

CURSO DE MODELOS DINÂMICOS

Professor : Dani Gamerman

Assistente : Alexandra Mello Schmidt

Instituto de Pesquisa Econômica Aplicada

IPEA - DIPES

Agosto/Setembro 1997

1 Introdução

1.1 Noções básicas

Série temporal - conjunto de observações ordenadas (no tempo).

Tempo pode ser espaço, profundidade, ... ;

Observações vizinhas são dependentes;
modelagem

Estudo de séries temporais : análise dessa dependência.

Técnicas específicas a séries temporais

Exemplos de aplicações :

Economia - preços diários de ações
taxa de desemprego mensal

Medicina - níveis de eletrocardiograma
de eletroencefalograma

Epidemiologia - casos semanais de sarampo
mensais de AIDS

Metereologia - temperatura diária
registro de marés

⋮

1.2 Classificação

Série temporal é conjunto de observações $\{Y(t), t \in T\}$

Y - variável de interesse

T - conjunto de índices

1.3 Tipos de séries Temporais

1. Discreta : $T = \{t_1, t_2, \dots, t_n\}$

Ex : Exportações mensais de 1970 a 1990

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

Notação : Y_t

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

Ex : Registro da maré no Rio durante 1 dia

$T = [0, 24]$ se unidade de tempo é a hora

Notação : $Y(t)$

3. Multivariada : Observações são $\{Y_1(t), \dots, Y_k(t), t \in T\}$

Ex : Vendas semanais (Y_{1t}) e gastos com propaganda (Y_{2t}).

Y também pode ser discreto ou contínuo.

Muitas vezes, Y é discreto mas pode ser tratado como contínuo.

Ex : Número de casos notificados de AIDS.

Nesse curso, séries são univariadas ou multivariadas, discretas e observadas em tempos equiespaçados.

Podemos identificar T com $\{1, 2, \dots, n\}$

1.4 Objetivos de uma análise de séries temporais

Os principais objetivos são :

- (i) compreender o mecanismo gerador da série;
- (ii) prever o comportamento futuro da série.

Compreender o mecanismo da série possibilita :

- descrever eficientemente o comportamento da série;
- encontrar periodicidades na série;
- descobrir razões para o comportamento da série;
(possivelmente através de variáveis auxiliares)
- controlar a trajetória da série.

Predizer o futuro possibilita :

- fazer planos a longo, médio e curto prazo;
- tomar decisões apropriadas.

Objetivos (i) e (ii) estão interligados.

Só é possível prever bem rotineiramente se o modelo é adequado e vice-versa.

A não ser nos raros casos de modelos determinísticos, futuro envolve incerteza → previsões não são perfeitas.

Objetivo é reduzir ao máximo os **erros de previsão**.

1.5 Modelagem, aprendizado e previsão

Central à análise de séries temporais está a construção de um modelo.

Modelo - esquema de descrição (e explicação) que organiza informação (e experiências) de forma a propiciar aprendizagem e previsão.

Bom **modelo** permite **aprendizado** levando a **previsões** adequadas.

Devido à incerteza presente, modelo é probabilístico.

Deve também ser econômico (parsimônia).

Descrição deve ser relativamente simples e flexível para poder se adaptar ao futuro (incerto) e facilitar aprendizado.

Aprendizado é processamento de informação através do modelo.

Previsão é hipótese, conjectura ou especulação sobre o futuro.

A natureza dinâmica de séries temporais faz com que modelos tenham que ter adaptabilidade no tempo.

Tem que ser parametrizado de forma a permitir mudanças locais em sua estrutura.

Veremos que essa mudança pode ser modelada estocasticamente, isto é, usando probabilidade.

Quando isso é feito dentro do contexto Bayesiano, temos uma série de vantagens como intervenção, função de transferência, estimação retrospectiva (alisamento) sendo obtidas naturalmente.

A idéia básica é definir modelos que representem a estrutura da série (componente cíclica, de tendência, \dots).

Daí o nome de modelos estruturais.

Quando método de inferência é Bayesiano, modelos dinâmicos

Característica básica : parâmetros que caracterizam o modelo mudam com o tempo probabilisticamente.

Ex: Queremos prever Y que sabemos ser influenciada por X .

Relação mais simples : $Y = X\theta + \epsilon$

Talvez melhor assumir θ mudando com o tempo

Regra de aprendizagem - Teorema de Bayes

- Modelos possíveis M_1, M_2, \dots, M_p
com probabilidades $P(M)$, $M = M_1, M_2, \dots, M_p$;
- Observa-se Y cuja descrição é $P(Y | M)$; Verossimilhança do modelo M ;
- Após observar $Y = y$ temos $P(M | Y = y)$,
 $M = M_1, \dots, M_p$ dado por

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

Posteriori \propto Verossimilhança \times Priori

1.6 Aspectos importantes de sistemas de previsão

Básicos : previsão e estimação.

1. INTERVENÇÃO (ANTECIPATÓRIA)

Analista pode modificar modelo de acordo com informação prévia durante processo de observação.

Ex. : tabelamento dos juros

2. MONITORAÇÃO

Performance do modelo cai



Monitoração sinaliza



Mudanças são feitas

- Intervenção retrospectiva

3. RETROSPECÇÃO

Previsão : ver o que passado diz sobre futuro.

Também pode-se ver o que futuro diz sobre passado.

Importância secundária : controle.

2 Distribuições de Probabilidade

2.1 Operações com funções de densidade de probabilidade

Sejam x, y, z variáveis aleatórias e seja $f(x, y, z)$ a função densidade de probabilidade conjunta dessas três variáveis aleatórias.

Valem as seguintes identidades :

$$f(x|y) = \frac{f(x, y)}{f(y)} \quad (1)$$

$$f(x) = \int_y f(x, y)dy \quad (2)$$

$$f(x, y | z) = \frac{f(x, y, z)}{f(z)} \quad (3)$$

$$f(x, y) = \int_z f(x, y, z)dz \quad (4)$$

$$f(x | y) = \int_z f(x, z | y)dz \quad (5)$$

$$f(x | y, z) = \frac{f(y | x, z)f(x | z)}{f(y | z)} \quad (6)$$

Resultados continuam válidos para x, y , e z vetores.

2.2 Distribuição Normal Univariada

A v.a. aleatória x tem distribuição normal univariada com média μ e variância σ^2 , denotada por $N(\mu, \sigma^2)$, se sua densidade é dada por:

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2\sigma^2}(x - \mu)^2 \right\}, \quad x \in R \quad (7)$$

A distribuição normal padronizada é obtida quando $\mu = 0$ e $\sigma^2 = 1$.

$$E(x) = \mu \text{ e } V(x) = \sigma^2.$$

2.3 Distribuição Normal Multivariada

$x = (x_1, x_2, \dots, x_p)'$ tem distribuição normal multivariada (ou p -variada) com média μ e variância Σ , denotada por $N(\mu, \Sigma)$, se sua densidade é dada por :

$$f_N(x; \mu, \Sigma) = (2\pi)^{p/2} |\Sigma|^{1/2} \exp \left\{ -\frac{1}{2}(x - \mu)' \Sigma^{-1} (x - \mu) \right\} \quad (8)$$

onde $|A|$ denota o determinante da matriz A . A distribuição normal padronizada é obtida quando $\mu = 0$ e $\Sigma = I_p$, a matriz identidade de ordem p . Nesse caso, as componentes x'_i s são normais padronizadas independentes. A distribuição normal univariada é o caso particular onde $p = 1$.

Propriedades importantes

Transformações lineares : se A é uma matriz $r \times p$ e b é um vetor r -dimensional então

$$y = Ax + b \sim N(A\mu + b, A\Sigma A') \quad (9)$$

Distribuições Marginais : se o vetor x é dividido em 2 blocos x_1 contendo os primeiros r componentes de x e x_2 contendo os outros $p - r$ componentes então procedendo a mesma partição em μ e Σ na forma

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad e \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \quad (10)$$

obtém-se que $x_i \sim N(\mu_i, \Sigma_{ii})$, $i = 1, 2$.

Distribuições condicionais : mantendo as mesmas partições em x , μ e Σ obtém-se que

$$x_1 | x_2 \sim N(\mu_{1.2}, \Sigma_{11.2}) \quad (11)$$

onde $\mu_{1.2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)$ e $\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$. Resultados análogos são obtidos para a distribuição de $x_2 | x_1$ bastando a troca dos índices 1 e 2. Para esses resultados, é obviamente necessário que as submatrizes Σ_{22} e Σ_{11} respectivamente tenham posto máximo, caso contrário suas inversas não existem.

Reconstrução da conjunta :

se $x_1 \mid x_2 \sim N(\mu_1 + B_1(x_2 - \mu_2), B_2)$ e $x_2 \sim N(\mu_2, \Sigma_{22})$

então

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \sim N(\mu, \Sigma) \quad \text{com} \quad (12)$$

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \text{e} \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

onde $\Sigma_{11} = B_2 + B_1\Sigma_{22}B_1'$ e $\Sigma_{21}' = \Sigma_{12} = B_1\Sigma_{22}$.

Formas quadráticas : $(x - \mu)'\Sigma^{-1}(x - \mu) \sim \chi_p^2$.

2.4 Distribuição Gama

x tem distribuição Gama com parâmetros α e β se sua função de densidade for dada por:

$$p(x|\alpha; \beta) \propto x^{\alpha-1} \exp\{-\beta x\}, 0 < x < \infty. \quad (13)$$

Notação : $x \sim G(\alpha, \beta)$.

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

2.5 Distribuição Gama Invertida

x tem distribuição Gama invertida com parâmetros α e β se $x^{-1} \sim G(\alpha, \beta)$.

Sua função de densidade é dada por:

$$p(x|\alpha; \beta) \propto \left(\frac{1}{x}\right)^{\alpha+1} \exp\left\{-\frac{\beta}{x}\right\}, 0 < x < \infty. \quad (14)$$

Notação : $x \sim GI(\alpha, \beta)$.

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

2.6 Distribuição t-Student Univariada

x tem distribuição t-Student univariada com ν graus de liberdade, com parâmetros μ e σ se sua função densidade é dada por :

$$p(x|\mu, \sigma, \nu) \propto [\nu\sigma^2 + (x - \mu)^2]^{-(\nu+1)/2}, -\infty < x < \infty. \quad (15)$$

Quando $\mu = 0$ e $\sigma = 1$ diz-se que a variável aleatória tem distribuição t-Student padrão. Neste caso,

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

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

2.7 t-Multivariada

\mathbf{x} ($n \times 1$) segue uma distribuição t-multivariada com parâmetros $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$ e ν graus de liberdade, se sua função densidade é dada por:

$$p(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) \propto [\nu + (\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})]^{-(\nu+n)/2} \quad (16)$$

para $\mathbf{x} \in R^n$ onde $\boldsymbol{\mu} \in R^n$, $\boldsymbol{\Sigma} > 0$ ($n \times n$) e $\nu > 0$.

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

A média e variância de \mathbf{x} são :

$$E(\mathbf{x}) = \boldsymbol{\mu} \quad \text{se} \quad \nu > 1$$
$$V(\mathbf{x}) = \frac{\nu}{\nu - 2}\boldsymbol{\Sigma} \quad \text{se} \quad \nu > 2.$$

2.8 Wishart

Diz-se que uma matriz $\mathbf{\Omega}$ segue uma distribuição Wishart com parâmetro $\mathbf{\Sigma}$ e ν graus de liberdade, se e só se, sua função densidade é dada por :

$$p(\mathbf{\Omega}|\mathbf{\Sigma}, \nu) \propto |\mathbf{\Omega}|^{(\nu-n-1)/2} \exp \left\{ -\frac{1}{2} \text{tr} \mathbf{\Sigma} \mathbf{\Omega} \right\} \quad (17)$$

onde $\nu \geq n$, $\mathbf{\Sigma} > 0$ ($n \times n$) e $\text{tr} \mathbf{\Sigma}$ é o traço da matriz $\mathbf{\Sigma}$.

Notação : $\mathbf{\Omega} \sim W(\mathbf{\Sigma}, \nu)$.

2.9 Wishart Invertida

Diz-se que uma matriz $\mathbf{\Omega}$ ($n \times n$) segue uma distribuição Wishart-Invertida com parâmetro $\mathbf{\Sigma}$ e ν graus de liberdade, se e só se, $\mathbf{\Omega}^{-1} \sim W(\mathbf{\Sigma}, \nu)$. Sua função densidade é dada por :

$$p(\mathbf{\Omega}|\mathbf{\Sigma}, \nu) \propto |\mathbf{\Omega}|^{-(\nu+n+1)/2} \exp \left\{ -\frac{1}{2} \text{tr} \mathbf{\Sigma} \mathbf{\Omega}^{-1} \right\} \quad (18)$$

onde $\nu \geq n$, $\mathbf{\Sigma} > 0$ ($n \times n$) e $\text{tr} \mathbf{\Sigma}$ é o traço da matriz $\mathbf{\Sigma}$.

Notação : $\mathbf{\Omega} \sim WI(\mathbf{\Sigma}, \nu)$

Algumas propriedades das distribuições citadas acima :

1. Se $\mathbf{\Omega} \sim W(\mathbf{\Sigma}, \nu) \Rightarrow \mathbf{A} \mathbf{\Omega} \mathbf{A}' \sim W(\mathbf{A} \mathbf{\Sigma} \mathbf{A}', \nu)$;
2. Se $\mathbf{\Omega}^{-1} \sim WI(\mathbf{\Sigma}, \nu) \Rightarrow \mathbf{A} \mathbf{\Omega} \mathbf{A}' \sim WI(\mathbf{A} \mathbf{\Sigma} \mathbf{A}', \nu)$;

2.10 Normal-Gama

Suponha que

$$x|y \sim N(\mu, y^{-1}V) \quad \text{e} \quad y \sim G\left(\frac{\nu}{2}, \frac{d}{2}\right) \quad (\Leftrightarrow y^{-1} \sim GI\left(\frac{\nu}{2}, \frac{d}{2}\right)).$$

\Downarrow

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

\Downarrow

$$x \sim t_\nu(\mu, SV) \quad \text{onde} \quad S = \frac{d}{\nu}$$

2.11 Normal-Wishart

Suponha que $x|Y \sim N(\mu, VY^{-1})$ e $Y \sim W(\Sigma, \nu)$

\Downarrow

$$(x, Y) \sim NW(\mu, V, \nu, \Sigma)$$

\Downarrow

$$x \sim t_\nu\left(\mu, \frac{V\Sigma}{\nu}\right)$$

3 Inferência Bayesiana

Simple e resumida apresentação da metodologia

Ilustração com modelos dinâmicos.

Teorema de Bayes

Observações y : descritas por densidade ou função de probabilidade $f(y|\theta)$

Função de verossimilhança: $l(\theta) = f(y|\theta)$

Quantidade θ : indexador de f (parâmetro)

Situação canônica: amostra aleatória simples $y = (y_1, \dots, y_n)$ é extraída de $f(y|\theta)$.

Exemplo (1). medições sobre uma grandeza física θ com erros e_i descritos pela $N(0, \sigma^2)$, σ^2 é conhecida.

$$y_i = \theta + e_i, i = 1, \dots, n \text{ e}$$

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

θ é mais do que um simples indexador

Essa situação se repete em casos mais gerais

É bastante provável que o pesquisador saiba caracterizar sua incerteza a respeito de θ probabilisticamente.

Isso pode ser feito através de uma densidade $p(\theta)$

Historicamente, muita controvérsia a respeito mas menos importante agora

Processo de inferência baseado na distribuição de θ após observar y

→ distribuição a posteriori (em oposição à priori)

Obtida através do teorema de Bayes via

$$p(\theta|y) = \frac{f(y|\theta)p(\theta)}{f(y)} \quad \text{ou}$$
$$\pi(\theta) \propto l(\theta)p(\theta)$$

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

Exemplo (cont.) modelo pode ser completado com priori $p(\theta) = N(\mu, \tau^2)$, μ e τ^2 conhecidos.

