t, & Q^*_t &= 1 + R^*_t \\
m_t &= a_t + A_t e_t, & C^*_t &= R^*_t - A_t^2 Q^*_t \\
A_t &= R^*_t {Q^*_t}^{-1}, & e_t &= Y_t - f_t
\end{aligned}$$

(b) For the obs. precision $\phi = V^{-1}$:

$$\begin{aligned}(\phi \mid D_{t-1}) &\sim G(n_{t-1}/2, d_{t-1}/2),\\ (\phi \mid D_t) &\sim G(n_t/2, d_t/2),\end{aligned}$$

where

$n_t = n_{t-1} + 1$ and $d_t = d_{t-1} + e_t^2 {Q_t^*}^{-1}$

(c) Uncondicional on V:

$$\begin{aligned}(\mu_{t-1} \mid D_{t-1}) &\sim t_{n_{t-1}}[m_{t-1}, C_{t-1}],\\ (\mu_t \mid D_{t-1}) &\sim t_{n_{t-1}}[a_t, R_t],\\ (Y_t \mid D_{t-1}) &\sim t_{n_{t-1}}[f_t, Q_t],\\ (\mu_t \mid D_t) &\sim t_{n_t}[m_t, C_t],\end{aligned}$$

where

$$\begin{aligned}C_{t-1} &= S_{t-1}C_{t-1}^*, \, R_t = S_{t-1}R_t^*, \, Q_t = S_{t-1}Q_t^*\\ C_t &= S_tC_t^*, \, S_{t-1} = d_{t-1}/n_{t-1} \text{ and } S_t = d_t/n_t\end{aligned}$$

Following relations are also obtained

$$
\begin{aligned}
m_t &= a_t + A_t e_t \\
C_t &= S_t / S_{t-1} [R_t - A_t^2 Q_t] = A_t S_t \\
n_t &= n_{t-1} + 1 \\
d_t &= d_{t-1} + S_{t-1} e_t^2 Q_t^{-1} \\
Q_t &= S_{t-1} + R_t \\
A_t &= R_t Q_t^{-1}
\end{aligned}
$$

Proof: Given the model definition, results in (a) follow directly from results with known variance.

The rest of the demonstration is by induction, and uses results from the NG distribution.

Assume by induction that the prior for the precision $\phi(= V^{-1})$ in (b) holds,
i.e. $(\phi \mid D_{t-1}) \sim G[n_{t-1}/2, d_{t-1}/2]$.

From (a), $(Y_t \mid D_{t-1}, \phi) \sim N[f_t, Q_t^*/\phi]$,

and

$$p(Y_t \mid D_{t-1}, \phi) \propto \phi^{1/2} \exp(-0.5\phi e_t^2/Q_t^*).$$

By Bayes theorem, the posterior for ϕ is

$$p(\phi \mid D_t) \propto p(\phi \mid D_{t-1}) p(Y_t \mid D_{t-1}, \phi).$$

Using the prior in (b) and the likelihood above gives

$$p(\phi \mid D_t) \propto \phi^{(n_{t-1}+1)/2-1} \exp[-(d_{t-1}+e_t^2/Q_t^*)\phi/2].$$

Clearly, this is the density of a $G(n_t/2, d_t/2)$ distribution, where $n_t = n_{t-1} + 1$ and $d_t = d_{t-1} + e_t^2 {Q_t^*}^{-1}$.

To prove (c), just remember that

$$\begin{aligned} \mu \mid \phi &\sim N[m, \phi^{-1}C^*] \text{ and } \phi \sim G[n/2, d/2] \Rightarrow \\ \mu &\sim t_n(m, SC^*) \end{aligned}$$

where $S = d/n$. All results in (c) follow from the marginalization of the distributions in (a) with respect to the appropriate prior/posterior Gamma distribution for ϕ.

Results for the general model are obtained in the same way.

The model is defined by

$$\begin{aligned} Y_t &= \mathbf{F}_t'\boldsymbol{\theta}_t + \nu_t \quad , \quad \nu_t \sim N(0, V) \\ \boldsymbol{\theta}_t &= \mathbf{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t \quad , \quad \boldsymbol{\omega}_t \sim N[0, V\mathbf{W}_t^*] \\ (\boldsymbol{\theta}_0 \mid D_0, V) &\sim N[\mathbf{m}_0, V\mathbf{C}_0^*] \text{ and} \\ (V \mid D_0) &\sim IG(n_0/2, d_0/2). \end{aligned}$$

(Again, we assume $V_t = V_{t-1} = V$, $\forall t$.)

Same independence assumptions remain, but now conditional on V (or $\phi = V^{-1}$).
Prior mean of ϕ is $E[\phi \mid D_t] = n_0/d_0 = S_0^{-1}$ where S_0 is a prior estimate of V.

Inferential process consists again in

$$\boldsymbol{\theta}_{t-1}, V \mid D_{t-1} \stackrel{evolution}{\rightarrow} \boldsymbol{\theta}_t, V \mid D_{t-1} \stackrel{updating}{\rightarrow} \boldsymbol{\theta}_t, V \mid D_t$$

posterior prior posterior

$\downarrow$ prediction

$$Y_t \mid D_{t-1}$$

Theorem. With model above, the following distributions are obtained

(a) Conditional on V:

$$\begin{aligned}
(\boldsymbol{\theta}_{t-1} \mid D_{t-1}, V) &\sim N[\mathbf{m}_{t-1}, V\mathbf{C}^*_{t-1}], \\
(\boldsymbol{\theta}_t \mid D_{t-1}, V) &\sim N[\mathbf{a}_t, V\mathbf{R}^*_t], \\
(Y_t \mid D_{t-1}, V) &\sim N[f_t, VQ^*_t], \\
(\boldsymbol{\theta}_t \mid D_t, V) &\sim N[\mathbf{m}_t, V\mathbf{C}^*_t],
\end{aligned}$$

with

$$\begin{aligned}
\mathbf{a}_t &= \mathbf{G}_t\mathbf{m}_{t-1}, & \mathbf{R}^*_t = \mathbf{G}_t\mathbf{C}^*_{t-1}\mathbf{G}'_t + \mathbf{W}^*_t \\
f_t &= \mathbf{F}'_t\mathbf{a}_t, & Q^*_t = 1 + \mathbf{F}'_t\mathbf{R}^*_t\mathbf{F}_t \\
\mathbf{m}_t &= \mathbf{a}_t + \mathbf{A}_t e_t, & \mathbf{C}^*_t = \mathbf{R}^*_t - \mathbf{A}_t\mathbf{A}'_t Q^*_t \\
\mathbf{A}_t &= \mathbf{R}^*_t\mathbf{F}_t {Q^*_t}^{-1}, & e_t = Y_t - f_t
\end{aligned}$$

(b) For the obs. precision $\phi = V^{-1}$:

$$
\begin{aligned}
(\phi \mid D_{t-1}) &\sim G(n_{t-1}/2, d_{t-1}/2), \\
(\phi \mid D_t) &\sim G(n_t/2, d_t/2),
\end{aligned}
$$

where

$n_t = n_{t-1} + 1$ and $d_t = d_{t-1} + e_t^2 {Q_t^*}^{-1}$

(c) Unconditional on V:

$$
\begin{aligned}
(\boldsymbol{\theta}_{t-1} \mid D_{t-1}) &\sim t_{n_{t-1}}[\mathbf{m}_{t-1}, \mathbf{C}_{t-1}], \\
(\boldsymbol{\theta}_t \mid D_{t-1}) &\sim t_{n_{t-1}}[\mathbf{a}_t, \mathbf{R}_t], \\
(Y_t \mid D_{t-1}) &\sim t_{n_{t-1}}[f_t, Q_t], \\
(\boldsymbol{\theta}_t \mid D_t) &\sim t_{n_t}[\mathbf{m}_t, \mathbf{C}_t],
\end{aligned}
$$

where

$$
\begin{aligned}
\mathbf{C}_{t-1} &= S_{t-1}\mathbf{C}_{t-1}^*, \; \mathbf{R}_t = S_{t-1}\mathbf{R}_t^*, \; Q_t = S_{t-1}Q_t^* \\
\mathbf{C}_t &= S_t\mathbf{C}_t^*, \; S_{t-1} = d_{t-1}/n_{t-1} \text{ and } S_t = d_t/n_t
\end{aligned}
$$

The following relations can be obtained

$$\mathbf{m}_t = \mathbf{a}_t + \mathbf{A}_t e_t \qquad \mathbf{C}_t = S_t/S_{t-1}[\mathbf{R}_t - \mathbf{A}_t\mathbf{A}_t' Q_t]$$

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

$$Q_t = S_{t-1} + \mathbf{F}_t'\mathbf{R}_t\mathbf{F}_t \qquad \mathbf{A}_t = \mathbf{R}_t\mathbf{F}_t Q_t^{-1}$$

Proof:

Direct generalization of 1st order model.

For (a), use results for model with known V.

For (b), use same results obtained for 1st order model.