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

onde \bar{y} é a média aritmética dos y_i 's.

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

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

ou seja, $\pi(\theta) = N(\mu_1, \tau_1^2)$.

$\tau^2 \rightarrow \infty$: priori não-informativa $p(\theta) \propto cte.$ e

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

Exemplo (2) Vamos supor agora que a **variância** da observação também é **desconhecida**. Nosso modelo passa a ser

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

Nosso parâmetro agora é bidimensional e para completar a distribuição a priori precisamos de uma distribuição marginal para σ^2 .

Por conveniência especificamos a priori para $\phi = \sigma^{-2}$ como $\phi \sim G\left(\frac{n}{2}, \frac{d}{2}\right) \Leftrightarrow \sigma^2 \sim GI\left(\frac{n}{2}, \frac{d}{2}\right)$. Logo $(\theta, \phi) \sim NG(a, R, n, d)$.

Como

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

Como

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

Daí, aplicando o teorema de Bayes obtemos

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

Como $f(\theta, \phi|y) = f(\theta|\phi, y)f(\phi|y) \propto f(\theta|\phi, y)$ pois $f(\phi|y)$ não depende de θ

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

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

Observe que a distribuição de $(\theta|\phi, y)$ tem a mesma forma que a distribuição de $\theta|\phi$.

$(\phi|y) \sim G\left(\frac{n+1}{2}, \frac{1}{2}\left[d + \frac{(y-a)^2}{R+1}\right]\right)$ que também tem a mesma forma que a distribuição a priori de ϕ .

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

Como $y = \theta + \epsilon$ onde $(\epsilon|\sigma^2) \sim N(0, \sigma^2)$ obtemos que

$$E(y|\sigma^2) = E(\theta|\sigma^2) + E(\epsilon|\sigma^2) = a$$

$$V(y|\sigma^2) = V(\theta|\sigma^2) + V(\epsilon|\sigma^2) = \sigma^2(R+1)$$

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

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

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

Priori não-informativa: muita controvérsia entre Bayesianos
⇒ Variância grande

Previsão de uma observação futura y após observar x

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

se y e x são independentes condicionalmente a θ

No caso multivariado: $\theta = (\theta_1, \dots, \theta_p)$

Densidade marginal a posteriori de θ_i é dada por

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

onde $\theta_{-i} = (\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_p)$

Para cada θ_i : profusão de distribuições condicionais

Sumarizadores importantes da posteriori:

a) locação: média, moda e mediana

b) dispersão: variância, desvio-padrão, precisão e curvatura na moda

Em geral, expressão da posteriori é muito complexa:
impossível obtenção analítica dessas quantidades

Exemplo priori é alterada da normal para a Cauchy

$$\pi(\theta) \propto \exp \left\{ -\frac{1}{2} \frac{(\bar{y} - \theta)^2}{\sigma^2/n} \right\} \frac{1}{\tau^2 + (\theta - \mu)^2}$$

Modelos de Regressão

Modelo de regressão linear normal

Observações $y = (y_1, \dots, y_n)'$ descritas por

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

x_{i1}, \dots, x_{ip} - valores das p variáveis explicativas para a i -ésima observação

β_1, \dots, β_p - coeficientes de regressão

Modelo pode ser escrito na forma matricial

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

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

Modelo Bayesiano completado com priori conjugada para β e $\phi = \sigma^{-2}$

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

Aplicando o teorema de Bayes, vem a posteriori

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

Detalhes de modelos de regressão

Estimação de máxima verossimilhança

Os estimadores de MV de $\boldsymbol{\beta}$ e σ^2 no modelo de regressão são

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

mas $S^2 = n\hat{\sigma}^2/(n-p)$ é o estimador não-viciado de σ^2 .

Suas distribuições amostrais são dadas por

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

Estimação Bayesiana

Os parâmetros da distribuição a posteriori de $\boldsymbol{\beta}$ e σ^2 são

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

Se a priori é não-informativa ($n_0 \rightarrow 0$ e $\mathbf{B}_0^{-1} \rightarrow \mathbf{0}$) então

$$\mathbf{b}_1 = \hat{\boldsymbol{\beta}}, \mathbf{B}_1 = (\mathbf{X}'\mathbf{X})^{-1}, n_1 = n-p \text{ e } n_1 S_1 = (n-p)S^2$$

As expressões das distribuições a posteriori são

$$\begin{aligned}\boldsymbol{\beta} &\sim t_{n-p}(\hat{\boldsymbol{\beta}}, S^2(\mathbf{X}'\mathbf{X})^{-1}) \text{ e} \\ (n-p)\frac{S^2}{\sigma^2} &\sim \chi_{n-p}^2 \Leftrightarrow \sigma^2 \sim GI\left(\frac{n-p}{2}, \frac{n-p}{2}S^2\right)\end{aligned}$$

Modelos lineares generalizados

extensão de modelos de regressão para a família exponencial

$$f(y_i|\theta_i) = a(y_i) \exp\{y_i\theta_i + b(\theta_i)\}$$

$$E(y_i|\theta_i) = \mu_i$$

$$g(\mu_i) = x_{i1}\beta_1 + \dots + x_{ip}\beta_p \quad , \quad i = 1, \dots, n$$

onde a função de ligação g é diferenciável.

$$\textit{Exemplo} \quad y_i|\pi_i \sim \textit{bin}(n_i, \pi_i), \quad i = 1, \dots, n$$

probabilidades π_i determinadas pelos valores de variável x

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

F - qq função de distribuição

Priori natural: $\beta \sim N(b_0, B_0) \rightarrow$ não é conjugada

4 Aplicações de inferência Bayesiana

Aula prática utilizando o software *First Bayes*

1. Amostra Aleatória Simples :

- (a) Uma amostra de 23 observações da densidade da terra será analisada. Inicialmente é suposto que σ é conhecido e assume-se distribuição normal. Assim a posteriori será obtida;
- (b) Serão criados conjuntos de dados com amostras parciais do exercício anterior para verificar a evolução da posteriori;
- (c) A posteriori do item (a) será novamente analisada, entretanto com a hipótese de que σ é desconhecido.

2. Modelo de regressão

- (a) Um modelo de regressão será ajustado para os dados dos 6 primeiros intervalos de tempo (em minutos) entre erupções do gêiser *Old Faithful*. O objetivo é verificar a relação entre o tamanho da erupção e a duração da mesma.

5 Regressão Multivariada

Considere o modelo

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{U} \quad (19)$$

onde

- $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_q)$ é uma matriz $(n \times q)$ com n observações de q variáveis dependentes (ou respostas);
- \mathbf{X} é uma matriz $(n \times k)$ com n observações de k variáveis explicativas;
- $\mathbf{B} = (\beta_1, \dots, \beta_q)$ é uma matriz $(k \times q)$ de parâmetros de regressão;
- $\mathbf{U} = (\mathbf{u}_1, \dots, \mathbf{u}_q) = (\mathbf{v}_1, \dots, \mathbf{v}_n)'$ é uma matriz $(n \times q)$ de erros aleatórios, não correlacionada com \mathbf{X} .

Assume-se ainda que:

$$\mathbf{v}_i \sim N(0, \Sigma), \quad \Sigma > 0 \quad \text{é} \quad q \times q \quad (20)$$

$$E(\mathbf{v}_i \mathbf{v}_j') = 0 \quad \forall i \neq j \quad i, j = 1, \dots, n \quad (21)$$

isto é as observações são independentes.

Notação : $\mathbf{U} \sim N(0, \mathbf{I}_n, \Sigma)$ isto é, \mathbf{U} segue uma distribuição normal matrizvariada com média 0, matriz de variância-covariância a esquerda \mathbf{I}_n (isso significa que as observações são independentes) e matriz de variância-covariância a direita Σ

(isso significa que as variáveis dependentes têm suas interdependências descritas por Σ).

Alternativamente podemos escrever (19) como :

$$\mathbf{y}_j = \mathbf{X}\boldsymbol{\beta}_j + \mathbf{u}_j \quad j = 1, \dots, q \quad (22)$$

Verossimilhança de \mathbf{U} :

$$p(\mathbf{U}|\Sigma) = p(\mathbf{U}'|\Sigma) = p(\mathbf{v}_1, \dots, \mathbf{v}_n|\Sigma) = \prod_{j=1}^n p(\mathbf{v}_j|\Sigma)$$

$\mathbf{v}_j \sim N(0, \Sigma)$ independentes. Assim temos que

$$\begin{aligned} p(\mathbf{U}|\Sigma) &\propto |\Sigma|^{-n/2} \exp \left\{ -\frac{1}{2} \sum_{j=1}^n \mathbf{v}_j' \Sigma^{-1} \mathbf{v}_j \right\} \\ p(\mathbf{U}|\Sigma) &\propto |\Sigma|^{-n/2} \exp \left\{ -\frac{1}{2} \text{tr} \Sigma^{-1} \mathbf{U}' \mathbf{U} \right\}. \end{aligned} \quad (23)$$

Como $\mathbf{U} = \mathbf{Y} - \mathbf{X}\mathbf{B}$

$$p(\mathbf{Y}|\mathbf{B}, \Sigma) \propto |\Sigma|^{-n/2} \exp \left\{ -\frac{1}{2} \text{tr} \Sigma^{-1} (\mathbf{Y} - \mathbf{X}\mathbf{B})' (\mathbf{Y} - \mathbf{X}\mathbf{B}) \right\} \quad (24)$$

A verossimilhança do modelo (19) é escrita como em (24), e será denotada por $L(\mathbf{B}, \Sigma|\mathbf{Y})$.

Estimação de Máxima Verossimilhança

Os estimadores de máxima verossimilhança de \mathbf{B} e Σ , no modelo (19) são

$$\hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \quad (25)$$

$$\hat{\Sigma} = \frac{1}{n}\mathbf{Y}'(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{Y} \quad (26)$$

supondo que

1. $\text{posto}(\mathbf{X}) = k$ para que $(\mathbf{X}'\mathbf{X})^{-1}$ exista; e
2. $n \geq q + k$.

Distribuições Amostrais

Sob a hipótese de que $\mathbf{U} \sim N(0, \mathbf{I}_n, \Sigma)$, pode ser demonstrado que a distribuição amostral de $(\hat{\mathbf{B}}|\Sigma)$ é

$$(\hat{\mathbf{B}}|\Sigma) \sim N(\mathbf{B}, (\mathbf{X}'\mathbf{X})^{-1}, \Sigma) \quad (27)$$

$$n\hat{\Sigma} \sim W(\Sigma, n - k) \quad (28)$$

A distribuição marginal de $\hat{\mathbf{B}}$ é uma distribuição t-matrizvariada com a seguinte especificação

$$\hat{\mathbf{B}} \sim T_{n-k}(\mathbf{B}, (\mathbf{X}'\mathbf{X})^{-1}, \hat{\Sigma}) \quad (29)$$

Inferência Bayesiana

A verossimilhança do modelo (19) pode ser reescrita como

$$L(\mathbf{B}, \Sigma | \mathbf{Y}) \propto |\Sigma|^{-(\nu+q+1)/2} \exp\left\{-\frac{n}{2} \text{tr} \Sigma^{-1} \hat{\Sigma}\right\} \quad (30)$$

$$\times |\Sigma|^{-k/2} \exp\left\{-\frac{1}{2} \text{tr} \Sigma^{-1} (\mathbf{B} - \hat{\mathbf{B}})' (\mathbf{X}'\mathbf{X}) (\mathbf{B} - \hat{\mathbf{B}})\right\}$$

onde $\nu = n - (q + k + 1)$.

Priori Informativa

Fazendo uma análise conjugada, a priori informativa seria da seguinte forma:

$$(\mathbf{B} | \Sigma) \sim N(\mathbf{B}_0, \mathbf{A}_0, \Sigma) \quad (31)$$

$$\Sigma \sim WI(\mathbf{G}, m) \quad (32)$$

Lembrando que

$$p(\mathbf{B}, \Sigma | \mathbf{Y}) \propto L(\mathbf{B}, \Sigma | \mathbf{X}, \mathbf{Y}) \times p(\mathbf{B}, \Sigma), \quad (33)$$

As densidades a posteriori de $(\mathbf{B} | \Sigma)$ e Σ são normal matriz-variada e Wishart-Invertida, respectivamente,

$$(\mathbf{B} | \Sigma, \mathbf{Y}) \sim N(\mathbf{B}_1, \mathbf{A}_1, \Sigma) \quad (34)$$

$$(\Sigma | \mathbf{Y}) \sim WI(\mathbf{G}_1, n + m) \quad (35)$$

onde

$$\begin{aligned} \mathbf{G}_1 &= n\hat{\Sigma} + \mathbf{G} + \hat{\mathbf{B}}' \mathbf{X}' \mathbf{X} \hat{\mathbf{B}} + \mathbf{B}'_0 \mathbf{A}_0^{-1} \mathbf{B}_0 - \mathbf{B}'_1 \mathbf{A}_1^{-1} \mathbf{B}_1 \\ \nu + m + 2q + k + 2 &= n - (q + k + 1) + m + 2q + k + 2 \\ &= n + m + q + 1 \end{aligned}$$

A posteriori marginal tem distribuição t-matricial,

$$(\mathbf{B}|\mathbf{Y}) \sim T_{n+m}(\mathbf{B}_1, \mathbf{A}_1, \mathbf{G}_1) \quad (36)$$

Previsão

$$\mathbf{Y}_1 = \mathbf{X}_1\mathbf{B} + \mathbf{U}_1, \quad \mathbf{U}_1 \sim N(0, \mathbf{I}_{n_1}, \Sigma)$$

então

$$\mathbf{Y}_1|\mathbf{Y} \sim T_{n+m}(\mathbf{Y}_1^*, \mathbf{A}_2, \mathbf{G}_2)$$

onde

$$\begin{aligned} \mathbf{Y}_1^* &= (\mathbf{I} - \mathbf{X}_1(\mathbf{X}'_1\mathbf{X}_1 + \mathbf{A}_1^{-1})^{-1}\mathbf{X}'_1)^{-1}\mathbf{X}_1(\mathbf{X}'_1\mathbf{X}_1 + \mathbf{A}_1^{-1})^{-1} \\ &\quad \times (\mathbf{X}'\mathbf{Y} + \mathbf{A}_0^{-1}\mathbf{B}_0) \end{aligned}$$

$$\mathbf{A}_2^{-1} = \mathbf{I} - \mathbf{X}_1(\mathbf{X}'_1\mathbf{X} + \mathbf{X}'\mathbf{X} + \mathbf{A}_0^{-1})^{-1}\mathbf{X}'_1$$

$$\begin{aligned} \mathbf{G}_2 &= \mathbf{Y}'\mathbf{Y} + \mathbf{B}'_0\mathbf{A}_0^{-1}\mathbf{B}_0 + \mathbf{G} \\ &\quad + (\mathbf{X}'\mathbf{Y} + \mathbf{A}_0^{-1}\mathbf{B}_0)'(\mathbf{X}'_1\mathbf{X}_1 + \mathbf{X}'\mathbf{X} + \mathbf{A}^{-1})^{-1}(\mathbf{X}'\mathbf{Y} + \mathbf{A}_0^{-1}\mathbf{B}_0) \\ &\quad - \mathbf{Y}_1^*\mathbf{X}_1(\mathbf{X}'_1\mathbf{X}_1 + \mathbf{X}'\mathbf{X} + \mathbf{A}_0^{-1})^{-1}(\mathbf{X}'\mathbf{Y} + \mathbf{A}_0^{-1}\mathbf{B}_0) \end{aligned}$$

5.1 Modelo Auto-Regressivo Vetorial (VAR)

Considere o seguinte modelo para a série y_t :

$$y_t = c + y_{t-1}\phi_1 + y_{t-2}\phi_2 + \cdots + y_{t-p}\phi_p + \epsilon_t, \quad (37)$$

onde

$$E(\epsilon_t) = 0 \quad \text{e} \quad E(\epsilon_t \epsilon_k) = \begin{cases} \sigma^2 & \text{se } t=k \\ \mathbf{0} & \text{caso contrário} \end{cases}. \quad (38)$$

Considere agora um conjunto de variáveis contidas num vetor linha \mathbf{y}_t de dimensão q . Um modelo auto-regressivo vetorial de ordem p , denotado como $VAR(p)$, é uma generalização multivariada das equações (37) e (38):

$$\mathbf{y}_t = \mathbf{c} + \mathbf{y}_{t-1}\Phi_1 + \mathbf{y}_{t-2}\Phi_2 + \cdots + \mathbf{y}_{t-p}\Phi_p + \epsilon_t. \quad (39)$$

1. \mathbf{c} é um vetor de constantes ($1 \times q$);
2. Φ_l é uma matriz ($q \times q$) de coeficientes autoregressivos para $l = 1, 2, \cdots, p$;
3. ϵ_t é um vetor ($1 \times q$) de ruídos brancos, isto é:

$$E(\epsilon_t) = \mathbf{0}$$
$$E(\epsilon_t \epsilon_k') = \begin{cases} \Omega & \text{se } t=k \\ \mathbf{0} & \text{caso contrário} \end{cases}$$

onde Ω é uma matriz simétrica positiva definida.

Ex. : Um modelo VAR de 3 variáveis pode ser especificado para as variáveis : investimento, renda e consumo.

Reescrevendo o modelo (39) na forma matricial (19),

Matriz de observações :

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}_n \\ \vdots \\ \mathbf{y}_1 \end{bmatrix} = \begin{bmatrix} y_{n1} & \cdots & y_{nq} \\ \vdots & \ddots & \vdots \\ y_{11} & \cdots & y_{1q} \end{bmatrix}_{n \times q}$$

Matriz de covariáveis ($n \times qp + 1$):

$$\begin{aligned} \mathbf{X} &= \begin{bmatrix} 1 & \overbrace{y_{n-1,1}, \cdots, y_{n-1,q}}^{\mathbf{y}_{n-1}} & \cdots & y_{n-p,1}, \cdots, y_{n-p,q} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & y_{0,1}, \cdots, y_{0,q} & \cdots & y_{1-p,1}, \cdots, y_{1-p,q} \end{bmatrix} \\ &= \begin{bmatrix} 1 & \mathbf{y}_{n-1} & \cdots & \mathbf{y}_{n-p} \\ 1 & \mathbf{y}_{n-2} & \cdots & \mathbf{y}_{n-p-1} \\ \vdots & \ddots & & \\ 1 & \mathbf{y}_0 & \cdots & \mathbf{y}_{1-p} \end{bmatrix} \end{aligned}$$

com blocos $1 \times q$.

Matriz de parâmetros :

$$\Phi_l = \begin{bmatrix} \phi_{l,11} & \cdots & \phi_{l,1q} \\ \vdots & \ddots & \vdots \\ \phi_{l,q1} & \cdots & \phi_{l,qq} \end{bmatrix} = [\phi_{l1} \cdots \phi_{lq}]$$

$$\mathbf{B}_{qp+1 \times q} = \begin{bmatrix} c_1 & \cdots & c_q \\ \phi_{11} & \cdots & \phi_{1q} \\ \phi_{21} & \cdots & \phi_{2q} \\ \vdots & \cdots & \vdots \\ \phi_{p1} & \cdots & \phi_{pq} \\ \beta_1 & \cdots & \beta_q \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} \mathbf{c} \\ \Phi_1 \\ \vdots \\ \Phi_p \end{bmatrix}$$

Inferência Bayesiana

Priori de Litterman : $\phi_{l,jk} \sim N(a_{ljk}, R_{ljk})$, independentes.

l : lag;

j : variável dependente;

k : variável resposta;

Defina :

$\hat{\sigma}_i^2$ - variância residual do ajuste de um modelo autoregressivo univariado com p defasagens na variável i .

$$a_{ljk} = \begin{cases} 1 & l = 1, j = k \\ 0 & c.c. \end{cases}$$
$$R_{ljk} = \frac{\lambda_{jk} \hat{\sigma}_j^2}{l^d \hat{\sigma}_k^2}$$

λ_{jk} - coeficiente de Litterman

Observe que quanto maior o lag l , o parâmetro da l^a defasagem da variável k para a explicação da variável j estará mais próximo de zero (a priori).

Em geral, $d=1$.

6 Modelos de Primeira Ordem

Algumas notações úteis

Como veremos o processo de inferência será sequencial, isto é, é feito a cada tempo t ou a cada observação y_t .

O processo se inicia portanto baseado na informação que detemos antes de observar os dados, quando $t = 0$.

Essa informação está concentrada no conjunto D_0 que é o conjunto contendo os Dados (subjativos ou não) que dispomos antes de observar a série a partir de $t=1$.

Toda afirmação sobre o futuro ($t > 0$) será condicional a D_0 .

Particular interesse está nas distribuições preditivas de $y_t|D_0$.

Quando chegamos ao tempo t , nossa informação está concentrada em D_t e é baseado nesse conjunto que faremos a inferência.

Em particular, queremos obter as distribuições preditivas de $y_{t+h}|D_t, h > 0$.

Quando o tempo muda, também muda a informação de que dispomos. Assim quando o tempo passa de t para $t+1$ nossa informação inclui não apenas D_t mas também a nova observação y_{t+1} . Se isso é tudo que aprendemos então $D_{t+1} = \{y_{t+1}, D_t\}$.

Mais geralmente podemos ter mais informação chegando além daquela obtida com as observações e podemos generalizar $D_{t+1} = \{I_{t+1}, D_t\}$ onde I_{t+1} contém toda informação

adicional obtida.

Caso a única informação adicional obtida em cada tempo t é a própria observação y_t então $D_t = \{D_0, y_1, y_2, \dots, y_t\}$.

Nesse caso, diremos estar na presença de um sistema **fechado**.

Sistema **aberto** passa por oposição a ser o sistema que admite entradas (de informação) outras que não apenas a série observada.

A idéia aqui é que as observações flutuam em torno de uma média. Essa média porém **não é estática** e está sujeita a pequenas variações ao longo do tempo.

O modelo para as observações é $(y_t|\mu_t) \sim N(\mu_t, V_t)$ onde suporemos a princípio que V_t é **conhecido** para todo t .

A novidade aqui é que as variações de μ_t são modeladas através de um passeio aleatório, isto é, $\mu_t = \mu_{t-1} + \omega_t$ onde $\omega_t \sim N(0, W_t)$.

A variação de μ com o tempo é essencialmente estocástica e vai depender dos erros (ou perturbações) da evolução ω_t .

Esse modelo embora seja o mais simples incorpora muito dos principais conceitos de modelagem dinâmica.

O primeiro conceito é o de evolução paramétrica caracterizada pela equação que relaciona sucessivos valores dos parâmetros.

Os erros ω_t controlam a evolução através de sua variância W_t . Quanto maior (menor) o seu valor, mais errático (suave) ele será.

A média zero garante a "constância local".

Podemos também pensar em μ como uma função contínua do tempo com uma expansão de Taylor :

$$\mu_{t+\Delta} = \mu_t + \textit{termos de ordem maior}.$$

No modelo adotado, os termos de ordem maior são substituídos por um ruído. Daí o nome modelo de primeira ordem.

O tipo de trajetória descrita depende da relação V_t/W_t :

W/V pequeno \rightarrow boa parte do movimento da série é devido às observações.

W/V grande \rightarrow movimentos devidos às observações mas também às variações do nível μ .

Um aspecto importante de um modelo é o seu comportamento preditivo. Para o presente modelo, previsões h-passos-a-frente no tempo t são baseadas em

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

Podemos também estudar o comportamento do modelo através da função de previsão $f_t(h) = E(y_{t+h}|D_t) = E(\mu_{t+h}|D_t) = E(\mu_t|D_t) = m_t$.

As previsões para o futuro são constantes como no modelo de alisamento exponencial simples.

Podemos formalizar o modelo através de :

$$\text{Equação de observação : } Y_t = \mu_t + \nu_t \quad \nu_t \sim N[0, V_t] \quad (40)$$

$$\text{Equação de Sistema : } \mu_t = \mu_{t-1} + \omega_t \quad \omega_t \sim N[0, W_t]$$

$$\mu_0 | D_0 \sim N[m_0, C_0]$$

onde ν_t, ω_t são todos independentes entre si e de $\mu_0 | D_0$

O sistema de inferência funciona da seguinte forma

$$\begin{array}{ccccc} \mu_{t-1} | D_{t-1} & \xrightarrow{\text{EVOLUCAO}} & \mu_t | D_{t-1} & \xrightarrow{\text{ATUALIZACAO}} & \mu_t | D_t \\ \text{post.} & & \text{priori} & & \text{post.} \\ & & \downarrow & & \\ & & Y_t | D_{t-1} & & \\ & & \text{previsao} & & \end{array}$$

A **evolução** é feita através da equação do sistema.

A **atualização** é feita através da incorporação da informação obtida em y_t usando o **teorema de Bayes**.

A **previsão** é feita com a distribuição marginal (ou preditiva) de y_t dada D_{t-1} , a informação disponível antes de observar y_t .

Teorema 6.1 *No MLD univariado, a previsão 1 passo a frente e as distribuições a posteriori, $\forall t$, são dadas por :*

(a) *Posteriori em $t - 1$:*

Para alguma média m_{t-1} e variância C_{t-1} ,

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

(b) *Priori em t :*

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

onde :

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

$$R_t = C_{t-1} + W_t$$

(c) *Previsão 1 passo a frente :*

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

onde :

$$f_t = m_{t-1}$$

$$Q_t = R_t + V_t$$

(d) *Posteriori em t :*

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

onde :

$$m_t = m_{t-1} + A_t e_t \quad A_t = R_t / Q_t \quad e_t = Y_t - f_t$$

$$C_t = R_t - A_t^2 Q_t.$$

Demonstração : Baseada nas contas da seção 2.1.

(b) Priori em t :

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

(c) Previsão 1 passo a frente:

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

(d) Posteriori em t :

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

Observações :

- (i) Como $f_t = f_{t-1}(1)$, e_t é o **erro de previsão** (1-passo-a-frente)
- (ii) Podemos reescrever m_t como $m_{t-1} + A_t(y_t - m_{t-1}) = A_t y_t + (1 - A_t)m_{t-1}$. Logo A_t é o **peso** (adaptativo) dado à observação mais recente y_t .

As distribuições preditivas h-passos-a-frente são dadas por

$$y_{t+h}|D_t \sim N(f_t(h), Q_t(h)) \text{ onde}$$
$$f_t(h) = m_t \quad Q_t(h) = C_t + \sum_{j=1}^h W_{t+j} + V_{t+h}$$

Modelo é constante se $V_t = V$ e $W_t = W, \forall t$.

Já vimos que ele é fechado (a informações externas) se

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

É um modelo mais restrito mas, interessantes propriedades são obtidas:

- (i) Quando $t \rightarrow \infty$, $A_t \rightarrow A$ e $C_t \rightarrow AV$ onde

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

O processo tem comportamento limite determinado por r .

(ii) Para t grande $m_t \approx Ay_t + (1 - A)m_{t-1}$.

Se a priori for vaga com C_0^{-1} próximo de zero então A_t é monotonicamente decrescente em t .

Se a priori for precisa com C_0 próximo de zero então A_t é monotonicamente crescente em t .

O estimador m_t de μ_t coincide para t grande com os sugeridos nos alisamentos exponencial simples de Holt e exponencial geral de Brown. O valor de A sofre portanto a mesma interpretação : quanto maior o seu valor, maior o peso dado a observações recentes.

(iii) Como temos as relações $m_t = m_{t-1} + A_t e_t$, $m_{t-1} = y_{t-1} - e_{t-1}$ temos que

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

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

$$\begin{aligned} y_t - y_{t-1} &= A_{t-1}e_{t-1} + e_t - e_{t-1} \\ &= e_t - (1 - A_{t-1})e_{t-1} \end{aligned}$$

Fazendo t grande e $\theta = 1 - A$ temos $y_t - y_{t-1} \approx e_t - \theta e_{t-1}$

Como no tempo t , $e_t | D_{t-1} \sim N(0, Q_t) \approx N(0, Q)$ para t grande.

y_t admite a representação de um modelo ARIMA(0,1,1) assintoticamente com os erros das médias móveis dados pelos erros de previsão.

(iv) Como $R_t = C_{t-1} + W$ no limite temos $R = C + W$

Sabemos também que $V^{-1} + R_t^{-1} = C_t^{-1}$ e no limite $V^{-1} + R^{-1} = C^{-1}$.

Além disso, no limite $V = C/A \Rightarrow AC^{-1} + R^{-1} = C^{-1} \Rightarrow R^{-1} = C^{-1}(1 - A) \Rightarrow R = C/(1 - A)$ que substituída acima fornece $W = \frac{A}{1-A}C$

Logo W implica num crescimento de $100.A/(1 - A)\%$ da variância do sistema. Como o limite A é atingido rapidamente, pode-se adotar como norma uma taxa constante de crescimento da variância. Fazendo $\delta = 1 - A$ como o fator de desconto constante teríamos $R_t = C_{t-1}/\delta$ ou $R_t^{-1} = \delta C_{t-1}^{-1}$.

Tomando o inverso da variância como medida da informação do sistema teríamos uma queda constante $\delta \in (0, 1)$ para cada passagem no tempo. Os valores correspondentes de W_t seriam dados através de

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

Esse conceito de fator de desconto será generalizado mais tarde onde será relacionado com o fator de desconto do alisamento exponencial geral de Brown.

7 Modelos Dinâmicos (Lineares)

Os modelo já vistos são todos casos particulares de uma estrutura linear mais geral. Veremos que essa estrutura engloba além dos modelos já vistos, uma série de modelos também úteis.

O modelo é novamente descrito por duas equações :

$$\text{Eq. Obs. : } \mathbf{y}_t = \mathbf{F}'_t \boldsymbol{\theta}_t + \boldsymbol{\nu}_t \quad \boldsymbol{\nu}_t \sim N[0, \mathbf{V}_t]$$

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

$$\text{Info. inicial : } (\boldsymbol{\theta}_0 | D_0) \sim N[\mathbf{m}_0, \mathbf{C}_0] \quad (41)$$

- \mathbf{y}_t vetor de observações $r \times 1$;
- $\boldsymbol{\theta}_t$ vetor de parâmetros $n \times 1$;
- \mathbf{F}_t matriz conhecida $n \times r$;
- \mathbf{G}_t matriz conhecida $n \times n$;
- \mathbf{V}_t matriz de covariância conhecida $r \times r$;
- \mathbf{W}_t matriz de covariância conhecida $n \times n$;

Podemos também definir o modelo através da quádrupla $\{\mathbf{F}, \mathbf{G}, V, \mathbf{W}\}_t$.

Exemplos :

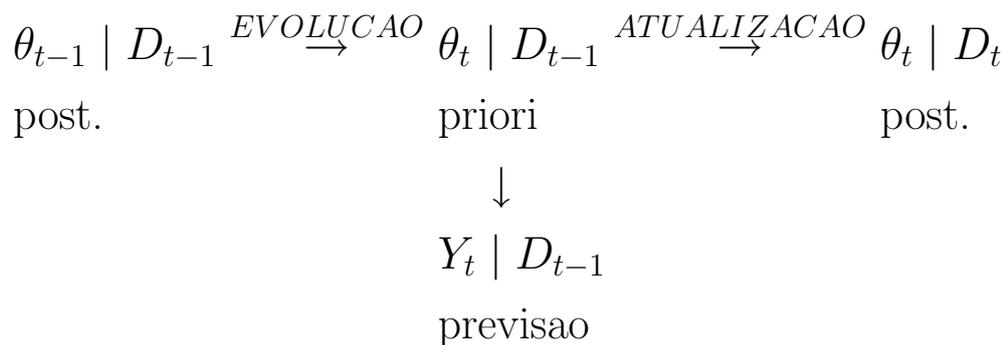
(i) Modelo de 1a. Ordem :

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

(ii) Modelo de regressão dinâmica através da origem com n regressores :

$\mathbf{F}'_t = (x_{t,1}, \dots, x_{t,n})$, $\mathbf{G} = \mathbf{I}_n$, $\boldsymbol{\theta}'_t = (\beta_{t,1}, \dots, \beta_{t,n})$ onde \mathbf{I}_n é matriz identidade de ordem n .