For (c), use results for NG distribution in the multivariate case,i.e.,

$\boldsymbol{\theta} \mid \phi \sim N[\mathbf{m}, \phi^{-1}\mathbf{C}^*]$ and $\phi \sim G[n/2, d/2] \Rightarrow$

$\boldsymbol{\theta} \sim t_n[\mathbf{m}, S\mathbf{C}^*]$ where $S = d/n$.

Retrospective Analysis or Smoothing

So far we obtained the distributions of $\boldsymbol{\theta}_t | D_{t-1}$ (prior) and $\boldsymbol{\theta}_t | D_t$ (posterior or on-line).

If system is used in real time, on-line distribution is the best.

Sometimes, data up to time n (beyond t) is available.

We would like to use this information to obtain a more precise distribution for $\boldsymbol{\theta}_t$, i.e., the distribution of $\boldsymbol{\theta}_t | D_n$, $n > t$.

This distribution is useless for prediction ...

but knowing the distribution of $\boldsymbol{\theta}_t | D_n$, $n > t$ can be useful for retrospective analyses, providing a better control of the system.

Theorem. Consider the DLM $\{\mathbf{F}_t, \mathbf{G}_t, V_t, \mathbf{W}_t\}$. For all t and k ($1 \leq k < t$),

$\boldsymbol{\theta}_{t-k} | D_t \sim N(\mathbf{a}_t(-k), \mathbf{R}_t(-k))$ where

$\mathbf{a}_t(-k) = \mathbf{m}_{t-k} + \mathbf{B}_{t-k}[\mathbf{a}_t(-k+1) - \mathbf{a}_{t-k+1}]$

$\mathbf{R}_t(-k) = \mathbf{C}_{t-k} - \mathbf{B}_{t-k}[\mathbf{R}_{t-k+1} - \mathbf{R}_t(-k+1)]\mathbf{B}'_{t-k}$

with $\mathbf{B}_t = \mathbf{C}_t \mathbf{G}'_{t+1} \mathbf{R}^{-1}_{t+1}$,

and initial values

$\mathbf{a}_t(0) = \mathbf{m}_t$ e $\mathbf{R}_t(0) = \mathbf{C}_t$.

Proof: (by induction)

$p(\boldsymbol{\theta}_{t-k}, \boldsymbol{\theta}_{t-k+1}|D_t) =$

$p(\boldsymbol{\theta}_{t-k}|\boldsymbol{\theta}_{t-k+1}, D_t)\, p(\boldsymbol{\theta}_{t-k+1}|D_t)$

But $p(\boldsymbol{\theta}_{t-k}|\boldsymbol{\theta}_{t-k+1}, D_t) = p(\boldsymbol{\theta}_{t-k}|\boldsymbol{\theta}_{t-k+1}, D_{t-k})$

To obtain $p(\boldsymbol{\theta}_{t-k}|\boldsymbol{\theta}_{t-k+1}, D_{t-k})$, one needs first to note that

$$\begin{pmatrix} \boldsymbol{\theta}_{t-k} \\ \boldsymbol{\theta}_{t-k+1} \end{pmatrix} |D_{t-k} \Big) \sim N\left[\begin{pmatrix} \mathbf{m}_{t-k} \\ \mathbf{a}_{t-k+1} \end{pmatrix}, \begin{pmatrix} \mathbf{C}_{t-k} & \mathbf{S}'_{t-k+1} \\ \mathbf{S}_{t-k} & \mathbf{R}_{t-k+1} \end{pmatrix}\right].$$

So, $\boldsymbol{\theta}_{t-k}|\boldsymbol{\theta}_{t-k+1}, D_{t-k} \sim N(\mathbf{h}_t(k), \mathbf{H}_t(k))$ where

$$\begin{aligned} \mathbf{h}_t(k) &= \mathbf{m}_{t-k} + \mathbf{B}_{t-k}(\boldsymbol{\theta}_{t-k+1} - \mathbf{a}_{t-k+1}) \\ \mathbf{H}_t(k) &= \mathbf{C}_{t-k} - \mathbf{B}_{t-k}\mathbf{R}_{t-k+1}\mathbf{B}'_{t-k} \end{aligned}$$

(i) The dependence of $\boldsymbol{\theta}_{t-k}$ in $\boldsymbol{\theta}_{t-k+1}$ is linear on the mean and

(ii) $\boldsymbol{\theta}_{t-k+1}|D_t \sim N[\mathbf{a}_t(-k+1), \mathbf{R}_t(-k+1)]$, by the induction hypothesis

(i) & (ii) $\Rightarrow$ $\boldsymbol{\theta}_{t-k}|D_t \sim N(\mathbf{a}_t(-k), \mathbf{R}_t(-k))$ with

$$\mathbf{a}_t(-k) = \mathbf{m}_{t-k} + \mathbf{B}_{t-k}[\mathbf{a}_t(-k+1) - \mathbf{a}_{t-k+1}]$$
$$\mathbf{R}_t(-k) = \mathbf{C}_{t-k} - \mathbf{B}_{t-k}[\mathbf{R}_{t-k+1} - \mathbf{R}_t(-k+1)]\mathbf{B}'_{t-k}$$

Corollary. If observational variances are unknown and $V_t = V = \phi^{-1}$, $\forall t$, then

$$\boldsymbol{\theta}_{t-k} \,|\, \phi, D_t \;\sim\; N[\mathbf{a}_t(-k), \phi^{-1}\mathbf{R}_t(-k)] \text{ and}$$
$$\boldsymbol{\theta}_{t-k} \,|\, D_t \;\sim\; t_{n_t}[\mathbf{a}_t(-k), S_t\mathbf{R}_t(-k)]$$

Proof: Explores same results about the NG distribution.

Linear Bayes Optimality

Assume $\mathbf{y}$ is a vector of observations and $\boldsymbol{\theta}$ is a vector of parameters.

We want to estimate $\boldsymbol{\theta}$ by $\mathbf{d}$ is such a way that the loss $L(\boldsymbol{\theta}, \mathbf{d})$ has minimal expected value. Obviosly, if $\mathbf{y}$ is related to $\boldsymbol{\theta}$, we would like to use this information in $\mathbf{d}$.

The problem then is to choose $\mathbf{d}$ that **minimizes**

$$\mathsf{E}[L(\boldsymbol{\theta}, \mathbf{d})|\mathbf{y}].$$

Assume the only available information from $\mathbf{y}$ and $\boldsymbol{\theta}$ are the first two moments

$$\begin{bmatrix} \mathbf{y} \\ \boldsymbol{\theta} \end{bmatrix} \sim \left[\begin{pmatrix} \mathbf{f} \\ \mathbf{a} \end{pmatrix}, \begin{pmatrix} \mathbf{Q} & \mathbf{S}' \\ \mathbf{S} & \mathbf{R} \end{pmatrix} \right]$$

Without additional information about the distribution, one cannot calculate $\mathsf{E}[\boldsymbol{\theta}|\mathbf{y}]$. To bypass this problem, we work with the expected loss $\mathsf{E}[L(\boldsymbol{\theta}, \mathbf{d})]$ and restrict attention to estimators from the class of linear estimators

$$\mathbf{d} = \mathbf{d}(\mathbf{y}) = \mathbf{h} + \mathbf{H}\mathbf{y}.$$

The linear Bayes estimator (LBE) is defined as the estimator that minimizes the expected loss among linear estimators.

It can be shown that the LBE is given by

$$\mathbf{m} = \mathbf{a} + \mathbf{S}\mathbf{Q}^{-1}(\mathbf{y} - \mathbf{f})$$

and the associated loss is

$$\mathbf{C} = \mathbf{R} - \mathbf{S}\mathbf{Q}^{-1}\mathbf{S}'.$$

$\mathbf{m}$ and $\mathbf{C}$ are then used as an approximation to the moments of $\boldsymbol{\theta}|\mathbf{y}$, and called posterior linear mean and posterior linear variance of $\boldsymbol{\theta}|\mathbf{y}$. Note that they coincide with posterior mean and variance if the joint distribution of $(\mathbf{y}, \boldsymbol{\theta})$ is normal.

Importance of linear optimality to DLM

The distributions of interest in DLM's can be obtained by this method without the normality hypotheses. Consider the DLM

$$\begin{aligned} Y_t &= \mathbf{F}_t' \boldsymbol{\theta}_t + v_t, \quad v_t \sim [0, V_t] \\ \boldsymbol{\theta}_t &= \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim [0, \mathbf{W}_t] \\ \boldsymbol{\theta}_0 | D_0 &\sim [\mathbf{m}_0, \mathbf{C}_0] \end{aligned}$$

where all distributions are partially specified through means and variances only.

Then, the general theorem remains valid but without the normality assumption, i.e.

$$\begin{aligned}
\boldsymbol{\theta}_t|D_{t-1} &\sim [\mathbf{a}_t, \mathbf{R}_t] \\
Y_t|D_{t-1} &\sim [f_t, Q_t] \\
\boldsymbol{\theta}_t|D_t &\sim [\mathbf{m}_t, \mathbf{C}_t]
\end{aligned}$$

where the expressions of the moments remain the same.

The first two results follow from linearity of expectation.

Last result follows from linear optimality.