O ciclo de inferência permanece o mesmo que no modelo de 1^a ordem



As distribuições de interesse são obtidas através do seguinte teorema

Teorema 7.1 *No MLD univariado, a previsão 1 passo a frente e as distribuições a posteriori, $\forall t$, são dadas por :*

(a) *Posteriori em $t - 1$:*

Para alguma média \mathbf{m}_{t-1} e matriz de variância \mathbf{C}_{t-1} ,

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

(b) *Priori em t :*

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

$$\mathbf{a}_t = \mathbf{G}_t \mathbf{m}_{t-1}$$

$$\mathbf{R}_t = \mathbf{G}_t \mathbf{C}_{t-1} \mathbf{G}_t' + \mathbf{W}_t$$

(c) *Previsão 1 passo a frente :*

$$\mathbf{Y}_t \mid D_{t-1} \sim N[\mathbf{f}_t, \mathbf{Q}_t], \text{ onde :}$$

$$\mathbf{f}_t = \mathbf{F}_t' \mathbf{a}_t$$

$$\mathbf{Q}_t = \mathbf{F}_t' \mathbf{R}_t \mathbf{F}_t + \mathbf{V}_t$$

(d) *Posteriori em t :*

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

$$\mathbf{m}_t = \mathbf{a}_t + \mathbf{A}_t \mathbf{e}_t \quad \mathbf{A}_t = \mathbf{R}_t \mathbf{F}_t \mathbf{Q}_t^{-1} \quad \mathbf{e}_t = \mathbf{Y}_t - \mathbf{f}_t$$

$$\mathbf{C}_t = \mathbf{R}_t - \mathbf{A}_t \mathbf{A}_t' \mathbf{Q}_t.$$

Demonstração : Livro texto

7.1 Distribuições Preditivas

Definição 7.1 Para qualquer tempo corrente t , a função de previsão $f_t(k)$ é definida para todos os inteiros $k \geq 0$ como

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

onde

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

é a função de resposta média.

Teorema 7.2 Para cada tempo t e $k \geq 1$, as distribuições para $\boldsymbol{\theta}_{t+k}$ e Y_{t+k} dado D_t , k passos a frente, são dadas por:

(a) *Distribuição de Estado :*

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

(b) *Distribuição de Previsão :*

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

com momentos definidos recursivamente por :

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

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

Corolário 7.1 *No caso especial em que a matriz de evolução \mathbf{G}_t é constante, $\mathbf{G}_t = \mathbf{G}$ para todo t , então, para $k \geq 0$,*

$$\mathbf{a}_t(k) = \mathbf{G}^k \mathbf{m}_t \quad (42)$$

tal que

$$f_t(k) = \mathbf{F}'_{t+k} \mathbf{G}^k \mathbf{m}_t \quad (43)$$

Se, em adição, $\mathbf{F}_t = \mathbf{F}$ para todo t , o modelo é um MLDST, e a função de previsão tem a forma

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

Importante deste resultado é que previsões são governadas por potências de \mathbf{G} .

Modelos dinâmicos Bayesianos de Componentes Comuns (MDBCC)

Todas as q componentes de \mathbf{y} tem as mesmas variáveis explicativas, mas com coeficientes diferentes.

Modelo :

Para cada variável temos que :

$$y_{tj} = \mathbf{F}'_t \boldsymbol{\theta}_{tj} + v_{tj}, \quad j = 1, \dots, q,$$

na forma matricial temos

$$\begin{aligned} \mathbf{y}'_t &= \mathbf{F}'_t \boldsymbol{\Theta}_t + \mathbf{v}_t, & \mathbf{v}_t &\sim N(0, V_t \boldsymbol{\Sigma}) \\ \boldsymbol{\Theta}_t &= \mathbf{G}_t \boldsymbol{\Theta}_{t-1} + \boldsymbol{\Omega}_t, & \boldsymbol{\Omega}_t &\sim N(0, \mathbf{W}_t, \boldsymbol{\Sigma}) \end{aligned}$$

Se $\boldsymbol{\Theta}_t = \boldsymbol{\Theta}_{t-1} = \boldsymbol{\Theta}$, $\forall t$, recai-se no modelo de regressão multivariada (19) com

$$\mathbf{Y} = \begin{pmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} \mathbf{F}_1 \\ \vdots \\ \mathbf{F}_n \end{pmatrix} \quad e \quad \mathbf{B} = \boldsymbol{\Theta}$$

Ex. : BVAR dinâmico $\rightarrow \Phi_l$ mudam com o tempo.

8 Aplicações de BVAR e de modelos de 1a. Ordem

Aula prática usando o software *PRVWIN*

1. análise de um modelo VAR para a indústria geral, utilizando variáveis que precedem a indústria.
2. análise de um modelo de 1a. ordem da série produto da indústria geral do Brasil;

9 Modelos de Segunda Ordem (Crescimento Linear)

A primeira extensão sobre os modelos de 1a. ordem é permitir que a expansão de Taylor inclua o termo linear. Nesse caso,

$$\mu_{t+\Delta t} = \mu_t + \Delta t \mu'_t + \text{termos de ordem superior em } \Delta t.$$

Podemos definir além do nível μ , o seu crescimento β . Teríamos então a equação de diferença

$$\mu_{t+1} = \mu_t + \beta_{t+1} + \omega_{1,t+1} \text{ com } \omega_{1,t+1} \sim N(0, W_{1,t+1})$$

onde $\omega_{1,t+1}$ é a perturbação que substitui os termos de ordem superior. Assim, como anteriormente, a forma mais simples de modelar a evolução do crescimento β é através de um passeio aleatório

$\beta_{t+1} = \beta_t + \omega_{2,t+1}$ com $\omega_{2,t+1} \sim N(0, W_{2,t+1})$ independente de $\omega_{1,t+1}$.

O crescimento está sujeito a pequenas flutuações locais.

A equação do sistema consiste portanto das duas equações acima. O modelo completo é

$$\text{Eq. Obs. : } y_t = \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} \mu_t \\ \beta_t \end{pmatrix} + v_t, v_t \sim N(0, V_t)$$

$$\text{Eq. Sist. : } \begin{pmatrix} \mu_t \\ \beta_t \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_{t-1} \\ \beta_{t-1} \end{pmatrix} + \omega_t$$

$$\text{Info. Inicial : } \begin{pmatrix} \mu_0 \\ \beta_0 \end{pmatrix} \sim N \left(\begin{pmatrix} m_0 \\ b_0 \end{pmatrix}, \mathbf{C}_0 \right)$$

onde

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

alternativamente podemos ter $V(\boldsymbol{\omega}_t) = \text{diag}(W_{1,t}, W_{2,t})$.

Observe que o modelo procura descrever séries com tendência "localmente linear", pois a taxa de crescimento β está sujeita a pequenas flutuações. Vale enfatizar também que β pode ser negativo caracterizando um decrescimento na série apesar do nome adotado.

O comportamento preditivo pode ser avaliado através de

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

Mas

$$\begin{aligned} \mu_{t+h} &= \mu_{t+h-1} + \beta_{t+h-1} + \text{soma de erros} \\ &= \mu_{t+h-2} + 2\beta_{t+h-2} + \text{soma de erros} \\ &= \mu_t + h\beta_t + \text{soma de erros} \end{aligned}$$

Como todos os erros considerados tem média 0,

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

A função de previsão é dada por

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

que é função linear de h, como esperado.

O processo de inferência é realizado como descrito nas seções anteriores.

Usando a teoria normal e o fato de $D_t = \{y_t, D_{t-1}\}$ temos que:

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

$$\mathbf{m}_t = \mathbf{a}_t + \mathbf{S}_t \mathbf{Q}_t^{-1} e_t \text{ e } \mathbf{C}_t = \mathbf{R}_t - \mathbf{S}_t \mathbf{Q}_t^{-1} \mathbf{S}_t'$$

onde $e_t = y_t - f_t$ e $\mathbf{S}_t = \mathbf{R}_t \mathbf{F}_t$.

Desta forma, temos que

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

Como $y_t = f_t + e_t$ e $f_t = m_{t-1} + b_{t-1}$, podemos escrever

$$y_t = m_{t-1} + b_{t-1} + e_t$$

$$m_t = m_{t-1} + b_{t-1} + A_{t,1}e_t$$

$$b_t = b_{t-1} + A_{t,2}e_t$$

Tomando duas diferenças na primeira equação acima e usando as seguintes chega-se a

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

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

Pode-se mostrar que, assim como no modelo de 1a. ordem, podemos obter resultados limites para o modelo constante ($V_t = V$ e $W_t = W$).

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

O modelo no limite atinge a forma constante (independente de t)

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

que é a forma de um processo ARIMA(0,2,2) com os erros de médias móveis dados pelos erros de previsão.

Além disso, as estimativas dos parâmetros são atualizados segundo

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

O modelo de alisamento biparamétrico exponencial de Holt atualiza as estimativas segundo

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

Logo podemos reescrever $m_t = m_{t-1} + b_{t-1} + \alpha e_t$ e $b_t = b_{t-1} + \gamma(m_t - m_{t-1} - b_{t-1}) = b_{t-1} + \alpha\gamma e_t$.

Se tomamos $A_1 = \alpha$ e $A_2 = \alpha\gamma$ obtemos as equações de Holt como limite em um modelo constante.

O alisamento exponencial geral aplicado a uma reta estima μ e β através da minimização de

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

pode-se mostrar que os valores de μ e β que minimizam a expressão acima são atualizadas à medida que uma nova observação é obtida através de

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

onde $e_t = y_t - m_{t-1} - b_{t-1}$ é o mesmo erro de previsão já considerado.

Podemos definir

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

Logo

$$V \left[\begin{pmatrix} \mu_t \\ \beta_t \end{pmatrix} \middle| D_{t-1} \right] = \mathbf{R}_t = \mathbf{G}\mathbf{C}_{t-1}\mathbf{G}' + V(\boldsymbol{\omega}_t) = \mathbf{P}_t + V(\boldsymbol{\omega}_t)$$

O papel principal das perturbações $\boldsymbol{\omega}_t$ é justamente o de "aumentar" a variância do sistema já que normalmente tem média 0.

9.1 Modelos de Tendência Polinomial

Os modelos polinomiais de 1a. e 2a. ordens são casos particulares de uma estrutura mais geral. Um modelo de tendência polinomial de n-ésima ordem é definido por :

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

$\theta_{t,j}$ muda em relação a seu antecessor $\theta_{t-1,j}$ por $\theta_{t-1,j+1}$.

Exemplos :

1. MLD de tendências quadráticas :

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

$\boldsymbol{\theta}_t = (\mu_t, \beta_t, \gamma_t)'$, μ_t representa o nível, β_t tendência e γ_t representa a mudança da tendência. Este modelo é chamado de tendência quadrática porque sua função de previsão k passos a frente é da seguinte forma :

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

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

2. MLD de 2a. ordem : $\boldsymbol{\theta}_t = (\mu_t, \beta_t)'$

3. MLD de 1a. ordem : $\boldsymbol{\theta}_t = \mu_t$

10 Modelos Sazonais

Sazonalidade é efeito comum a muitas séries e tem que ser considerada.

Consideramos aqui apenas descrição de efeitos cíclicos.

Possíveis explicações podem ser dadas por covariáveis.

10.1 Definições

Seja $g(t)$ qualquer função real definida para os inteiros não-negativos, onde t é o indexador de tempo.

1. $g(t)$ é cíclica ou periódica se, para algum inteiro $p \geq 1$, $g(t + np) = g(t)$, para todo inteiro $t \geq 0$ e $n \geq 0$.
2. O menor inteiro p para o qual o ítem anterior é válido é chamado de período de $g(\cdot)$.
3. $g(\cdot)$ exibe um ciclo completo em qualquer intervalo de tempo que contenha p pontos consecutivos, como $(t, t + p - 1)$, $\forall t \geq 0$.
4. Os fatores sazonais de $g(\cdot)$ são os p valores de qualquer ciclo completo :

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

Note que, para $t > 0$, $g(t) = g(j)$ onde j é o resto da divisão de t por p . ($j = p \mid t$)

5. O vetor de fatores sazonais no tempo t é simplesmente o vetor de fatores sazonais permutados, de modo que o primeiro elemento é aquele para o tempo t

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

quando o fator sazonal corrente é θ_j . Em particular, para quaisquer inteiros n e $k = np$,

$$\boldsymbol{\theta}_k = (\theta_0, \dots, \theta_{p-1})'.$$

6. Em qualquer ciclo, o tempo correspondente ao fator sazonal θ_j tem um rótulo, $M(j)$. Claramente, os rótulos são cíclicos com período p e $j = p \mid t$ se, e só se, $M(j) = t$.

10.2 Caracterização de fatores sazonais ou cíclicos

Duas possibilidades : usando indicadores ou funções trigonométricas.

Podemos associar a cada ponto no tempo um vetor de fatores sazonais.

fator θ_1 , sazonais no tempo

Caso o modelo consista apenas na parte sazonal temos que o nível da série no tempo t , μ_t é dado pelo primeiro componente de θ_t e portanto pode ser obtido como $\mu_t = \mathbf{E}'_p \theta_t$ onde

$$\mathbf{E}'_p = (1, \underbrace{0, 0, \dots, 0}_{(p-1) \text{ termos}}).$$

Como o tempo t corresponde ao j -ésimo período, $t + 1$ corresponde ao $(j + 1)$ -ésimo. Permutando o vetor θ_t obtemos $\theta'_{t+1} = (\theta_{j+1}, \dots, \theta_p, \theta_1, \dots, \theta_j)$.

$$\mu_{t+1} = \mathbf{E}'_p \theta_{t+1}.$$

A passagem de θ_t para θ_{t+1} é realizada através da matriz

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

isto é, $\theta_{t+1} = \mathbf{P}_p \theta_t$ onde \mathbf{P}_p satisfaz a $\mathbf{P}_p^{k+np} = \mathbf{P}_p^k$ e, em particular, $\mathbf{P}_p^{np} = \mathbf{I}_p$.

Essa formulação leva naturalmente à definição do modelo sazonal como

$$\text{Eq. das Obs. : } y_t = \mathbf{E}'_p \theta_t + \nu_t \quad \nu_t \sim N[0, V_t] \quad (44)$$

$$\text{Eq. Sist. : } \theta_t = \mathbf{P}_p \theta_{t-1} + \omega_t \quad \omega_t \sim N[0, \mathbf{W}_t] \quad (45)$$

$$\text{Inf. Inicial : } (\boldsymbol{\theta}_0|D_0) \sim N(\mathbf{m}_0, \mathbf{C}_0) \quad (46)$$

Observações :

- (i) Se $\mathbf{W}_t = c\mathbf{I}_p$, os erros $\boldsymbol{\omega}_t$ tem componentes não correlacionadas e o problema se reduz à análise de p modelos de 1a. ordem, se a priori $\mathbf{C}_0 = \text{diag}(c_1, \dots, c_p)$. Isso porque não haverá passagem de informação entre diferentes níveis de ciclo sazonal;
- (ii) Para obter a função de previsão vamos supor que $\boldsymbol{\theta}_t = (\theta_1, \dots, \theta_p)'_t$ e, portanto, $E(\boldsymbol{\theta}_t|D_t) = \mathbf{m}_t = (m_{t,1}, \dots, m_{t,p})'$. A previsão h-passos-a-frente é dada por

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

Usualmente a modelagem por indicadores sazonais é feita pelo estabelecimento de um nível médio e variações sazonais com relação a este nível médio μ_t . Nesse caso, os fatores sazonais satisfazem a $\mathbf{1}'\boldsymbol{\theta}_t = \sum_{j=1}^p \theta_{t,j} = 0$.