Intervention

Every time the model behaves in non-standard form, we should intervene (management by exception).

Intervention may be

(i) anticipatory (feed forward)

(ii) retrospective or corrective (feed back)

Consider known variances for the moment.

$D_t = \{y_t, I_t, D_{t-1}\}$

Anticipatory Intervention

1. Ignore observation y_t

If y_t is deemed not compatible with series

$\Rightarrow$ non-informative

$\Rightarrow$ may be ignored.

So, $I_t = \{y_t \text{ missing}\}$

$\Rightarrow D_t = D_{t-1}$

$\Rightarrow \mathbf{m}_t = \mathbf{a}_t$ and $\mathbf{C}_t = \mathbf{R}_t$.

This can be formalized in DLM's taking $V_t \to \infty$.

Anticipatory Intervention (cont.)

2. Additional evolution disturbance

If we know something happened, noise is added to the system. In general, we take a larger $\mathbf{R}_t$ for some or all components of $\boldsymbol{\theta}_t$. So, take $I_t = \{\boldsymbol{\eta}_t\}$ where the additional noise $\boldsymbol{\eta}_t \sim N[\mathbf{h}_t, \mathbf{H}_t]$ is independent of $\boldsymbol{\omega}_t$.

Formally,

$$\boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \underbrace{\boldsymbol{\omega}_t + \boldsymbol{\eta}_t}_{\boldsymbol{\omega}_t^* \sim N(\boldsymbol{h}_t, \mathbf{W}_t + \mathbf{H}_t)}$$

and $\boldsymbol{\theta}_t | D_{t-1}, I_t \sim N[\mathbf{a}_t^*, \mathbf{R}_t^*]$ with

$$\begin{aligned} \mathbf{a}_t^* &= \mathbf{a}_t + \mathbf{h}_t \\ \mathbf{R}_t^* &= \mathbf{R}_t + \mathbf{H}_t \end{aligned}$$

If some components of $\mathbf{h}_t$ and $diag\mathbf{H}_t$ are zero, we do not expect changes (from standard) in the corresponding $\theta_{t,j}$.

Just like $\mathbf{W}_t$, the role of $\mathbf{H}_t$ is to increase uncertainty.

So, it may also be specified through discounts.

In this case, use a smaller discount than usual to allow a greater increase in uncertainty.

Anticipatory Intervention (cont.)

3. Arbitrary subjective intervention

Model provides $\mathbf{a}_t$, $\mathbf{R}_t$, $\forall t$.

Assume that $I_t = \{\mathbf{a}_t^*, \mathbf{R}_t^*\}$, for some $\mathbf{a}_t^*$ and $\mathbf{R}_t^*$.

So, $\boldsymbol{\theta}_t | I_t, D_{t-1} \sim N(\mathbf{a}_t^*, \mathbf{R}_t^*)$.

Eg.: $\mathbf{R}_t^* = 0 \Rightarrow \Pr(\boldsymbol{\theta}_t = \mathbf{a}_t^* | D_{t-1}, I_t) = 1$.

This model is not consistent with MLD framework and prevents filtering.

Reconciliation with DLM's is possible with the result below.

Lemma: Let $E(\boldsymbol{\theta}_t) = \mathbf{a}_t$ and $V(\boldsymbol{\theta}_t) = \mathbf{R}_t$, $\mathbf{K}_t$ $n \times n$ non-singular, upper triangular and $\boldsymbol{\theta}_t^* = \mathbf{K}_t\boldsymbol{\theta}_t + \mathbf{h}_t$. Then, $E(\boldsymbol{\theta}_t^*) = \mathbf{a}_t^*$ and $V(\boldsymbol{\theta}_t^*) = \mathbf{R}_t^*$ if $\mathbf{K}_t = \mathbf{U}_t\mathbf{Z}_t^{-1}$ and $\mathbf{h}_t = \mathbf{a}_t^* - \mathbf{K}_t\mathbf{a}_t$, where $\mathbf{U}_t(\mathbf{Z}_t)$ is the only non-singular upper triangular square root matrix of $\mathbf{R}_t^*$ $(\mathbf{R}_t)$, i.e., $\mathbf{R}_t^* = \mathbf{U}_t\mathbf{U}_t'$ $(\mathbf{R}_t = \mathbf{Z}_t\mathbf{Z}_t')$.

Proof: From linear algebra, if $\mathbf{R}_t$ and $\mathbf{R}_t^*$ are symmetric p.d. $\Rightarrow$ $\mathbf{U}_t$ and $\mathbf{Z}_t$ exist and are unique.

From the definition of $\mathbf{h}_t$,

$$
\begin{aligned}
E(\boldsymbol{\theta}_t^*) &= \mathbf{K}_t\mathbf{a}_t + \mathbf{h}_t = \mathbf{K}_t\mathbf{a}_t + \mathbf{a}_t^* - \mathbf{K}_t\mathbf{a}_t = \mathbf{a}_t^* \\
V(\boldsymbol{\theta}_t^*) &= \mathbf{K}_t\mathbf{R}_t\mathbf{K}_t' = \mathbf{K}_t\mathbf{Z}_t\mathbf{Z}_t'\mathbf{K}_t' = (\mathbf{K}_t\mathbf{Z}_t)(\mathbf{K}_t\mathbf{Z}_t)' \\
&= \mathbf{U}_t\mathbf{U}_t'
\end{aligned}
$$

We obtain from the lemma a "second evolution" of $\boldsymbol{\theta}_t$ to $\boldsymbol{\theta}_t^*$ with the required moments. So,

$$\begin{aligned}
\boldsymbol{\theta}_t^* &= \mathbf{K}_t\boldsymbol{\theta}_t + \mathbf{h}_t \\
&= \mathbf{K}_t(\mathbf{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t) + \mathbf{h}_t \\
&= \mathbf{K}_t\mathbf{G}_t\boldsymbol{\theta}_{t-1} + \mathbf{K}_t\boldsymbol{\omega}_t + \mathbf{h}_t \\
&= \mathbf{G}_t^*\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t^*
\end{aligned}$$

where $\mathbf{G}_t^* = \mathbf{K}_t\mathbf{G}_t$, $\boldsymbol{\omega}_t^* = \mathbf{K}_t\boldsymbol{\omega}_t + \mathbf{h}_t \sim N(\mathbf{h}_t, \mathbf{W}_t^*)$ and $\mathbf{W}_t^* = \mathbf{K}_t\mathbf{W}_t\mathbf{K}_t'$.

Model is now formally in the structure of a DLM.

Anticipatory Intervention (cont.)

4. Inclusion of intervention effects

So far, interventions have maintained the same model parameters.

Sometimes, intervention requires additional parameters in order to single them out and estimate them.

Eg.: Consider the LGM with $\boldsymbol{\theta}_t = (\mu_t, \beta_t)$

Let $\boldsymbol{\theta}_t^* = (\mu_t, \beta_t, \gamma_t)$ where

$$\begin{aligned}
\boldsymbol{\theta}_t^* &= \begin{pmatrix} \mu_t^* \\ \beta_t \\ \gamma_t \end{pmatrix} = \begin{pmatrix} \mu_t + \gamma_t \\ \beta_t \\ \gamma_t \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \mu_t \\ \beta_t \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \gamma_t \\
\boldsymbol{\theta}_t^* &= \mathrm{K}_t \boldsymbol{\theta}_t + \boldsymbol{\xi}_t
\end{aligned}$$

Note that the dimension of the parameter vector changed.

A "second evolution" for model parameters is thus defined.

$$\boldsymbol{\theta}_t | D_{t-1} \sim N(\mathbf{a}_t, \mathbf{R}_t) \Rightarrow \boldsymbol{\theta}_t^* | I_t, D_{t-1} \sim N(\mathbf{a}_t^*, \mathbf{R}_t^*)$$

with

$$\mathbf{a}_t^* = \mathbf{K}_t \mathbf{a}_t + E(\boldsymbol{\xi}_t) \quad \mathbf{R}_t^* = \mathbf{K}_t \mathbf{R}_t \mathbf{K}_t' + V(\boldsymbol{\xi}_t)$$

The (total) evolution is given by

$$\boldsymbol{\theta}_t^* = \mathbf{K}_t \boldsymbol{\theta}_t + \boldsymbol{\xi}_t = \mathbf{K}_t[\mathbf{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t] + \boldsymbol{\xi}_t = \mathbf{G}_t^* \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t^*$$

where $\mathbf{G}_t^* = \mathbf{K}_t \mathbf{G}_t$, $\boldsymbol{\omega}_t^* = \boldsymbol{\xi}_t + \mathbf{K}_t \boldsymbol{\omega}_t \sim N[E(\boldsymbol{\xi}_t), \mathbf{W}_t^*]$ with $\mathbf{W}_t^* = \mathbf{K}_t \mathbf{W}_t \mathbf{K}_t' + V(\boldsymbol{\xi}_t)$.

Monitoring

Atention now is turned to monitoring model performance (poor forecasts, changes in parameters).

Techniques linked to functions of the forecast error e$_t$ $\Rightarrow$ Bayes factors (BF).