O nível de cada tempo t é dado por

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

Essa modelagem impõe restrições sobre a modelagem de $\boldsymbol{\theta}_t$:

- 1) Priori inicial : Como $\boldsymbol{\theta}_0 | D_0 \sim N(\mathbf{m}_0, \mathbf{C}_0)$ e $\mathbf{1}'\boldsymbol{\theta}_0 = 0$ temos que

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

Se a especificação inicial não satisfaz às condições acima, pode-se impor a condição adicionalmente para a partir da especificação incorreta $\boldsymbol{\theta}_0 | D_0 \sim N(\mathbf{m}_0^*, \mathbf{C}_0^*)$ obter

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

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

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

Prova: Defina $\boldsymbol{\psi} = \mathbf{L}\boldsymbol{\theta}$ onde $\mathbf{L} = \begin{pmatrix} \mathbf{I} \\ \mathbf{1}' \end{pmatrix}$. $\boldsymbol{\psi} = \begin{pmatrix} \boldsymbol{\theta} \\ \mathbf{1}'\boldsymbol{\theta} \end{pmatrix}$ e obtenha $\boldsymbol{\theta} | \mathbf{1}'\boldsymbol{\theta}$.

- 2) Restrições são preservadas pelo modelo dinâmico com $\mathbf{1}'\boldsymbol{\omega}_t = 0$.

Uma vez garantidas essas restrições podemos redefinir o modelo sazonal como

$$\begin{aligned} \text{Eq. Observações: } y_t &= \mu_t + \theta_{t,j} + \nu_t \quad \nu_t \sim N[0, V_t] \\ &= \mu_t + \mathbf{E}'_p \boldsymbol{\theta}_t + \nu_t \\ &= \mathbf{F}' \boldsymbol{\beta}_t + \nu_t \end{aligned}$$

$$\begin{aligned} \text{Eq. do Sistema : } \mu_t &= \mu_{t-1} + \omega_{0,t}, \quad \omega_{0,t} \sim N(0, W_{0,t}) \\ \boldsymbol{\theta}_t &= \mathbf{P}_p \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t \quad \boldsymbol{\omega}_t \sim N(0, \mathbf{W}_t) \end{aligned} \quad (47)$$

$$\text{Inf. Inicial : } (\mu_0 | D_0) \sim N(m_{00}, \mathbf{C}_{00}) \quad (\boldsymbol{\theta}_0 | D_0), \sim N(\mathbf{m}_0, \mathbf{C}_0)$$

$$\text{Restrição adicional : } \mathbf{1}' \boldsymbol{\theta}_t = 0, \quad \forall t.$$

Supondo novamente que $E[\boldsymbol{\beta}_t | D_t] = (m_{t,0}, m_{t,1}, \dots, m_{t,p})'$ onde $\sum_{j=1}^p m_{t,j} = 0$ temos a função de previsão dada por

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

11 Representação Trigonométrica da Sazonalidade

A sazonalidade de p períodos pode ser representada tanto por fatores (ou indicadores) sazonais $\theta_1, \dots, \theta_p$ quanto por combinações de funções trigonométricas pois podemos escrever θ_j , $j = 1, \dots, p$ como

$$\theta_j = \begin{cases} a_0 + \sum_{r=1}^{q-1} [a_r \cos(\omega r t) + b_r \sin(\omega r t)], & p \text{ ímpar, } q = (p+1)/2 \\ a_0 + \sum_{r=1}^{q-1} [a_r \cos(\omega r t) + b_r \sin(\omega r t)] + a_q \cos(\pi t), & p \text{ par, } q = p/2 \end{cases}$$

onde o tempo t corresponde ao j -ésimo período no ciclo sazonal $\omega = 2\pi/p$.

Cada componente da soma constitui um harmônico.

Ele pode ser reescrito como $A_r \cos(\omega r t + \phi_r)$, onde

A_r é a amplitude do harmônico de ordem r ;

ϕ_r é a fase do harmônico de ordem r , $r = 1, \dots, q$.

A grande vantagem dessa representação é a economia que ela pode fornecer pois se algum harmônico for pouco relevante (A_r pequeno) ele pode ser excluído.

Ex. : Em séries mensais de certos setores da economia, o primeiro harmônico é suficiente.

Para compreender a modelagem dinâmica vamos nos concentrar no modelo mais simples, i.e., com apenas 1 harmônico ($r = 1$).

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

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

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

$$\begin{pmatrix} a_0 \\ \beta_1 \\ \gamma_1 \end{pmatrix} \rightarrow \begin{pmatrix} a_0 \\ a \\ b \end{pmatrix} \text{ é 1 a 1.}$$

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

Analogamente,

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

Defina

$$\mathbf{G}_1 = \begin{pmatrix} \cos \omega & \sin \omega \\ -\sin \omega & \cos \omega \end{pmatrix} \text{ e } \mathbf{G} = \text{diag}(1, \mathbf{G}_1) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \omega & \sin \omega \\ 0 & -\sin \omega & \cos \omega \end{pmatrix}$$

Daí

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

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

e $-\beta_{1,t} \sin \omega + \gamma_{1,t} \cos \omega = \gamma_{1,t+1}$ por contas semelhantes.

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

Este raciocínio pode ser estendido a todos os harmônicos pois

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

Definindo $\beta_{r,t} = a_r \cos(\omega r t) + b_r \sin(\omega r t)$, $r = 1, \dots, q-1$

$$\gamma_{r,t} = -a_r \sin(\omega r t) + b_r \cos(\omega r t)$$

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

temos $\theta_j = \mathbf{F}'_t \boldsymbol{\beta}_t$ e $\boldsymbol{\beta}_{t+1} = \mathbf{G}\boldsymbol{\beta}_t$ onde

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

O dinamismo pode ser completado através da incorporação dos erros de observação e do sistema levando a

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

$$\text{Eq. Sist. : } \boldsymbol{\beta}_{t+1} = \mathbf{G} \boldsymbol{\beta}_t + \boldsymbol{\omega}_{t+1} \quad \boldsymbol{\omega}_{t+1} \sim N[0, \mathbf{W}_{t+1}]$$

$$\text{Inf. Inicial : } (\boldsymbol{\beta}_0 | D_0) \sim N(\mathbf{m}_0, \mathbf{C}_0)$$

onde

$$\mathbf{G} = \begin{cases} \text{diag}(1, \mathbf{G}_1, \dots, \mathbf{G}_{q-1}) , \text{ p ímpar;} \\ \text{diag}(1, \mathbf{G}_1, \dots, \mathbf{G}_{q-1}, -1) , \text{ p par.} \end{cases}$$

$$\mathbf{F} = \begin{cases} (1 \ 1 \ 0 \ 1 \ 0 \ \dots \ 1 \ 0) , \text{ p ímpar;} \\ (1 \ 1 \ 0 \ 1 \ 0 \ \dots \ 1 \ 0 \ 1) , \text{ p par.} \end{cases}$$

Escolha dos harmônicos

A descrição completa da sazonalidade envolve a especificação e inclusão no modelo de $q - 1$ harmônicos completos.

Como já dissemos, nem sempre necessitamos de todos eles.

Com dados mensais, podemos modelar com o 1^o e o 4^o harmônicos, por exemplo. O primeiro representaria o ciclo anual e o 4^o corresponderia a um padrão trimestral.

Caso não saibamos se harmônicos devem ou não ser excluídos basta verificar sua amplitude. Para isso, observe que os parâmetros $\beta_{r,t}$ e $\gamma_{r,t}$ correspondentes ao r -ésimo harmônico satisfaçam a $\sqrt{\beta_{r,t}^2 + \gamma_{r,t}^2} = \sqrt{a_r^2 + b_r^2} = A_r$, a amplitude do harmônico. Estimativas de $\beta_{r,t}$ e $\gamma_{r,t}$ fornecem indicação da significância de A_r .

12 Aplicação

Aula prática usando o software PRV.

1. Análise da série produto da indústria geral do Brasil incluindo componentes de tendência e sazonalidade;
2. Comparação da capacidade preditiva dos modelos até aqui analisados.

13 Superposição de Modelos e Fatores de Descontos

A estrutura linear dos modelos permite que sejam combinadas várias componentes em um único modelo.

Considere os modelos 1 e 2 dados por :

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

Se definimos y_t como a soma das duas séries temos que y_t satisfaz a $y_t = y_{1,t} + y_{2,t} = \mathbf{F}'_{1,t}\boldsymbol{\theta}_{1,t} + \mathbf{F}'_{2,t}\boldsymbol{\theta}_{2,t} + v_{1,t} + v_{2,t}$.

Se definimos

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

$v_t = v_{1,t} + v_{2,t}$ e $V_t = V_{1,t} + V_{2,t}$ temos

$y_t = \mathbf{F}'_t\boldsymbol{\theta}_t + v_t$, $v_t \sim N(0, V_t)$ onde supusemos que $v_{1,t}$ e $v_{2,t}$ são independentes.

Se além disso fizermos $\mathbf{G}_t = \text{diag}(\mathbf{G}_{1,t}, \mathbf{G}_{2,t})$ temos

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

onde $\boldsymbol{\omega}_t = \begin{pmatrix} \boldsymbol{\omega}_{1,t} \\ \boldsymbol{\omega}_{2,t} \end{pmatrix}$ tem distribuição normal com média $\mathbf{0}$ e variância $\mathbf{W}_t = \text{diag}(\mathbf{W}_{1,t}, \mathbf{W}_{2,t})$.

O exemplo mais comum é a superposição de uma tendência polinomial com uma parte sazonal.

Nesse caso, supondo crescimento linear e modelagem trigonométrica com m harmônicos temos

$$\begin{aligned}\mathbf{F}'_{1,t} &= (1 \ 0), \quad \mathbf{F}'_{2,t} = \underbrace{(1 \ 0 \ 1 \ 0 \ \cdots \ 1 \ 0)}_{2m}, \\ \mathbf{F}'_t &= \underbrace{(1 \ 0 \ 1 \ 0 \ \cdots \ 1 \ 0)}_{2(m+1)}.\end{aligned}$$

$$\mathbf{G}_{1,t} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{G}_{2,t} = \text{diag}(\mathbf{G}^{(1)}, \dots, \mathbf{G}^{(m)}),$$

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

$$\text{com } \mathbf{G}^{(j)} = \begin{pmatrix} c_j & s_j \\ -s_j & c_j \end{pmatrix},$$

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

Teorema 13.1 *Para um inteiro $h \geq 2$, considere h séries temporais Y_{it} geradas pelos MLD $M_i : \{\mathbf{F}_i, \mathbf{G}_i, \mathbf{W}_i, \mathbf{V}_i\}_t$, com vetores de estado $\boldsymbol{\theta}_{it}$ de dimensão n_i , para $i = 1, 2, \dots, h$. Denote os erros de observação e evolução em M_i por v_{it} e $\boldsymbol{\omega}_{it}$, respectivamente. Assuma que, $\forall i \neq j, (1 \leq i, j \leq h)$, v_{it} e $\boldsymbol{\omega}_{it}$ são mutuamente independentes das séries v_{jt} e $\boldsymbol{\omega}_{jt}$. Então a série definida por :*

$$Y_t = \sum_{i=1}^h Y_{it} \quad (48)$$

segue um MLD $\{\mathbf{F}, \mathbf{G}, V, \mathbf{W}\}_t$ com vetor de estado dado por $\boldsymbol{\theta}'_t = (\boldsymbol{\theta}'_{1t}, \dots, \boldsymbol{\theta}'_{ht})$, de dimensão $n = n_1 + \dots + n_h$ e quádrupla definida por :

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

Demonstração : Temos que $Y_t = \mathbf{F}'_t \boldsymbol{\theta}_t + v_t$ onde $v_t = \sum_{i=1}^h v_{it}$. v_t é normalmente distribuído com média zero e as hipóteses de independência levam à variância V_t , como necessário. Para o vetor de estado $\boldsymbol{\theta}_t$ temos $\boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t$ onde $\boldsymbol{\omega}'_t = (\boldsymbol{\omega}'_{1,t}, \dots, \boldsymbol{\omega}'_{h,t})$. Novamente, pelas hipóteses usuais, $\boldsymbol{\omega}_t \sim N[\mathbf{0}, \mathbf{W}_t]$ que é independente de v_t , assim temos definido o MLD.

Fatores de Desconto

Já vimos no modelo polinomial de 1a. ordem, que a especificação da matriz \mathbf{W}_t de variância da perturbação do sistema $\boldsymbol{\omega}_t$ pode ser feita indiretamente através do uso de descontos.

Mais especificamente, se a equação do sistema é

$\boldsymbol{\theta}_t = \mathbf{G}\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t$, $\boldsymbol{\omega}_t \sim N(\mathbf{0}, \mathbf{W}_t)$ temos que, para $V(\boldsymbol{\theta}_{t-1}|D_{t-1}) = \mathbf{C}_{t-1}$, $\mathbf{R}_t = V(\boldsymbol{\theta}_t|D_{t-1}) = \mathbf{P}_t + \mathbf{W}_t$ onde $\mathbf{P}_t = V(\mathbf{G}_t\boldsymbol{\theta}_{t-1}|D_{t-1}) = \mathbf{G}_t\mathbf{C}_{t-1}\mathbf{G}'_t$. Daí, $\mathbf{W}_t = \mathbf{R}_t - \mathbf{P}_t$.

Se definimos δ de tal forma que $\mathbf{R}_t = \mathbf{P}_t/\delta$ podemos interpretar δ como a percentagem de informação que passa de $t-1$ para t e nesse caso

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

No caso de superposição de modelos (ex.: tendência + sazonalidade) podemos estender o raciocínio acima.

No caso geral de k modelos, poderíamos definir

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

e utilizar um desconto δ_i , $i = 1, 2, \dots, k$ para cada bloco de forma que $\mathbf{R}_{i,t} = \mathbf{P}_{i,t}/\delta_i$ e $\mathbf{W}_i = \mathbf{P}_{i,t}(\delta_i^{-1} - 1)$.

O modelo total obtido pela superposição dos k componentes teria então $\mathbf{W}_t = \text{diag}(\mathbf{W}_{1,t}, \dots, \mathbf{W}_{k,t})$, onde $\mathbf{W}_{i,t}$ seria dado como acima.

Estratégias práticas de desconto

Desconto é uma técnica projetada para evolução 1-passo-a-frente.

Usa raciocínio multiplicativo para obter fator aditivo. Se ela for usada k vezes teremos decaimento exponencial da informação, que é inconsistente com o decaimento aritmético especificado pelo modelo.

Para passar de $\mathbf{C}_t \rightarrow \mathbf{R}_{t+1}$ usamos desconto δ levando a \mathbf{W}_{t+1} .

Observando y_{t+1} , passamos de $\mathbf{C}_{t+1} \rightarrow \mathbf{R}_{t+2}$ usando desconto δ que leva a $\mathbf{W}_{t+2} \neq \mathbf{W}_{t+1}$.

Por outro lado se queremos prever y_{t+2} precisamos de $V(\omega_{t+2}|D_t)$ que tomamos também como \mathbf{W}_{t+1} .

Raciocínio similar vale para k passos a frente.

As variâncias \mathbf{W}_t usadas na filtragem são as especificadas 1-passo-a-frente.

14 Modelos com variâncias desconhecidas

Até agora, foi assumido em todos os modelos que a variância das observações, dada por V_t , era conhecida para todos os tempos.

Na prática, isso raramente acontece e o que deve ser feito é estimar a variância a cada tempo.

Seguindo a mesma estrutura do método, a variância será estimada sequencialmente e uma forma de evolução será obtida para relacioná-la em tempos sucessivos. Na sequência trabalharemos com $V_t = V_{t-1} = V = \phi^{-1}$, $\forall t$, a precisão observacional.

Antes de tratar do caso geral, analisaremos modelos de 1a. ordem.