Consider models M_0 (standard), $M_1, M_2, \cdots$ with forecasts $p(y_t|D_{t-1}, M_i) = p_i(y_t|D_{t-1})$.

The BF for M_0 against M_1 based on y_t is $H_t = \frac{p_0(y_t|D_{t-1})}{p_1(y_t|D_{t-1})}$.

BF is ratio of predictive likelihoods:
The larger (smaller) H_t, the greater (smaller) the evidence for M_0 (against M_1).

The BF for M_0 versus M_1 based on the last k observations $y_t, y_{t-1}, \cdots, y_{t-k+1}$ is

$$\begin{aligned} H_t(k) &= \frac{p_0(y_t, \cdots, y_{t-k+1} | D_{t-k})}{p_1(y_t, \cdots, y_{t-k+1} | D_{t-k})} \\ &= \prod_{r=t-k+1}^{t} \frac{p_0(y_r | D_{r-1})}{p_1(y_r | D_{r-1})} \end{aligned}$$

Remarks:

(i) $H_t(1) = H_t$;

(ii) $H_t^{-1}(k)$ is the BF for M_1 versus M_0 based on $y_t, \cdots, y_{t-k+1}$;

(iii) $H_t(t)$ is BF based on data up to y_t;

(iv) $H_t(k) = H_t H_{t-1}(k-1)$ or
$\log H_t(k) = \log H_t + \log H_{t-1}(k-1)$

Due to the dynamic nature of situation, we usually work with $H_t(k)$ for small k.

$H_t = H_t(1)$ small $\Rightarrow y_t$ is a possible *outlier*.

$H_t(k)$ small $\Rightarrow$ last k points indicate structural changes not captured by M_0.

In practice, we should concentrate on the worst value of $H_t(k)$, $1 \le k \le t$. The theorem below tell how this is updated.

Define $L_t = H_t(l_t) = \min_k H_t(k)$

Idea is to use L_t and l_t for detection.

Theorem. (i) $L_t = H_t \min\{1, L_{t-1}\}$;

(ii) $L_t = H_t(l_t)$ with

$$l_t = \begin{cases} 1 + l_{t-1} & \text{if } L_{t-1} < 1 \\ 1 & \text{if } L_{t-1} \geq 1 \end{cases}$$

If L_t is too small, we intervene. In general, the threshold τ for intervention is around 0.2. Before that, nothing is done even if $L_t < 1$. (If $L_t > 1$, all is fine.)

If $\tau < L_{t-1} < 1 \Rightarrow L_t < L_{t-1}$ if $H_t < 1$.

If $L_t < \tau$, the monitor should signal.

If $L_t < \tau$ and

(i) $l_t = 1 \Rightarrow L_t = H_t \Rightarrow y_t$ is *outlier* or model starts to deteriorate at t;

(ii) $l_t > 1 \Rightarrow$ model started to deteriorate at time $t - l_t + 1$.

Choice of alternative models M_1

Assume $f_t = 0$ and $Q_t = 1$ or alternatively take

$$e_t^* = \frac{y_t - f_t}{\sqrt{Q_t}} | D_{t-1} \sim N(0, 1), \text{ under } M_0$$

If variance is unknown, change N by $t_{n_{t-1}}$.

Possible alternatives for M_1 are:

(i) $e_t \sim N(h, 1)$ - shift in level;

(ii) $e_t \sim N(0, k^2)$ - shift in scale.

Under (i), $H_t = \exp\{(h^2 - 2he_t)/2\}$ and choice of h in based on BF.

$H_t = 1$ (indiference between M_0 and M_1) $\Leftrightarrow h = 2e_t$.

This is reasonable for $e_t = 1.5 \Rightarrow h \approx 3$.

Assuming threshold $\tau = e^{-2} \Leftrightarrow h^2 - 2he_t - 2\log\tau = 0$.

Threshold should be reached when $e_t \approx 2,5 \Rightarrow h = 1$ or 4.

h between 3 and 4 seem ok: BF indiferent when $e_t = 1.5$ and rejection of M_0 when $e_t = 2.5$.

Under (ii), $H_t = k\exp\{-e_t^2(1 - k^{-2})/2\}$.

Note: value of k not relevant for large $|e_t|$.

$k = 3$ or 4 and $\tau = 0,15$ seem ok: lead to rejection of M_0 when $|e_t| = 2.5$.

Retrospective Intervention

Automated detection and diagnostic scheme:

(A) Calculation of H_t.

If $H_t > \tau$, go to (B).

If $H_t < \tau$, y_t is *outlier* and omitted, or y_t is start of change. Go to (C).

(B) Calculation of L_t and l_t.

If $L_t > \tau$ (or $l_t > 4$), go to (D).

If $L_t < \tau$, go to (C).

If $L_t < 1$ and $l_t \leq 4$, go to (C).

(C) Monitor signals. Intervene and reset monitor to $l_t = 0$ and $L_t = 1$.

(D) Keep standard analysis.

Simplest form of intervention is through increase in prior uncertainty.

Dynamic regression models

Regression model in time series has

y_t - response time series

x_t - time series of regressors that influence y_t.

It is also possible that x_s, $s < t$ or even y_s, $s < t$ influence y_t.

Simple model: $\mu_t = \alpha + \beta x_t$

Assume that exists f such that $\mu_t = f(x_t, t)$. In general, if it exists it is unknown.

One approximation is to take $f(x_t, t) = \alpha_t + \beta_t x_t$, i.e., linear function of x_t but co-efficients change.

Good choices of α_t and β_t lead to good approximation.

Model has local form to represent changes in α and β. Simple form to relate α's and β's is random walk

$$\begin{pmatrix} \alpha_t \\ \beta_t \end{pmatrix} = \begin{pmatrix} \alpha_{t-1} \\ \beta_{t-1} \end{pmatrix} + \boldsymbol{\omega}_t$$

The multiple regression DLM

Ideas can be extended to the case of **n regressors** x_1, x_2, $\cdots$, x_n where $x_1 = 1$.

$$\text{Obs. eq.: } y_t = \alpha_t + \beta_{t,1} x_{t,2} + \cdots + \beta_{t,n-1} x_{tn} + v_t$$

$$\begin{aligned}
\text{Syst. eq.: } \alpha_t &= \alpha_{t-1} + \omega_{t,1} \\
\beta_{t,1} &= \beta_{t-1,1} + \omega_{t,2} \quad \omega_{t,i} \sim N(0, W_{t,i}) \\
&\vdots \\
\beta_{t,n-1} &= \beta_{t-1,n-1} + \omega_{t,n}
\end{aligned}$$

or placing in matrix form

$$\begin{aligned}
\text{Obs. equation: } y_t &= \mathbf{F}_t' \boldsymbol{\theta}_t + v_t \quad v_t \sim N(0, V_t) \\
\text{System equation: } \boldsymbol{\theta}_t &= \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t \quad \boldsymbol{\omega}_t \sim N(0, \mathbf{W}_t)
\end{aligned}$$

where $\mathbf{F}_t' = (1, x_{t,2}, \cdots, x_{tn})$,

$\boldsymbol{\theta}_t' = (\alpha_t, \beta_{t,1}, \cdots, \beta_{t,n-1})'$, $\boldsymbol{\omega}_t' = (\omega_{t,1}, \cdots, \omega_{t,n})'$

If $\mathbf{W}_t = 0 \Rightarrow \boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1} = \boldsymbol{\theta}$, $\forall t$

$\Rightarrow$ Static (usual) regression.

If $\mathbf{W}_t \neq 0$, $\boldsymbol{\theta}_t$ changes with time.

$\mathbf{W}_t$ large (small), lots of (little) change in $\boldsymbol{\theta}_t$ with time.

only 1 regressor: $\mathbf{F}'_t = (1, x_t)$, $\boldsymbol{\theta}'_t = (\alpha_t, \beta_t)'$

For forecast, future values of regressors are needed.

One possibility: joint modeling of y_{t+k} and $\mathbf{x}_{t+k} \Rightarrow$ multivariate model.

Another form is to obtain $p(\mathbf{x}_{t+k}|D_t)$ and

$$\begin{aligned} p(y_{t+k}|D_t) &= \int p(y_{t+k}, \mathbf{x}_{t+k}|D_t) d\mathbf{x}_{t+k} \\ &= \int p(y_{t+k}|\mathbf{x}_{t+k}, D_t) p(\mathbf{x}_{t+k}|D_t) d\mathbf{x}_{t+k} \end{aligned}$$

If we only have $E[\mathbf{x}_{t+k}|D_t] = \mathbf{h}_t(k)$ and $V[\mathbf{x}_{t+k}|D_t] = \mathbf{H}_t(k)$ then using that $y_{t+k}|\mathbf{x}_{t+k}, \boldsymbol{\beta}_{t+k}, D_t \sim N(\mathbf{x}'_{t+k}\boldsymbol{\beta}_{t+k}, V_t)$ gives

$$
\begin{aligned}
E[y_{t+k}|D_t] &= E[E[y_{t+k}|\mathbf{x}_{t+k}, \boldsymbol{\beta}_{t+k}, D_t]] \\
&= E[\mathbf{x}'_{t+k}\boldsymbol{\beta}_{t+k}|D_t] \\
&= E[E[\mathbf{x}'_{t+k}\boldsymbol{\beta}_{t+k}|\mathbf{x}_{t+k}, D_t]] \\
&= E[\mathbf{x}'_{t+k}E[\boldsymbol{\beta}_{t+k}|\mathbf{x}_{t+k}, D_t]] \\
&= \mathbf{h}_t(k)\mathbf{a}_t(k) \quad \text{and} \\
V[y_{t+k}|D_t] &= E[V[y_{t+k}|\mathbf{x}_{t+k}, \boldsymbol{\beta}_{t+k}, D_t]] \\
&+ V[E[y_{t+k}|\mathbf{x}_{t+k}, \boldsymbol{\beta}_{t+k}, D_t]] \\
&= \frac{n_t}{n_t - 2}[S_t + \mathbf{h}_t(k)'\mathbf{R}_t(k)\mathbf{h}_t(k) \\
&+ tr\{\mathbf{R}_t(k)\mathbf{H}_t(k)'\}] + \mathbf{m}'_t\mathbf{H}_t(k)\mathbf{m}_t
\end{aligned}
$$

In dynamic regression models, $\mathbf{a}_t(k) = \mathbf{m}_t$ and $\mathbf{R}_t(k) = \mathbf{C}_t + \sum_{i=1}^k \mathbf{W}_{t+i}$.

Transfer response functions

Mean response function may depend on current and lagged values of a regressor
$\mu_t = \theta_0 + \theta_1 x_t + \theta_2 x_{t-1} + \cdots + \theta_{p+1} x_{t-p}$

Model can be placed in a dynamic structure making θ's time dependent. System equation would typically be a random walk
$\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t$.

In the above, y_t depend on x_t only up to p time periods. One can also model an effect that smoothly decays to 0 through transformations of x.

Example: $y_t = \theta_{0,t} + \xi_t + v_t$ where

$\xi_t = \lambda \xi_{t-1} + \psi x_t$.

λ represents memory of effect

ψ represents the strength of x.

It can be shown that $\xi_{t+k} = \lambda^k \xi_t + \psi \sum_{i=1}^{k} \lambda^{k-i} x_{t+i}$

If $x_{t+r} = 0$, $r \neq 1$, $\xi_{t+k} = \lambda^k \xi_t + \psi \lambda^{k-i} x_{t+1}$.

In the usual regression model $y_t = \alpha + \theta x_t + e_t$, the transfer function of x_t in y_{t+r} is

$$\begin{cases} \theta x, & r = 0 \\ 0, & r \neq 0 \end{cases}$$

In autoregressive models of order p, we have

$$\begin{cases} \theta_{r+1} x, & r = 0, 1, \cdots, p \\ 0, & \text{c.c.} \end{cases}$$

Definition: A general transfer function model is

$$\begin{aligned} y_t &= \mathbf{F}_t'\boldsymbol{\theta}_t + v_t \\ \boldsymbol{\theta}_t &= \mathbf{G}\boldsymbol{\theta}_{t-1} + x_t\boldsymbol{\psi}_t + \mathbf{w}_{t,1} \\ \boldsymbol{\psi}_t &= \boldsymbol{\psi}_{t-1} + \mathbf{w}_{t,2} \end{aligned}$$

This model can be placed in DLM form. Define a state vector $\tilde{\boldsymbol{\theta}}_t' = (\boldsymbol{\theta}_t', \boldsymbol{\psi}_t')$ and the model quadruple as

$$\tilde{\mathbf{F}}_t' = (\mathbf{F}_t', \mathbf{0}') \ , \ \tilde{\mathbf{G}}_t = \begin{pmatrix} \mathbf{G} & x_t\mathbf{I}_n \\ \mathbf{0} & \mathbf{I}_n \end{pmatrix}, V \text{ and}$$

$$\tilde{\mathbf{W}}_t = \begin{pmatrix} \mathbf{W}_{t,1} + x_t^2\mathbf{W}_{t,2} & x_t\mathbf{W}_{t,2} \\ x_t\mathbf{W}_{t,2} & \mathbf{W}_{t,2} \end{pmatrix}$$

The forecast function is

$$f_t(k) = E[y_{t+k}|D_t] = \mathbf{F}_{t+k}' E[\boldsymbol{\theta}_{t+k}|D_t]$$

$$\begin{aligned}\text{But } \boldsymbol{\theta}_{t+k} &= \mathbf{G}\boldsymbol{\theta}_{t+k-1} + \boldsymbol{\psi}_{t+k} x_{t+k} + \mathbf{w}_{t,1} \\ &= \mathbf{G}^k \boldsymbol{\theta}_t + \sum_{r=1}^{k} \mathbf{G}^{k-r} \boldsymbol{\psi}_{t+k-r} x_{t+r} + \text{ errors} \\ &= \mathbf{G}^k \boldsymbol{\theta}_t + \sum_{r=1}^{k} \mathbf{G}^{k-r} \boldsymbol{\psi}_t x_{t+r} + \text{ errors}\end{aligned}$$

So, $f_t(k) = \mathbf{F}'_{t+k}[\mathbf{G}^k \mathbf{m}_t + \sum_{r=1}^{k} \mathbf{G}^{k-r} \mathbf{h}_t x_{t+r}]$, where $\mathbf{h}_t = E[\boldsymbol{\psi}_t | D_t]$.

If $x_{t+r} = 0$, $r \neq 1 \Rightarrow$
$f_t(k) = \mathbf{F}'_{t+k} \mathbf{G}^k \mathbf{m}_t + \mathbf{F}'_{t+k} \mathbf{G}^{k-1} \mathbf{h}_t x_{t+1}$.

In the example, $\mathbf{G} = \lambda$, $\boldsymbol{\psi}_t = \psi$, $\mathbf{F} = 1$ and the transfer function is $\lambda^k \psi x$.

Significance tests for model parameters

For some models it is important to assess the significance of a subset $\boldsymbol{\theta}_{t,1}$ of parameters in explaining the response series y_t.

Example: $\boldsymbol{\theta}_{t,1} = (\beta_{t,j}, \gamma_{t,j})$, the parameters of the j-th harmonic of a seasonal component

Consider the region $R = \{\boldsymbol{\theta} | p(\boldsymbol{\theta}|D_t) > p(\mathbf{0}|D_t)\}$, where $\boldsymbol{\theta}$ can be

(i) $\boldsymbol{\theta}_t$,

(ii) $\theta_{t,j}$ or

(iii) $\boldsymbol{\theta}_{t,1}$ a subvector of components of $\boldsymbol{\theta}_t$.

Example: If V is known, $\boldsymbol{\theta}_t|D_t \sim N(\mathbf{m}_t, \mathbf{C}_t)$, $\boldsymbol{\theta}|D_t \sim N(\mathbf{m}, \mathbf{C})$ where if

(i) $\boldsymbol{\theta} = \theta_{t,j}$ then $\mathbf{m} = m_{t,j}$ and $\mathbf{C} = C_{t,jj}$;

(ii) $\boldsymbol{\theta} = \boldsymbol{\theta}_{t,1}$ then $\mathbf{m} = \mathbf{m}_{t,1}$ and $\mathbf{C} = \mathbf{C}_{t,1}$, where $\mathbf{m}_{t,1}$ and $\mathbf{C}_{t,1}$ are the respective components of vector $\mathbf{m}_t$ and matrix $\mathbf{C}_t$.

If $P(R|D_t)$ is

(i) high $\Rightarrow$ the point $\mathbf{0}$ is in the tail of the distribution $\Leftrightarrow$ $\mathbf{0}$ is very unilkely $\Leftrightarrow$ $\boldsymbol{\theta}$ is significant;

(ii) low $\Rightarrow$ $\mathbf{0}$ is in the tail of the distribution $\Leftrightarrow$ $\mathbf{0}$ is very likely $\Leftrightarrow$ $\boldsymbol{\theta}$ is not significant;

Calculation of $P(R|D_t)$

(i) V known

$\boldsymbol{\theta}|D_t \sim N(\mathbf{m}, \mathbf{C})$ and

$$p(\boldsymbol{\theta}|D_t) \;=\; k \exp(-Q(\boldsymbol{\theta})/2) \text{ where}$$
$$Q(\boldsymbol{\theta}) \;=\; (\boldsymbol{\theta}-\mathbf{m})'\mathbf{C}^{-1}(\boldsymbol{\theta}-\mathbf{m})$$

But $p(\boldsymbol{\theta}|D_t) \geq p(\mathbf{0}|D_t) \Leftrightarrow$

$Q(\boldsymbol{\theta}) \leq Q(\mathbf{0}) = \mathbf{m}'\mathbf{C}^{-1}\mathbf{m}$.

From page 20 of notes, $Q(\boldsymbol{\theta})|D_t \sim \chi^2_q$.

So, $P(R|D_t) = P(Q(\boldsymbol{\theta}) \leq Q(\mathbf{0})|D_t) =$

$P(\chi^2_q < \mathbf{m}'\mathbf{C}^{-1}\mathbf{m})$,

that can be easily calculated.