Modelo de 1a. Ordem

Esse modelo será ligeiramente modificado para:

$$\text{Eq. das observações: } y_t = \mu_t + v_t, v_t \sim N(0, V) \quad (49)$$

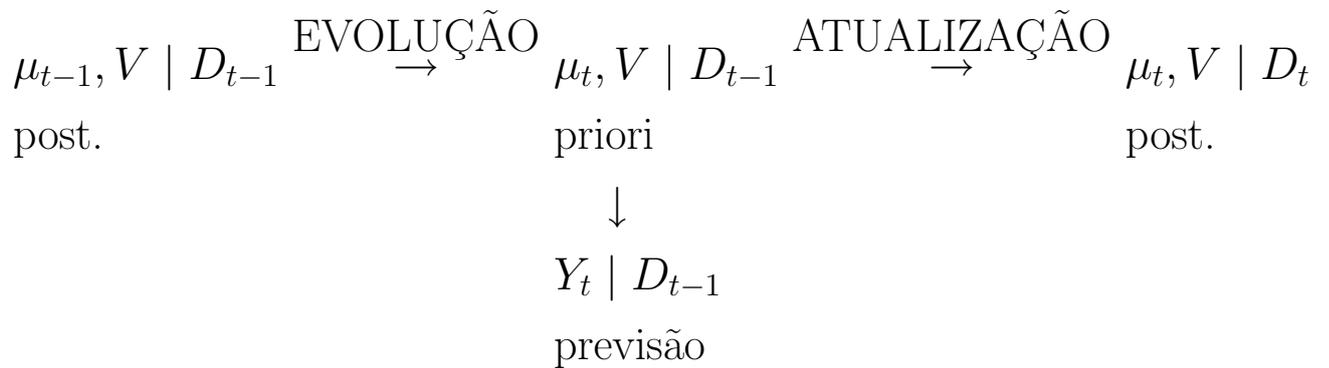
$$\text{Eq. do sistema: } \mu_t = \mu_{t-1} + \omega_t, \omega_t \sim N(0, VW_t^*)$$

$$\text{Info. Inicial: } \mu_0 | D_0, V \sim N(m_0, VC_0^*) \text{ e } V \sim GI\left(\frac{n_0}{2}, \frac{d_0}{2}\right)$$

A inclusão do termo V na variância de ω_t serve para dar tratabilidade analítica.

Continuam valendo as independências entre erros v_t e ω_t mas condicionadas a V .

O processo inferencial consiste novamente em



Em todas as etapas acima, ao invés de trabalharmos com a distribuição conjunta de (μ, V) trabalharemos com a condicional de $\mu|V$ e a marginal de V . Observe que $E[V^{-1}|D_0] = \frac{n_0/2}{d_0/2} = \frac{n_0}{d_0} = S_0^{-1}$ onde S_0 seria uma estimativa inicial de V . O valor de n_0 fornece a precisão dessa estimativa, pois $CV[V^{-1}|D_0] = \frac{1}{\sqrt{n_0}}\sqrt{2}$.

Teorema 14.1 *Com o modelo especificado como em (49), obtemos as seguintes distribuições para cada tempo $t \geq 1$.*

(a) *Condicional a V :*

$$\begin{aligned}(\mu_{t-1} \mid D_{t-1}, V) &\sim N[m_{t-1}, VC_{t-1}^*], \\(\mu_t \mid D_{t-1}, V) &\sim N[a_t, VR_t^*], \\(Y_t \mid D_{t-1}, V) &\sim N[f_t, VQ_t^*], \\(\mu_t \mid D_t, V) &\sim N[m_t, VC_t^*],\end{aligned}$$

com

$$\begin{aligned}a_t &= m_{t-1}, & R_t^* &= C_{t-1}^* + W_t^* \\f_t &= a_t, & Q_t^* &= 1 + R_t^* \\m_t &= a_t + A_t e_t, & C_t^* &= R_t^* - A_t^2 Q_t^* \\A_t &= R_t^* Q_t^{*-1}, & e_t &= Y_t - f_t\end{aligned}$$

(b) *Para a precisão $\phi = V^{-1}$:*

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

onde:

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

(c) *Incondicional a V :*

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

onde :

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

Podemos obter também as seguintes relações:

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

Demonstração : Dada a definição do modelo, os resultados em (a) seguem diretamente do teorema 6.1. Eles são apenas os resultados com a variância V conhecida.

O restante da demonstração é por indução, e usa os resultados da distribuição normal-gama.

Assuma que a priori para a precisão $\phi (= V^{-1})$ em (b) é verdadeira, de modo que $(\phi | D_{t-1}) \sim G[n_{t-1}/2, d_{t-1}/2]$. De (a) temos que :

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

tal que

$$p(Y_t | D_{t-1}, \phi) \propto \phi^{1/2} \exp(-0,5\phi e_t^2/Q_t^*).$$

Pelo Teorema de Bayes, temos a posteriori de ϕ dada por

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

Usando a priori de (b) e a verossimilhança acima temos

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

claramente esta é a função de densidade da Gama com parâmetros $n_t/2$ e $d_t/2$, onde $n_t = n_{t-1} + 1$ e $d_t = d_{t-1} + e_t^2 Q_t^{*-1}$. Isto estabelece a posteriori para ϕ em (b).

Para demonstrar (c) basta lembrar que se μ é uma variável aleatória então

$$\mu | \phi \sim N[m, \phi^{-1}C^*] \text{ e } \phi \sim G[n/2, d/2] \Rightarrow \mu \sim t_n[m, SC^*]$$

onde $S = d/n$. Os resultados em (c) seguem da marginalização das distribuições em (a) com respeito à distribuição gama apropriada a priori/posteriori para ϕ .

Os resultados para o modelo geral podem ser obtidos da seguinte forma. Para todo t , o modelo é definido por :

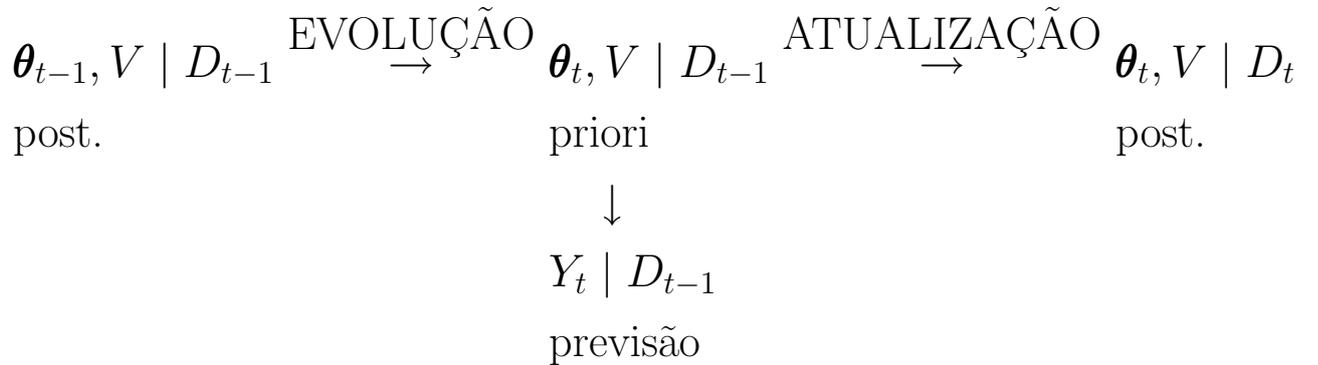
$$Y_t = \mathbf{F}'_t \boldsymbol{\theta}_t + \nu_t \quad , \quad \nu_t \sim N[0, V] \quad (50)$$

$$\boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t \quad , \quad \boldsymbol{\omega}_t \sim N[0, V \mathbf{W}_t^*]$$

$$(\boldsymbol{\theta}_0 | D_0, V) \sim N[\mathbf{m}_0, V \mathbf{C}_0^*] \text{ e } (V | D_0) \sim GI[n_0/2, d_0/2].$$

Novamente, estamos supondo que $V_t = V_{t-1} = V, \forall t$.

As hipóteses usuais de independência permanecem, porém agora condicionais a V , ou equivalentemente, $\phi = V^{-1}$. A priori de ϕ tem média $E[\phi | D_t] = n_0/d_0 = 1/S_0$ onde S_0 é uma estimativa a priori da variância observacional V . O processo inferencial consiste novamente em



Teorema 14.2 *Com o modelo especificado como em (50), obtemos as seguintes distribuições para cada tempo $t \geq 1$.*

(a) *Condicional a V :*

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

com

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

(b) *Para a precisão $\phi = V^{-1}$:*

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

onde :

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

(c) *Incondicional a V :*

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

onde :

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

Também podemos obter aqui as relações:

$$\begin{aligned}\mathbf{m}_t &= \mathbf{a}_t + \mathbf{A}_t e_t & \mathbf{C}_t &= S_t/S_{t-1}[\mathbf{R}_t - \mathbf{A}_t\mathbf{A}_t'Q_t] \\ n_t &= n_{t-1} + 1 & d_t &= d_{t-1} + S_{t-1}e_t^2Q_t^{-1} \\ Q_t &= S_{t-1} + \mathbf{F}_t'\mathbf{R}_t\mathbf{F}_t & \mathbf{A}_t &= \mathbf{R}_t\mathbf{F}_tQ_t^{-1}\end{aligned}$$

Demonstração : Generalização direta do modelo de 1a. ordem.

Para a parte (a), basta usar os resultados do teorema 7.1.

Para a parte (b), usam-se os mesmos resultados que no modelo de 1a. ordem.

Para a parte (c), basta usar os resultados para a distribuição normal-gama no caso multivariado.

Isto é, se $\boldsymbol{\theta}$ é um vetor aleatório ($n \times 1$) então

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

onde $S = d/n$.

15 Análise retrospectiva e Otimalidade Linear

Análise Retrospectiva ou Suavização

Até agora vimos como obter as distribuições de $\boldsymbol{\theta}_t|D_{t-1}$ (priori) e $\boldsymbol{\theta}_t|D_t$ (posteriori ou atualizada).

Se estamos usando o sistema no tempo real, a distribuição atualizada é o melhor de que dispomos.

No entanto, muitas vezes dispomos de dados que vão além do tempo t . Gostaríamos então de poder dispor dessa informação para obter uma distribuição para $\boldsymbol{\theta}_t$ mais precisa, ou seja, gostaríamos de obter $\boldsymbol{\theta}_t|D_n$, $n > t$. Essa distribuição é inútil para previsão mas muito útil para análises retrospectivas que permitem um melhor controle do sistema.

Teorema 15.1 *Seja um MLD $\{\mathbf{F}_t, \mathbf{G}_t, V_t, \mathbf{W}_t\}$. Para cada tempo t e para todo k ($1 \leq k < t$) temos que*

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

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

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

com $\mathbf{B}_t = \mathbf{C}_t\mathbf{G}'_{t+1}\mathbf{R}_{t+1}^{-1}$, e valores iniciais

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

Demonstração: (por indução)

$$\begin{aligned} p(\boldsymbol{\theta}_{t-k}|D_t) &= \int p(\boldsymbol{\theta}_{t-k}, \boldsymbol{\theta}_{t-k+1}|D_t) d\boldsymbol{\theta}_{t-k+1} \\ &= \int p(\boldsymbol{\theta}_{t-k+1}|D_t) p(\boldsymbol{\theta}_{t-k}|\boldsymbol{\theta}_{t-k+1}, D_t) d\boldsymbol{\theta}_{t-k+1} \end{aligned}$$

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

Para obter $p(\boldsymbol{\theta}_{t-k}|\boldsymbol{\theta}_{t-k+1}, D_{t-k})$ obtenha antes

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

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

$$\mathbf{h}_t(k) = \mathbf{m}_{t-k} + \mathbf{B}_{t-k}(\boldsymbol{\theta}_{t-k+1} - \mathbf{a}_{t-k+1})$$

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

A dependência de $\boldsymbol{\theta}_{t-k}$ em $\boldsymbol{\theta}_{t-k+1}$ é linear na média e

$\boldsymbol{\theta}_{t-k+1}|D_t \sim N(\mathbf{a}_t(-k+1), \mathbf{R}_t(-k+1))$, por hipótese

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

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

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

Corolário 15.1 *Se as variâncias observacionais forem desconhecidas e $V_t = V = \phi^{-1}$, $\forall t$, temos que*

$$\boldsymbol{\theta}_{t-k}|\phi, D_t \sim N(\mathbf{a}_t(-k), \phi^{-1}\mathbf{R}_t(-k)) \text{ e}$$

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

Prova: Explora os mesmos resultados a respeito da normal-gama.

Otimidade Linear Bayesiana

Suponha que \mathbf{y} é um vetor de observações e

$\boldsymbol{\theta}$ é um vetor de parâmetros.

Queremos estimar $\boldsymbol{\theta}$ por \mathbf{d} de forma que a perda $L(\boldsymbol{\theta}, \mathbf{d})$ esperada seja mínima. Obviamente se \mathbf{y} estiver relacionado $\boldsymbol{\theta}$, usaremos essa informação em \mathbf{d} .

O problema se resume em obter \mathbf{d} de forma a **minimizar**

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

Considere agora que a única informação disponível de \mathbf{y} e $\boldsymbol{\theta}$ são os dois primeiros momentos na forma

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

Sem informação adicional sobre a distribuição, não podemos calcular $E[\boldsymbol{\theta}|\mathbf{y}]$. Para contornar esse problema, trabalhamos com a perda esperada geral $E[L(\boldsymbol{\theta}, \mathbf{d})]$ e restringimos os estimadores à classe dos estimadores lineares da forma

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

Define-se o estimador linear Bayesiano (ELB) como o estimador que minimiza a perda esperada dentre os estimadores lineares.

Pode-se mostrar que ele é dado por

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

e a perda associada é dada por

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

\mathbf{m} e \mathbf{C} são então usados como primeira aproximação para os momentos de $\boldsymbol{\theta}|\mathbf{y}$, chamados de esperança linear a posteriori e variância linear a posteriori de $\boldsymbol{\theta}|\mathbf{y}$. Observe que eles coincidem com média e variância a posteriori se a distribuição conjunta de $(\mathbf{y}, \boldsymbol{\theta})$ for normal.

Importância de otimalidade linear para MLD

As distribuições de interesse em modelos lineares dinâmicos podem ser obtidas da aplicação desse método sem necessidade da hipótese de normalidade. Considere novamente o MLD

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

parcialmente especificado através de médias e variâncias. Então continua valendo o teorema 7.1 sem as hipóteses de normalidade.

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

onde as expressões dos momentos permanecem as mesmas.

Os dois primeiros resultados seguem da linearidade da esperança.

O último segue de otimalidade linear.

16 Aplicação

Uso do software PRVWIN.

Exercícios com a série do produto da indústria geral do Brasil.

1. Análise da série com fator de desconto para os blocos da tendência e sazonalidade;
2. Verificação do efeito da escolha dos fatores de desconto;
3. Análise retrospectiva do modelo de 1a. ordem e do modelo de tendência e sazonalidade.

17 Intervenção

Sempre que o comportamento sai do padrão, devemos intervir (management by exception).

Intervenção pode ser $\left\{ \begin{array}{l} \text{antecipatória (feed forward)} \\ \text{retrospectiva ou corretiva (feed back)} \end{array} \right.$

Notação: mesma de antes

Consideraremos variância conhecida para fixar idéias.

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

Intervenção Antecipatória

1. Ignorar observação y_t

Se y_t é julgado não compatível com a série \Rightarrow não-informativo \Rightarrow pode ser ignorado.

Logo, $I_t = \{y_t \text{ ausente}\} \Rightarrow D_t = D_{t-1} \Rightarrow \mathbf{m}_t = \mathbf{a}_t$ e $\mathbf{C}_t = \mathbf{R}_t$.

Isso pode ser formalizado no MLD fazendo $V_t \rightarrow \infty$.

2. Ruído de evolução adicional

Se sabemos que algo aconteceu, acrescentamos ruído ao sistema. Em geral, fazemos \mathbf{R}_t maior para alguns ou todos componentes de $\boldsymbol{\theta}_t$. Assim tomamos $I_t = \{\boldsymbol{\eta}_t\}$ onde o ruído adicional $\boldsymbol{\eta}_t \sim N[\mathbf{h}_t, \mathbf{H}_t]$ é independente de $\boldsymbol{\omega}_t$.