(ii) V unknown

$\boldsymbol{\theta}|D_t \sim t_{n_t}(\mathbf{m}, \mathbf{C})$ and

$$p(\boldsymbol{\theta}|D_t) = k[n + Q(\boldsymbol{\theta})]^{-(n+q)/2}$$

So, $p(\boldsymbol{\theta}|D_t) \geq p(\mathbf{0}|D_t) \Leftrightarrow$

$Q(\boldsymbol{\theta}) \leq Q(\mathbf{0}) = \mathbf{m}'\mathbf{C}^{-1}\mathbf{m}$.

But $Q(\boldsymbol{\theta})/q|D_t \sim F_{q,n_t}$

So $P(R|D_t) = P(Q(\boldsymbol{\theta}) \leq Q(\mathbf{0})|D_t) =$

$P(F_{q,n_t} < \mathbf{m}'\mathbf{C}^{-1}\mathbf{m}/q)$,

that can also be easily calculated.

Outliers and missing data

An outlier can be identified with an uexpected value for an observation

$\Rightarrow$ large forecast error.

We know that $e_t | D_{t-1} \sim t_{n_{t-1}}(0, Q_t)$.

Look and the value of $|e_t|$ or $|e_t|/\sqrt{Q_t}$

e_t influences estimates of $\boldsymbol{\theta}_t$ and V since

$\mathbf{m}_t = \mathbf{a}_t + \mathbf{A}_t e_t$ and $S_t = \dfrac{S_{t-1}}{(n_{t-1}+1)} \left[n_{t-1} + \dfrac{e_t^2}{Q_t} \right]$

e_t large $\Rightarrow$ y_t not described by the same model $\Rightarrow$ $(\boldsymbol{\theta}_t, V | D_t) \sim (\boldsymbol{\theta}_t, V | D_{t-1})$.

Obviously same result if y_t is missing.

Irregular time intervals

So far, time intervals equal to 1.

Not always true

Example:

only 1 value available at the end of month

$\Rightarrow$ observe y_{31}, y_{60}, y_{91}, $\cdots$

All that need to be done is to repeat missing data procedure.

Use of discounts should follow previous recommendation of 1-step-ahead rule.

Data Transformations

Very common in Statistics but ...

be careful with interpretation.

One idea is to transform y_t non-normal into $z_t = g(y_t)$ approximately normal.

If $E(y_t) = \mu_t$ and $V(y_t) = V_t$ then Taylor series expansions give

$$E(z_t) \approx g(\mu_t) + g''(\mu_t)V_t/2,$$

Analogously, $V(z_t) \approx \{g'(\mu_t)\}^2 V_t$.

Commonly used family of transformations,

$$g(x) = \begin{cases} (x^\lambda - 1)/\lambda, & \lambda \neq 0 \\ \log x, & \lambda = 0 \end{cases}$$

So, $V(z_t) = \{E(z_t)\}^{2(1-\lambda)} V$.

Important special cases:

(i) $\lambda = 0$ (log transformation) $\Rightarrow V(z_t) \propto E(z_t)^2$;

(ii) $\lambda = 0,5$ ($\sqrt{\quad}$ transformation) $\Rightarrow V(z_t) \propto E(z_t)$ (Poisson model).

Obviously, since $z_t \sim t_{n_t}(f_t, Q_t)$, forecast for y_t must be based on its distribution

Note that $E(z_t) \neq g[E(y_t)]$ but median $z_t = g$(median y_t).

If $y_t = \gamma_t x_t^{\beta_t} v_t$ then $z_t = \log y_t = \log \gamma_t + \beta_t \log x_t + \log v_t$ and additive linear model is OK. Very used in Econometrics where one measures the effect on y_t of the rate of growth (rather than the growth) of x_t.

Non-normal data

Generalized DLM $\rightarrow$

Extension of DLM to exponential family.

Observation and system equations now defined by

$$\begin{aligned} f(y_t|\theta_t) &= a(y_t)exp\{y_t\theta_t + b(\theta_t)\} \\ g(\mu_t) &= \mathbf{F}_t'\boldsymbol{\beta}_t \text{ where } E(y_t|\theta_t) = \mu_t \\ \boldsymbol{\beta}_t &= \mathbf{G}\boldsymbol{\beta}_{t-1} + \boldsymbol{\omega}_t \quad \boldsymbol{\omega}_t \sim N(\mathbf{0}, \mathbf{W}_t) \\ \boldsymbol{\beta}_0|D_0 &\sim N(\mathbf{m}_0, \mathbf{C}_0) \end{aligned}$$

$g(\mu_t)$ is the link function.

No conjugacy $\rightarrow$ no analytically exact inference

Generalizes the GLM (on page 37) by allowing a dynamic structure for $\boldsymbol{\beta}$.

Example: Dynamic logistic regression

Logistic regression can be extended as

$$y_t|\pi_t \sim bin(n_t, \pi_t),\ t = 1, ..., n$$

probabilities π_t determined by the values of variable x

$$\begin{aligned} \pi_t &= F(\alpha_t + \beta_t x_t) \quad , \quad t = 1, ..., n \\ \alpha_t &= \alpha_{t-1} + w_{1,t} \\ \beta_t &= \beta_{t-1} + w_{2,t} \end{aligned}$$

F - any distribution function

Estimation of cycles

Cycles: processes that repeat themselves according to a regular pattern. The simplest example already seen is the harmonic

$$y_t = acos(\theta t) + bsin(\theta t) = rcos(\theta t + \phi)$$

where $r^2 = a^2 + b^2$ and $cos(\phi) = a/r$.

The constant r is the series amplitude.

- θ is the frequency, since full wave makes $\theta/2\pi$ cycles in a time unit
- y_t completes cycle after λ time units
- $\lambda = 2\pi/\theta$ - length of harmonic wave
- ϕ is the phase angle and indicates how distant we moved at the origin.

Consider the AR(2) model

$y_t - a_1 y_{t-1} - a_2 y_{t-2} = 0.$

If $a_1^2 + 4a_2 < 0$, the roots of the polynomial equation $1 - a_1 B - a_2 B^2 = 0$ are imaginary.

The homogeneous solution has form $y_t = \beta_1 r^t cos(\theta t + \beta_2)$, where β_1 and β_2 are arbitrary constants, $r = (-a_2)^{1/2}$, and θ satisfies $cos(\theta) = a_1/[2(-a_2)^{1/2}]$.

We clearly have trigonometric functions as solutions, and they impose a cyclic pattern.

Stability: determined by $r = (-a_2)^{1/2}$.

If $\mid a_2 \mid = 1$, oscillations do not alter amplitude; homogeneous solution is periodic.

Oscillations decay if $\mid a_2 \mid < 1$ and explode if $\mid a_2 \mid > 1$.

Cycles can be represented by AR(2) forms.

When $\beta_t = a_1\beta_{t-1} + a_2\beta_{t-2}$, model can be described in system equations by taking $\boldsymbol{\theta}_t = (\beta_t, \beta_{t-1})'$ and so

$$\beta_t = (1,0)\boldsymbol{\theta}_t \quad \text{and} \quad \boldsymbol{\theta}_t = \mathbf{G}\boldsymbol{\theta}_{t-1},$$

$$\text{with } \mathbf{G} = \begin{pmatrix} a_1 & a_2 \\ 1 & 0 \end{pmatrix} \quad .$$

Given initial conditions $\boldsymbol{\theta}_1 = (\beta_1, \beta_0)'$, the solution $\beta_t = (1, 0)\mathbf{G}^{t-1}\boldsymbol{\theta}_1$ has form determined by eigen structure of matrix $\mathbf{G}$.

In the case of a conjugate pair of complex eigenvalues we have the solution of an AR(2) model.

Placing the AR(2) in the stochastic form
$\beta_t = a_1\beta_{t-1} + a_2\beta_{t-2} + \omega_t$,
where $\omega_t \sim N(0, W)$ gives

$$\beta_t = (1, 0)\boldsymbol{\theta}_t \quad \text{and} \quad \boldsymbol{\theta}_t = \mathbf{G}\boldsymbol{\theta}_{t-1} + (\omega_t, 0)'.$$

A series y_t exhibiting a cyclic behaviour of length λ and decay r may be modelled by

$$
\begin{aligned}
y_t &= \beta_t + v_t \quad v_t \sim N(0, V) \\
\beta_t &= a_1\beta_{t-1} + a_2\beta_{t-2} + w_t \quad w_t \sim N(0, W).
\end{aligned}
$$

$\omega_t = 0, \forall t \Rightarrow y_t$ are observations from a dampened cosine wave

The model in matrix form is

$$
\begin{aligned}
y_t &= \mathbf{F}'\boldsymbol{\theta}_t + v_t \\
\boldsymbol{\theta}_t &= \mathbf{G}\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t
\end{aligned}
$$

where $\boldsymbol{\theta}_t = (\beta_t, \beta_{t-1})$ and

$\mathbf{F} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $\mathbf{G} = \begin{pmatrix} a_1 & a_2 \\ 1 & 0 \end{pmatrix}$ and $\mathbf{W} = \begin{pmatrix} W & 0 \\ 0 & 0 \end{pmatrix}$,

where $W = V(\omega_t)$, $r = (-a_2)^{1/2}$ and

$\lambda = 2\pi/cos^{-1}(a_1/2r)$.