Formalmente,

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

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

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

Se alguma componente de \mathbf{h}_t e $\text{diag}\mathbf{H}_t$ são zero não esperamos mudança no correspondente $\theta_{t,j}$.

Assim como \mathbf{W}_t , a função de \mathbf{H}_t é aumentar a incerteza. Logo, podemos especificá-la através do uso de descontos. No caso, usamos desconto menor que o usual para permitir aumento maior da incerteza.

3. Intervenção subjetiva arbitrária

Modelo fornece $\forall t: \mathbf{a}_t, \mathbf{R}_t$. Suponha que $I_t = \{\mathbf{a}_t^*, \mathbf{R}_t^*\}$, \mathbf{a}_t^* e \mathbf{R}_t^* quaisquer. Logo $\boldsymbol{\theta}_t | I_t, D_{t-1} \sim N(\mathbf{a}_t^*, \mathbf{R}_t^*)$.

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

Esse modelo não é consistente com o MLD e não podemos fazer filtragem. Para fazer compatibilização usamos os resultados abaixo.

Lema: Sejam $E(\boldsymbol{\theta}_t) = \mathbf{a}_t$ e $V(\boldsymbol{\theta}_t) = \mathbf{R}_t$, \mathbf{R}_t $n \times n$ não-singular, triangular superior e $\boldsymbol{\theta}_t^* = \mathbf{K}_t \boldsymbol{\theta}_t + \mathbf{h}_t$. Então $E(\boldsymbol{\theta}_t^*) = \mathbf{a}_t^*$ e $V(\boldsymbol{\theta}_t^*) = \mathbf{R}_t^*$ se $\mathbf{K}_t = \mathbf{U}_t \mathbf{Z}_t^{-1}$ e

$\mathbf{h}_t = \mathbf{a}_t^* - \mathbf{K}_t \mathbf{a}_t$, onde $\mathbf{U}_t(\mathbf{Z}_t)$ é a única matriz não-singular triangular superior de \mathbf{R}_t^* (\mathbf{R}_t), isto é, $\mathbf{R}_t^* = \mathbf{U}_t \mathbf{U}_t'$ ($\mathbf{R}_t = \mathbf{Z}_t \mathbf{Z}_t'$).

Prova: Da álgebra linear, se \mathbf{R}_t e \mathbf{R}_t^* são simétricas p.d. $\Rightarrow \mathbf{U}_t$ e \mathbf{Z}_t existem e são únicas.

Da definição de \mathbf{h}_t ,

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

Assim do Lema obtemos uma "segunda evolução" de $\boldsymbol{\theta}_t$ para $\boldsymbol{\theta}_t^*$ com os momentos requeridos. Logo

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

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

Modelo está formalmente dentro da estrutura de MLD.

4. Inclusão de efeitos de intervenção

Até agora, as intervenções mantiveram os mesmos parâmetros.

Às vezes, queremos intervir através de parâmetros adicionais de forma a isolá-los e estimá-los.

Ex.: Considere o MCL com $\boldsymbol{\theta}_t = (\mu_t, \beta_t)$ fazemos

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

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

Note que o vetor paramétrico mudou de dimensão.

Define-se assim, uma "segunda evolução" para os parâmetros do sistema.

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

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

A evolução (total) fica dada por

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

onde $\mathbf{G}_t^* = \mathbf{K}_t \mathbf{G}_t$, $\boldsymbol{\omega}_t^* = \boldsymbol{\xi}_t + \mathbf{K}_t \boldsymbol{\omega}_t \sim N(E(\boldsymbol{\xi}_t), \mathbf{W}_t^*)$ com

$$\mathbf{W}_t^* = \mathbf{K}_t \mathbf{W}_t \mathbf{K}_t' + V(\boldsymbol{\xi}_t).$$

18 Monitoração

Atenção agora está voltada para monitorar performance do modelo (previsões ruins, mudança nos parâmetros).

Técnicas ligadas a alguma medida do erro de previsão $e_t \Rightarrow$ Fator de Bayes (FB).

Considere modelos M_0 (padrão), M_1, M_2, \dots com previsões $p(y_t|D_{t-1}, M_i) = p_i(y_t|D_{t-1}), i = 0, 1, 2, \dots$.

O FB para M_0 versus M_1 baseado em y_t é $H_t = \frac{p_0(y_t|D_{t-1})}{p_1(y_t|D_{t-1})}$.

O FB mede a razão das verossimilhanças preditivas: quanto maior (menor) for H_t , maior (menor) é a evidência a favor de M_0 (versus M_1).

O FB para M_0 versus M_1 baseado nas últimas k observações consecutivas $y_t, y_{t-1}, \dots, y_{t-k+1}$ é

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

Observações:

- (i) $H_t(1) = H_t$;
- (ii) $H_t^{-1}(k)$ é o FB para M_1 versus M_0 baseado em y_t, \dots, y_{t-k+1} ;
- (iii) $H_t(t)$ é o FB baseado em todos os dados até y_t ;
- (iv) $H_t(k) = H_t H_{t-1}(k-1)$ ou

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

Como o modelo é dinâmico trabalhamos mais com $H_t(k)$ para k pequeno.

$H_t = H_t(1)$ pequeno $\Rightarrow y_t$ é possível *outlier*.

$H_t(k)$ pequeno \Rightarrow últimos k pontos indicam mudanças estruturais não captadas por M_0 .

Na prática, devemos nos concentrar no pior valor de $H_t(k)$, $1 \leq k \leq t$. O teorema abaixo indica como atualizar este valor.

Teorema 18.1 $L_t = \min_k H_t(k) \Rightarrow$

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

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

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

Idéia é usar L_t e l_t para detecção.

Se L_t é muito pequeno, intervimos. Em geral, toma-se um limiar τ a partir do qual intervimos, por exemplo, $\tau \approx 0,2$. Antes disso, mesmo que $L_t < 1$ não mexemos no modelo. Se $L_t > 1$, não mexemos também.

Suponha $\tau < L_{t-1} < 1$. $L_t < L_{t-1}$ se $H_t < 1$.

Se $L_t < \tau$, monitor sinaliza.

Se $L_t < \tau$ e

(i) $l_t = 1 \Rightarrow L_t = H_t \Rightarrow y_t$ é *outlier* ou modelo começa a deteriorar em t ;

(ii) $l_t > 1 \Rightarrow$ modelo começou a mudar no tempo $t - l_t + 1$.

Escolha de modelos alternativos M_1

Suponha que $f_t = 0$ e $Q_t = 1$ ou alternativamente trabalhe com

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

Se a variância for desconhecida, basta trocar a N pela $t_{n_{t-1}}$.

Alternativas possíveis para M_1 são:

- (i) $e_t \sim N(h, 1)$ deslocamento no nível;
- (ii) $e_t \sim N(0, k^2)$ deslocamento na escala.

Sob (i), $H_t = \exp\{(h^2 - 2he_t)/2\}$ e a escolha de h é baseada no FB.

$$H_t = 1 \text{ (indiferença entre } M_0 \text{ e } M_1) \Leftrightarrow h = 2e_t.$$

Achamos isso razoável para um $e_t = 1,5 \Rightarrow h \approx 3$.

Supondo que o limiar é $\tau = e^{-2} \Leftrightarrow h^2 - 2he_t - 2 \log \tau = 0$. Queremos esse limiar atingido quando $e_t \approx 2,5 \Rightarrow h = 1$ ou 4.

h entre 3 e 4 parece razoável: FB indiferente quando $e_t = 1,5$ mas leva à rejeição de M_0 quando $e_t = 2,5$.

Sob (ii), $H_t = k \exp\{-e_t^2(1 - k^{-2})/2\}$. Nota-se que valor de k pouco relevante para $|e_t|$ grande.

$k = 3$ ou 4 e $\tau = 0,15$ parece razoável: leva à rejeição de M_0 quando $|e_t| = 2,5$.

Intervenção Retrospectiva

Detecção e diagnóstico automático.

Esquema funciona assim:

- (A) Cálculo de H_t . Se $H_t > \tau$, vai para (B). Se $H_t < \tau$, y_t é *outlier* e omitido, ou y_t anuncia início de mudança, vai para (C).
- (B) Cálculo de L_t e l_t . Se $L_t > \tau$ ou $l_t > 4$, vai para (D). Se $L_t < \tau$, intervem. Se $L_t < 1$ e $l_t \leq 4$, intervem. Vai para (C).
- (C) Monitor sinaliza. Intervem e resseta monitor para $l_t = 0$ e $L_t = 1$.
- (D) Faz análise usual.

Adaptação automática em caso de mudança paramétrica.

Forma mais simples de intervenção é através do aumento da incerteza a priori.

19 Modelo de Regressão Dinâmica, Função de Transferência e Testes

Modelos de Regressão Dinâmica

Regressão :

y_t - série temporal resposta

x_t - série temporal de regressores que influenciam y_t .

É também possível que x_s , $s < t$ ou mesmo y_s , $s < t$ influenciem y_t .

Modelo simples : $\mu_t = \alpha + \beta x_t$

Suponha que existe f tal que $\mu_t = f(x_t, t)$. Em geral, se existe é desconhecida.

Melhor aproximação é tomar $f(x_t, t) = \alpha_t + \beta_t x_t$, isto é, função linear de x_t mas coeficientes mudam.

É sempre possível (escolhendo bem α_t e β_t), aproximar f por $\alpha_t + \beta_t x_t$.

Modelo tem forma local para representar mudanças em α e β . Forma simples de relacionar α 's e β 's é

$$E \left[\begin{pmatrix} \alpha_t \\ \beta_t \end{pmatrix} \mid \begin{pmatrix} \alpha_{t-1} \\ \beta_{t-1} \end{pmatrix} \right] = \begin{pmatrix} \alpha_{t-1} \\ \beta_{t-1} \end{pmatrix}$$

Variação dos parâmetros :

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

O MLD de regressão múltipla

As idéias podem ser extendidas para o caso de **n regressores** x_1, x_2, \dots, x_n onde $x_1 = 1$.

$$\text{Eq. Obs.: } y_t = \alpha_t + \beta_{t,1}x_{t,2} + \dots + \beta_{t,n-1}x_{tn} + v_t \quad v_t \sim N(0, V_t)$$

$$\text{Eq. Sist.: } \alpha_t = \alpha_{t-1} + \omega_{t,1}$$

$$\beta_{t,1} = \beta_{t-1,1} + \omega_{t,2} \quad \omega_{t,i} \sim N(0, W_{t,i})$$

⋮

$$\beta_{t,n-1} = \beta_{t-1,n-1} + \omega_{t,n}$$

ou colocando na forma matricial

$$\text{Eq. Obs.: } y_t = \mathbf{F}'_t \boldsymbol{\theta}_t + v_t \quad v_t \sim N(0, V_t)$$

$$\text{Eq. Sist.: } \boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t \quad \boldsymbol{\omega}_t \sim N(0, \mathbf{W}_t)$$

onde $\mathbf{F}'_t = (1, x_{t,2}, \dots, x_{tn})$, $\boldsymbol{\theta}'_t = (\alpha_t, \beta_{t,1}, \dots, \beta_{t,n-1})'$

$$\boldsymbol{\omega}'_t = (\omega_{t,1}, \dots, \omega_{t,n})'$$

Se $\mathbf{W}_t = 0 \Rightarrow \boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1} = \boldsymbol{\theta}$, $\forall t \Rightarrow$ Modelo de regressão estática (usual).

Se $\mathbf{W}_t \neq 0$, $\boldsymbol{\theta}_t$ muda com o tempo.

\mathbf{W}_t "grande" (pequeno), muita (pouca) mudança de $\boldsymbol{\theta}_t$ com o tempo.

Se temos apenas 1 regressor, $\mathbf{F}'_t = (1, x_t)$, $\boldsymbol{\theta}'_t = (\alpha_t, \beta_t)'$

Para fazer previsão é preciso conhecer os valores futuros dos regressores. Uma forma possível é modelar conjuntamente y_{t+k} com $\mathbf{x}_{t+k} \Rightarrow$ modelo multivariado. Outra forma é obter $p(\mathbf{x}_{t+k}|D_t)$ e prever usando

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

Se tivermos apenas que $E[\mathbf{x}_{t+k}|D_t] = \mathbf{h}_t(k)$ e $V[\mathbf{x}_{t+k}|D_t] = \mathbf{H}_t(k)$ então usando o fato que $y_{t+k}|\mathbf{x}_{t+k}, \boldsymbol{\beta}_{t+k}, D_t \sim N(\mathbf{x}'_{t+k}\boldsymbol{\beta}_{t+k}, V_t)$

$$\begin{aligned} E[y_{t+k}|D_t] &= E[E[y_{t+k}|\mathbf{x}_{t+k}, \boldsymbol{\beta}_{t+k}, D_t]] \\ &= E[\mathbf{x}'_{t+k}\boldsymbol{\beta}_{t+k}|D_t] \\ &= E[E[\mathbf{x}'_{t+k}\boldsymbol{\beta}_{t+k}|\mathbf{x}_{t+k}, D_t]] \\ &= E[\mathbf{x}'_{t+k}E[\boldsymbol{\beta}_{t+k}|\mathbf{x}_{t+k}, D_t]] \\ &= \mathbf{h}_t(k)\mathbf{a}_t(k) \quad \text{e} \end{aligned}$$

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

No modelo de regressão dinâmica, $\mathbf{a}_t(k) = \mathbf{m}_t$ e $\mathbf{R}_t(k) = \mathbf{C}_t + \sum_{i=1}^k \mathbf{W}_{t+i}$.

Funções de Transferência

Podemos também ter os valores da série observada dependente em valores presentes e passados de uma série auxiliar.

Nesse caso, $y_t = \theta_0 + \theta_1 x_t + \theta_2 x_{t-1} + \dots + \theta_{p+1} x_{t-p} + v_t$.

O modelo pode ser colocado na estrutura dinâmica fazendo os parâmetros θ dependentes no tempo. A equação do sistema teria usualmente a forma de um passeio aleatório

$$\theta_t = \theta_{t-1} + \omega_t.$$

O modelo acima tem y_t dependendo apenas de x_t até p períodos de tempos atrás. Podemos também modelar um efeito que decai suavemente para 0 através de variáveis construídas a partir da variável independente x .

Exemplo: $y_t = \theta_{0,t} + \xi_t + v_t$ onde $\xi_t = \lambda \xi_{t-1} + \psi x_t$.

λ representa a memória do efeito

ψ representa a força ou penetração de x .

Pode-se mostrar que $\xi_{t+k} = \lambda^k \xi_t + \psi \sum_{i=1}^k \lambda^{k-i} x_{t+i}$

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

No modelo de regressão usual, $y_t = \alpha + \theta x_t + e_t$, a função de transferência de x_t em y_{t+r} é

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

Em modelos autoregressivos de ordem p temos $\begin{cases} \theta_{r+1} x, & r = 0, 1, \dots, p \\ 0, & \text{c.c.} \end{cases}$

Definição: Um modelo geral de função de transferência é dado por

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

Esse modelo pode ser colocado na forma de um MLD usual. Basta tomar o vetor de estados $\tilde{\boldsymbol{\theta}}'_t = (\boldsymbol{\theta}'_t, \boldsymbol{\psi}'_t)$ e definir o modelo pela quádrupla

$$\tilde{\mathbf{F}}'_t = (\mathbf{F}'_t, \mathbf{0}'), \quad \tilde{\mathbf{G}}_t = \begin{pmatrix} \mathbf{G} & x_t \mathbf{I}_n \\ \mathbf{0} & \mathbf{I}_n \end{pmatrix}, \quad V \text{ e } \tilde{\mathbf{W}}_t = \begin{pmatrix} \mathbf{W}_{t,1} + x_t^2 \mathbf{W}_{t,2} & x_t \mathbf{W}_{t,2} \\ x_t \mathbf{W}_{t,2} & \mathbf{W}_{t,2} \end{pmatrix}$$

A função de previsão é dada por

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

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

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

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

No caso do exemplo, $\mathbf{G} = \lambda, \boldsymbol{\psi}_t = \psi, \mathbf{F} = 1$ e a função de transferência é $\lambda^k \psi x$.