Variance laws

So far, normal model assumed with unknown mean and variance.

Basic property of normal: variance does not depend on mean was maintained.

Data may not behave like normal.

One solution is to normalize through transformation g.

Problem: when forecasting we must come back to original scale which may be complicated depending on g. Also, after transformation we lose interpretation on data.

Alternative: model the dependence of variance on mean.

Deterministic approach

Assume $V(y_t) = k_t V$ where k_t is known.

The only change in the inference cycle is that $Q_t = k_t S_{t-1} + \mathbf{F}_t' \mathbf{R}_t \mathbf{F}_t$.

(When $V(y_t) = V$, k_t was replaced by 1.)

Previous discussion may lead to $V(y_t) = k(\mu_t)V$.

Most common examples are:
i) $k(\mu_t) = \mu_t^p$, $p > 0$ ($p = 1 \Leftrightarrow$ Poisson)
ii) $1 + b\mu_t^p$ or $\mu_t(1 - \mu_t)$ (binomial data).

Exact treatament is difficult and destroys normal theory.

Simple alternative: replace $k(\mu_t)$ by $k(f_t)$ in the expression of Q_t since f_t is the best estimate of $\mu_t|D_{t-1}$.

Other examples of variance modeling are

1. ARCH (p) models given by

   $$V_t = \alpha + \gamma_1 v_{t-1}^2 + \gamma_2 v_{t-2}^2 + \cdots + \gamma_p v_{t-p}^2,$$

   V_t described by lagged obs. errors;

2. GARCH (p,q) models given by

   $$V_t = \alpha + \gamma_1 v_{t-1}^2 + \cdots + \gamma_p v_{t-p}^2 + \delta_1 V_{t-1} + \cdots + \delta_q V_{t-q},$$

   V_t also described by $AR(q)$ from

Stochastic modeling

Assume that $\phi = V^{-1}$ is now indexed by t and $\phi_t | D_{t-1} \sim G(n_{t-1}/2, d_{t-1}/2)$.

Taking $\phi_t = \phi_{t-1} + \psi_t$ where $E(\psi_t) = 0$ and $V(\psi_t) = U_t$

$\Rightarrow E(\phi_t) = E(\phi_{t-1}) = S_{t-1}^{-1}$ and

$V(\phi_t) = V(\phi_{t-1}) + V(\psi_t) = \frac{2n_{t-1}}{d_{t-1}^2} + U_t.$

Thinking multiplicatively (with discounts), $V(\phi_t | D_{t-1}) = V(\phi_{t-1} | D_{t-1})/\delta_V = 2n_{t-1}/d_{t-1}^2 \delta_V.$

$$\text{This defines } U_t = \frac{2n_{t-1}}{d_{t-1}^2 \delta_V} - \frac{2n_{t-1}}{d_{t-1}^2} = V(\phi_{t-1} | D_{t-1})(\delta_V^{-1} - 1).$$

So, U_t can be set with help of variance discount δ_V.

For an analyticaly tractable analysis, it is required that $\phi_t|D_{t-1} \sim G(c_1, c_2)$. Solving for c_1 and c_2,

$\frac{c_1}{c_2} = E(\phi_{t-1}|D_{t-1})$ and $\frac{c_1}{c_2^2} = V(\phi_{t-1}|D_{t-1})$
$\Rightarrow c_1 = \delta_V n_{t-1}/2$ and $c_2 = \delta_V d_{t-1}/2$.

This result is also obtained if
$\delta_V \frac{\phi_t}{\phi_{t-1}}|\phi_{t-1} \sim \text{Beta}\left(\delta_V \frac{n_{t-1}}{2}, (1-\delta_V)n_{t-1}/2\right)$

Analysis follows as before, after multiplying prior parameters of $\phi_t|D_{t-1}$ by δ_V. Typically, δ_V will be close to 1.

Remarks:

1. Filtering is approximately avaialable (based on moments). Result is $S_t^{-1}(k) = E(\phi_{t-k}|D_t) = S_{t-k}^{-1}+$ $\delta_{V,t-k+1}(S_t^{-1}(-k+1) - S_{t-k}^{-1}$ and $V(\phi_{t-k}|D_t) = \frac{2}{n_{t-k}S_{t-k}^2}(1 - \delta_{V,t-k+1}) +$ $\frac{2}{n_t(-k+1)S_t^2(-k+1)}\delta_{V,t-k+1}$ initialized with $S_t(0) = S_t$ and $n_t(0) = n_t$.

2. Another model used in known as stochatic volatility and is described by an AR(1) process in the log volatilities

$$\log V_t = \alpha + \gamma \log V_{t-1} + \psi_t\,,\ \psi_t \sim N(0, \sigma_V^2)$$

Hyperparameter Estimation

Introduction

DLM: defined by quadruple $M = \{\mathbf{F}_t, \mathbf{G}_t, V_t, \mathbf{W}_t\}$.

If all elements of the quadruple are known, inference is known.

If V is unknown, inference is also known.

However, if there are unknowns in vector $\mathbf{F}_t$ (eg.: forecasts in AR models), matrix $\mathbf{G}$ (eg.: transfer response), or even, if values of $\mathbf{W}$ or δ are unknown (common in practice),

$\Rightarrow$ Model is non-linear and previous theorems do not apply

Let $\boldsymbol{\lambda}$ be the vector containing all unknowns

These unknown quantities are called hyperparameters

Model is now defined in terms of $\boldsymbol{\lambda}$, i.e., $M(\boldsymbol{\lambda}) = \{\mathbf{F}_t(\boldsymbol{\lambda}), \mathbf{G}_t(\boldsymbol{\lambda}), V_t, \mathbf{W}_t(\boldsymbol{\lambda})\}$.

Note that conditional on $\boldsymbol{\lambda}$, model is linear and inference is done as before.

When there is presence of hyperparameters in the model, they must be integrated out for performing inference about quantities of interest: $\boldsymbol{\theta}_t$, y_{t+k}, $\cdots$.

$$
\begin{aligned}
p(\boldsymbol{\theta}_t|D_t) &= \int p(\boldsymbol{\theta}_t, \boldsymbol{\lambda}|D_t)d\boldsymbol{\lambda} \\
&= \int p(\boldsymbol{\theta}_t|\boldsymbol{\lambda}, D_t)p(\boldsymbol{\lambda}|D_t)d\boldsymbol{\lambda} \text{ where} \\
p(\boldsymbol{\lambda}|D_t) &\propto p(\boldsymbol{\lambda}|D_{t-1})p(y_t|\boldsymbol{\lambda}, D_{t-1})
\end{aligned}
$$

Even tough $p(\boldsymbol{\theta}_t|\boldsymbol{\lambda}, D_t)$ and $p(y_t|\boldsymbol{\lambda}, D_{t-1})$ are analytically available, the above integrations are generally not solved analytically.

Some form of approximation is needed.

Possible solutions:

(1) discretize $\boldsymbol{\lambda}$:

all calculations are available since integrals become sums. Every value of $\boldsymbol{\lambda}$ defines a model $\Rightarrow$ multiprocess models (ch. 12 of textbook);

(2) apply linearization;

(3) perform numerical integration;

(4) solve by simulation.

These techniques are outlined below.

Linearization

Solves non-linearities in $\mathbf{F}$ and/or $\mathbf{G}$.

Consider the model with $\boldsymbol{\theta}_t = \lambda\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t$, non-linear.

Parameters can be redefined as $\boldsymbol{\theta}_t^* = (\boldsymbol{\theta}_t, \lambda)$ evolving via

$$\begin{aligned} \boldsymbol{\theta}_t^* &= g(\boldsymbol{\theta}_{t-1}) + \boldsymbol{\omega}_t^* \text{ where} \\ g(\boldsymbol{\theta}_{t-1}^*) &= \lambda \begin{pmatrix} \boldsymbol{\theta}_{t-1} \\ 1 \end{pmatrix} \text{ and } \boldsymbol{\omega}_t^* = \begin{pmatrix} \boldsymbol{\omega}_t \\ 0 \end{pmatrix} \end{aligned}$$

Idea is to approximate g by a linear function using Taylor series approximation around $\mathbf{m}_{t-1}$, the best available estimate of $\boldsymbol{\theta}_{t-1}$ at time $t-1$.

But $g(\boldsymbol{\theta}_{t-1}) = g(\boldsymbol{m}_{t-1}) + \mathbf{G}_t(\boldsymbol{\theta}_{t-1} - \mathbf{m}_{t-1}) +$ higher order terms

where $\mathbf{G}_t = \frac{\partial g(\boldsymbol{\theta}_{t-1})}{\partial \boldsymbol{\theta}_{t-1}}\Big|_{\boldsymbol{\theta}_{t-1}=\mathbf{m}_{t-1}}$.

So,

$$\begin{aligned} \boldsymbol{\theta}_t &\doteq g(\boldsymbol{m}_{t-1}) - \mathbf{G}_t\mathbf{m}_{t-1} + \mathbf{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t \\ &= \boldsymbol{h}_t + \mathbf{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t \end{aligned}$$

In the example, $\mathbf{G}_t = \begin{pmatrix} \lambda\mathbf{1} & \boldsymbol{\theta}_{t-1} \\ 0 & 1 \end{pmatrix}\Big|_{\boldsymbol{\theta}_{t-1}=\mathbf{m}_{t-1}}$.

Therefore $\boldsymbol{\theta}_t | D_{t-1} \sim N(\mathbf{a}_t, \mathbf{R}_t)$ where

$$\begin{aligned} \mathbf{a}_t &= \mathbf{h}_t + \mathbf{G}_t\mathbf{m}_{t-1} \text{ and} \\ \mathbf{R}_t &= \mathbf{G}_t\mathbf{C}_{t-1}\mathbf{G}_t' + \mathbf{W}_t \end{aligned}$$

Numerical integration

Based on deterministic rules.

Difficult to use in large parametric spaces ($p > 10$).

1) Normal approximation of posterior π
Obtained after a Taylor expansion of $\log \pi$ around the posterior mode.
GLM: mode is obtained after an IRLS (iterative reweighted least squares) algorithm
At each iteration: normal regression model is built.

2) Laplace approximation

$$\begin{aligned} E[t(\boldsymbol{\theta})] &= \int t(\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta} \\ &= \frac{\int t(\boldsymbol{\theta})l(\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}}{\int l(\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}} \end{aligned}$$

Approx integrands by quadratic forms on log

Errors of order $O(n^{-1})$ cancel out

$\rightarrow$ error becomes $O(n^{-2})$.

3) Approximation by Gaussian quadrature

Approximate integrands in the form $\exp(-\frac{x^2}{2})h(x)$

If h is a polynomial function

$\rightarrow$ approximation is exact.

All approximations based on the normal distribution: reparametrizations are useful.

Methods based on stochastic simulation

Inference based on samples of the posterior $\pi(\boldsymbol{\theta})$ where $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_p)$.

Sample is always a partial substitute of the information in the density.

Deterministic methods: errors become smaller with increase of number of observations, that does not depend on the analyst.

Simulation methods: errors depend on the number of generated values, controlled by the analyst.

Consider sample $\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_n$ from π.

Sample from marginal of θ_i : $\theta_{i1}, \cdots, \theta_{in}$

A lot simpler than analytical integration

Sample of $t(\boldsymbol{\theta})$: $t_1 = t(\boldsymbol{\theta}_1), \cdots, t_n = t(\boldsymbol{\theta}_n)$

A lot simpler than obtaining the posterior of $t(\boldsymbol{\theta})$

Once a sample is available, one can obtain:

(i) point estimates,

(ii) credibility (or confidence) intervals,

(iii) marginal densities.

Example: $E_\pi(\boldsymbol{\theta}_i) \approx (1/n) \sum_j \boldsymbol{\theta}_{ij}$

Simulation via resampling

(Smith & Gelfand, 1992; Lopes, Schmidt & Moreira, 1998)

It is usually difficult to sample from posterior π.

Idea: use resampling techniques taking an auxiliary density q (that can be the prior of $\boldsymbol{\theta}$).

This density is known as importance density.

Using the rejection method (RM), there must exist c, such that $\pi(\boldsymbol{\theta}) \leq cq(\boldsymbol{\theta})$, $\forall \boldsymbol{\theta}$.

Resampling scheme:

(i) draw a value $\boldsymbol{\theta}$ from density $q(\boldsymbol{\theta})$;

(ii) accept this value with probability $w(\boldsymbol{\theta}) = \frac{\pi(\boldsymbol{\theta})}{cq(\boldsymbol{\theta})}$

If q=priori, $w(\boldsymbol{\theta}) = l(\boldsymbol{\theta})/l_{max}$.

Weighted resampling (SIR) scheme:

(i) draw sample $\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_n$ from density $q(\boldsymbol{\theta})$;

(ii) resample values from sample with probabilities

$$h_i = \frac{w(\boldsymbol{\theta}_i)}{\sum_{j=1}^n w(\boldsymbol{\theta}_j)} \quad , \quad i = 1, 2, \cdots, n.$$

If q=priori, $h_i = \frac{l(\boldsymbol{\theta}_i)}{\sum_{j=1}^n l(\boldsymbol{\theta}_j)}$

Problems:

a) RM: resample smaller than original sample;

b) both: handling conflict between prior and likelihood;

c) RM: likelihood maximization is required;

d) SIR: sample from π is approximated.

SIR in 2 steps

(Schmidt, Gamerman & Moreira, 1999)

1. use SIR as before as a 1st approximation, i.e., from q (eg, prior) obtain an approximate sample from π;

2. from sample, define the parameters of a new (refined) importance density q^*. Parameters may be calculated according to extremes in sample;

3. obtain a new resample from q^*. Resample with probabilities,

$$\omega_i = \frac{f_y(\theta_i)}{\sum_{i=1}^{N} f_y(\theta_i)} \quad ,$$

where

$$f_y(\theta_i) = \frac{l(\theta_i)p(\theta_i)}{q^*(\theta_i)}.$$

Approximations via MCMC

(Markov Chain Monte Carlo)

Consider DLM with

state vectors $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_n)$ and

hyperparameters $\boldsymbol{\lambda}$

Construct Markov chain with

transition according to posterior condicionals

$$\pi(\boldsymbol{\theta}|\boldsymbol{\lambda}) \text{ and } \pi(\boldsymbol{\lambda}|\boldsymbol{\theta})$$

$\Rightarrow$ equilibrium distribution is $\pi(\boldsymbol{\theta}, \boldsymbol{\lambda})$.

Draw a trajectory from this chain

$(\boldsymbol{\theta}^{(j)}, \boldsymbol{\lambda}^{(j)})$, $j = 0, 1, 2...$

This algorithm is known as Gibbs sampling.

After equilibrium, trajectory values are a sample from $\pi(\boldsymbol{\theta}, \boldsymbol{\lambda})$.

Example: consider the DLM where

- $\mathbf{F}_t$, $\mathbf{G}_t$ are known and
- $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_n)$, V and $\mathbf{W}$ are unknown

We need posterior conditional distributions $(\boldsymbol{\theta}, V|\mathbf{W})$ and $(\mathbf{W}|\boldsymbol{\theta}, V)$.

It can be shown that:

(i) $(\boldsymbol{\theta}, V|\mathbf{W})$ is NG $\rightarrow$ easy sampling (Fruhwirth-Schnatter, 1994; Carter & Kohn, 1994)

(ii) $(\mathbf{W}|\boldsymbol{\theta})$ is IW if prior for $\mathbf{W}$ is IW $\rightarrow$ easy sampling

When $\pi(\boldsymbol{\lambda}|\boldsymbol{\theta})$ is not easy to sample

- draw $\boldsymbol{\lambda}^{(n)}$ from a proposed transition $q(\boldsymbol{\lambda}|\boldsymbol{\lambda}^{(o)}, \boldsymbol{\theta})$

  Example: $\boldsymbol{\lambda}|\boldsymbol{\lambda}^{(o)} \sim N(\boldsymbol{\lambda}^{(o)}, c\mathbf{I})$

- accept $\boldsymbol{\lambda}^{(n)}$ with probability $\alpha(\boldsymbol{\lambda}^{(o)}, \boldsymbol{\lambda}^{(n)})$ where

$$\alpha(\boldsymbol{\lambda}^{(o)}, \boldsymbol{\lambda}^{(n)}) = \min\left\{1, \frac{\pi(\boldsymbol{\lambda}^{(n)}|\boldsymbol{\theta})}{\pi(\boldsymbol{\lambda}^{(o)}|\boldsymbol{\theta})} \frac{q(\boldsymbol{\lambda}^{(o)}|\boldsymbol{\lambda}^{(n)}, \boldsymbol{\theta})}{q(\boldsymbol{\lambda}^{(n)}|\boldsymbol{\lambda}^{(o)}, \boldsymbol{\theta})}\right\}$$

This is the Metropolis-Hastings algorithm (Metropolis et al., 1953; Hastings, 1970).

Again, Markov chain has equilibrium distribution given by $\pi(\boldsymbol{\theta}, \boldsymbol{\lambda})$.

Example: Dynamic GLM with

- $\mathbf{F}_t$, $\mathbf{G}_t$ known and
- $\boldsymbol{\theta}' = (\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_n)$ and $\mathbf{W}$ unknown

It can be shown that

(i) $(\mathbf{W}|\boldsymbol{\theta})$ is IW if prior for $\mathbf{W}$ is IW $\rightarrow$ easy sampling

(ii) $(\boldsymbol{\theta}|\mathbf{W})$ not known $\rightarrow$ difficult sampling

Build proposal based on

(i) the IRLS algorithm (Gamerman, 1998)

(ii) the prior (Knorr-Held, 2000)

(iii) likelihood approximations (Shephard and Pitt, 1997)

(iv) random walk forms