Testes de significância dos parâmetros do modelo

Em alguns modelos é interessante verificar a significância de um subconjunto de parâmetros, $\boldsymbol{\theta}_{t,1}$, para a explicação da série de interesse y_t .

Exemplo: $\boldsymbol{\theta}_{t,1} = (\beta_{t,j}, \gamma_{t,j})$, os parâmetros do j -ésimo harmônico da componente sazonal

Considere a região $R = \{\boldsymbol{\theta} | p(\boldsymbol{\theta} | D_t) > p(\mathbf{0} | D_t)\}$, onde $\boldsymbol{\theta}$ pode ser $\boldsymbol{\theta}_t$, $\theta_{t,j}$ ou $\boldsymbol{\theta}_{t,1}$ um subvetor de componentes de $\boldsymbol{\theta}_t$.

Ex.: Se V é conhecido, $\boldsymbol{\theta}_t | D_t \sim N(\mathbf{m}_t, \mathbf{C}_t)$,

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

- (i) $\boldsymbol{\theta} = \theta_{t,j}$ então $\mathbf{m} = m_{t,j}$ e $\mathbf{C} = C_{t,jj}$;
- (ii) $\boldsymbol{\theta} = \boldsymbol{\theta}_{t,1}$ então $\mathbf{m} = \mathbf{m}_{t,1}$ e $\mathbf{C} = \mathbf{C}_{t,1}$, onde $\mathbf{m}_{t,1}$ e $\mathbf{C}_{t,1}$ são as respectivas componentes do vetor \mathbf{m}_t e da matriz \mathbf{C}_t .

Observe que se

- (i) $P(R | D_t)$ for alta \Rightarrow o ponto $\mathbf{0}$ está na cauda da distribuição $\Leftrightarrow \mathbf{0}$ é pouco provável $\Leftrightarrow \boldsymbol{\theta}$ é significativa;
- (ii) $P(R | D_t)$ for baixa $\Rightarrow \mathbf{0}$ não está na cauda da distribuição $\Leftrightarrow \mathbf{0}$ é muito provável $\Leftrightarrow \boldsymbol{\theta}$ não é significativa;

Cálculo de $P(R|D_t)$

1. V conhecido

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

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

Mas $p(\boldsymbol{\theta}|D_t) \geq p(\mathbf{0}|D_t) \Leftrightarrow Q(\boldsymbol{\theta}) \leq Q(\mathbf{0}) = \mathbf{m}' \mathbf{C}^{-1} \mathbf{m}$.

Por outro lado, $Q(\boldsymbol{\theta})|D_t \sim \chi_q^2$ (página 12)

Logo,

$$P(R|D_t) = P(Q(\boldsymbol{\theta}) \leq Q(\mathbf{0})|D_t) = P(\chi_q^2 < \mathbf{m}' \mathbf{C}^{-1} \mathbf{m}),$$

que pode facilmente ser calculado.

2. V desconhecido

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

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

Mas $p(\boldsymbol{\theta}|D_t) \geq p(\mathbf{0}|D_t) \Leftrightarrow Q(\boldsymbol{\theta}) \leq Q(\mathbf{0}) = \mathbf{m}' \mathbf{C}^{-1} \mathbf{m}$.

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

Logo,

$$P(R|D_t) = P(Q(\boldsymbol{\theta}) \leq Q(\mathbf{0})|D_t) = P(F_{q, n_t} < \mathbf{m}' \mathbf{C}^{-1} \mathbf{m}/q),$$

que também pode facilmente ser calculado.

20 Aplicação

Uso do software PRVWIN.

Exercícios com a série do produto da indústria geral do Brasil.

1. Análise da série fazendo monitoração;
2. Análise da série fazendo intervenção;
3. Análise de uma série com função de transferência.

21 Transformações nos dados e séries não normais

Outlier e dados ausentes

Um outlier pode ser identificado como uma observação inesperada \Rightarrow erro de previsão grande. Sabemos que

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

Devemos olhar comportamento de $|e_t|$ ou $|e_t|/\sqrt{Q_t}$.

Observe que e_t influencia estimativas de $\boldsymbol{\theta}_t$ e V através de

$$\mathbf{m}_t = \mathbf{a}_t + \mathbf{A}_t e_t \text{ e } S_t = S_{t-1}(n_{t-1} + e_t^2/Q_t)/(n_{t-1} + 1).$$

e_t grande $\Rightarrow y_t$ não é descrito pelo mesmo modelo

$$\Rightarrow (\boldsymbol{\theta}_t, V | D_t) \sim (\boldsymbol{\theta}_t, V | D_{t-1}).$$

Obviamente, se y_t é perdido, resultado é o mesmo.

Intervalos de tempo irregulares

Até agora, consideramos intervalos de tempo iguais a 1.

Ex.: observamos apenas 1 observação no final do mês mas não sabemos quando \Rightarrow observamos $y_{31}, y_{60}, y_{91}, \dots$

Basta repetir procedimento de dados faltando para os outros tempos. Uso de desconto nesse caso deve seguir recomendações anteriores.

Transformações nos dados

Muito comum em estatística, deve ser usada com cuidado para não complicar interpretação.

Uma idéia é transformar y_t não normal em $z_t = g(y_t)$ aproximadamente normal.

Se $E(y_t) = \mu$ então aproximações de Taylor dão

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

analogamente, se $V(z_t) = V$ então $V(z_t) \approx \{g'(\mu_t)\}^2 V_t$.

Família comum de transformações,

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

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

Alguns casos particulares dessa transformação são:

1. $\lambda = 0$ (transformação log) $\Rightarrow V(z_t) \propto E(z_t)^2$;
2. $\lambda = 0,5$ (transf. $\sqrt{\quad}$) $\Rightarrow V(z_t) \propto E(z_t)$ (modelo Poisson).

Obviamente, se escolhermos transformações e $z_t \sim t_{n_t}(f_t, Q_t)$, a previsão de y_t é feita com distribuição da transformação inversa. É importante observar que $E(z_t) \neq g[E(y_t)]$, mas mediana $z_t = g(\text{mediana } y_t)$.

Se $y_t = \gamma_t x_t^{\beta_t} v_t$ então $z_t = \log y_t = \log \gamma_t + \beta_t \log x_t + \log v_t$ e modelo linear aditivo está OK. Muito usado em econometria onde deseja-se medir o efeito em y_t da taxa de crescimento (ao invés do crescimento) de x_t .

Dados não-normais

Modelos Lineares Dinâmicos Generalizados \rightarrow
Extensão para família exponencial.

A equação de observação e de sistema agora são definidas por

$$f(y_t|\theta_t) = a(y_t)\exp\{y_t\theta_t + b(\theta_t)\} \text{ com } E(y_t|\theta_t) = \mu_t$$

$$g(\mu_t) = \mathbf{F}'_t\boldsymbol{\beta}_t$$

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

$$\boldsymbol{\beta}_0|D_0 \sim N(\mathbf{m}_0, \mathbf{C}_0)$$

$g(\mu_t)$ é a função de ligação.

Não há conjugação \rightarrow não há inferência de forma exata.

Generaliza o modelo linear generalizado (pg. 24) ao permitir estrutura dinâmica para os $\boldsymbol{\beta}_t$.

Exemplo: Regressão logística dinâmica

Regressão logística pode ser extendida ao permitir uma estrutura dinâmica nos coeficientes de regressão

$$y_t|\pi_t \sim \text{bin}(n_t, \pi_t), \quad t = 1, \dots, n$$

probabilidades π_t determinadas pelos valores de variável x

$$\pi_t = F(\alpha_t + \beta_t x_t) \quad , \quad t = 1, \dots, n$$

$$\alpha_t = \alpha_{t-1} + w_{1,t}$$

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

F - qq função de distribuição

22 Ciclos e leis de variância

Estimação de ciclos

De uma forma geral, ciclos podem ser definidos como processos que se repetem segundo um padrão regular. O exemplo mais simples já visto é a função definida através de harmônicos

$$y_t = a\cos(\theta t) + b\sin(\theta t) = r\cos(\theta t + \phi)$$

onde $r^2 = a^2 + b^2$ e $\cos(\phi) = a/r$.

A constante r é a amplitude da série.

O coeficiente θ é a frequência, já que a onda completa $\theta/2\pi$ ciclos em uma unidade de tempo

$\Rightarrow y_t$ completa um ciclo no tempo $\lambda = 2\pi/\theta$

$\Rightarrow \lambda$ é o comprimento de onda do harmônico.

ϕ é o ângulo de fase e indica o quão distante a onda deslocou-se na origem.

Considere um modelo AR(2), $y_t - a_1y_{t-1} - a_2y_{t-2} = 0$. Se $a_1^2 + 4a_2 < 0$, as raízes da equação polinomial $1 - a_1B - a_2B^2 = 0$ são imaginárias. Desta forma, a solução homogênea obtida é da forma $y_t = \beta_1 r^t \cos(\theta t + \beta_2)$; onde β_1 e β_2 são constantes arbitrárias, $r = (-a_2)^{1/2}$, e θ satisfaz a $\cos(\theta) = a_1/[2(-a_2)^{1/2}]$. Está claro que temos funções trigonométricas como solução, e elas impõem um padrão cíclico.

Aqui a frequência do ciclo é determinada por θ .

A condição de estabilidade é determinada pela magnitude de $r = (-a_2)^{1/2}$.

Se $|a_2| = 1$, as oscilações não alteram a amplitude; a solução homogênea é periódica.

As oscilações decairão se $|a_2| < 1$ e explodirão se $|a_2| > 1$.

Pode-se representar comportamento cíclico através de modelos auto-regressivos de segunda ordem. Acima obtivemos as relações entre as duas representações.

Quando $\beta_t = a_1\beta_{t-1} + a_2\beta_{t-2}$, o modelo pode ser descrito na forma de equações de espaço-estado, fazendo $\boldsymbol{\theta}_t = (\beta_t, \beta_{t-1})'$ assim

$$\beta_t = (1, 0)\boldsymbol{\theta}_t \quad \text{e} \quad \boldsymbol{\theta}_t = \mathbf{G}\boldsymbol{\theta}_{t-1}, \text{ com } \mathbf{G} = \begin{pmatrix} a_1 & a_2 \\ 1 & 0 \end{pmatrix} .$$

Dadas condições iniciais $\boldsymbol{\theta}_1 = (\beta_1, \beta_0)'$, a solução $\beta_t = (1, 0)\mathbf{G}^{t-1}\boldsymbol{\theta}_1$ tem forma funcional em t , determinada pela estrutura dos auto-valores da matriz de sistema \mathbf{G} . No caso de um par de auto-valores complexos conjugados temos exatamente a solução de um modelo AR(2).

Colocando o modelo AR(2) na forma estocástica temos $\beta_t = a_1\beta_{t-1} + a_2\beta_{t-2} + \omega_t$, onde ω_t são perturbações com média zero, independentes para todo t e usualmente normalmente distribuídas.

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

Uma série y_t que exibe comportamento cíclico de comprimento λ e decaimento r pode ser modelada por

$$\begin{aligned} y_t &= \beta_t + v_t \quad v_t \sim N(0, V) \\ \beta_t &= a_1\beta_{t-1} + a_2\beta_{t-2} + w_t \quad w_t \sim N(0, W), \end{aligned} \quad (51)$$

onde v_t e w_t são as perturbações das equações de observação e de sistema, respectivamente.

$\omega_t = 0, \forall t \Rightarrow y_t$ são observações de uma função coseno amortecida

O modelo pode ser escrito na forma

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

onde o parâmetro de estado é $\boldsymbol{\theta}_t = (\beta_t, \beta_{t-1})$ e a quádrupla que define o modelo é dada por:

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

onde $W = V(\omega_t)$, $r = (-a_2)^{1/2}$ e $\lambda = 2\pi/\cos^{-1}(a_1/2r)$.

Modelagem da Lei de Variância

Até agora, assumimos modelo normal usual com média e possivelmente variância desconhecida.

A propriedade básica da normal de variância independente da média foi mantida.

Pode acontecer de os dados não se comportarem segundo uma normal. Se supusermos que existe transformação g que operada sobre a série original normaliza a série, basta realizarmos a transformação e trabalhar com a série normalizada.

O problema, como já vimos, é que na hora de prever queremos trabalhar com os dados originais e o problema pode complicar dependendo da forma de g . Além disso, ao realizar transformações perdemos a interpretabilidade dos dados.

Uma forma alternativa de abordar o problema é modelar a forma que a variância vai depender da média.

Modelagem Determinística

Podemos permitir diferentes variâncias observacionais.

Suponha que $V(y_t) = k_t V$ onde k_t é conhecido. Pode-se mostrar que a única mudança que ocorre na evolução do modelo é que $Q_t = k_t S_{t-1} + \mathbf{F}'_t \mathbf{R}_t \mathbf{F}_t$.

(Quando $V(y_t) = V, \forall t$, aparecia 1 no lugar de k_t .)

Da discussão anterior vimos que podemos ser levados a $V(y_t) = k(\mu_t) V$. Os exemplos mais comuns são:

- i) $k(\mu_t) = \mu_t^p, p > 0$ ($p = 1 \Leftrightarrow$ Poisson)
- ii) $1 + b\mu_t^p$ ou $\mu_t(1 - \mu_t)$ (dados binomiais).

Tratamento exato é difícil e destrói teoria normal. Alternativa simples é substituir $k(\mu_t)$ por $k(f_t)$ na expressão de Q_t pois esse é o único lugar que k_t entra, se for conhecido, e f_t é melhor estimativa de $\mu_t|D_{t-1}$.

Outros exemplos de modelagem da variância são os modelos

1. ARCH (p) onde o modelo é descrito por

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

isto é, a variância observacional é descrita por defasagens quadráticas do erro de observação;

2. GARCH (p,q) onde o modelo é descrito por

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

onde a variância observacional, além de ser descrita por p defasagens quadráticas do erro de observação, também é descrita por q defasagens suas (autoregressão)

Modelagem estocástica

Até agora no modelo, V não muda com o tempo.

Modelo mais geral permite mudança com o tempo.

Suponha então que $\phi = V^{-1}$ é indexado por t e

$$\phi_t | D_{t-1} \sim G(n_{t-1}/2, d_{t-1}/2).$$

Fazendo $\phi_t = \phi_{t-1} + \psi_t$ onde $E(\psi_t) = 0$ e $V(\psi_t) = U_t$

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

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

Pensando multiplicativamente, fazemos

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

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

Assim, U_t pode ser definido através do desconto δ_V usado para a variância.

Para a análise prosseguir precisamos que $\phi_t | D_{t-1} \sim G(c_1, c_2)$.

Resolvendo, temos

$$\frac{c_1}{c_2} = E(\phi_{t-1} | D_{t-1}) \text{ e } \frac{c_1}{c_2^2} = V(\phi_{t-1} | D_{t-1})$$

$$\Rightarrow c_1 = \delta_V n_{t-1} / 2 \text{ e } c_2 = \delta_V d_{t-1} / 2.$$

(Esse resultado é obtido se $\delta_V \phi_t / \phi_{t-1} | \phi_{t-1} \sim \text{Beta}(\delta_V \frac{n_{t-1}}{2}; (1 - \delta_V) n_{t-1} / 2)$).

Análise segue como antes, basta multiplicar parâmetros da priori $\phi_t | D_{t-1}$ por δ_V . Tipicamente, δ_V próximo de 1.

Observações:

1. Filtragem pode ser feita aproximadamente (baseada nos dois primeiros momentos). Resultado é

$$S_t^{-1}(k) = E(\phi_{t-k}|D_t) = S_{t-k}^{-1} + \delta_{V,t-k+1}(S_t^{-1}(-k+1) - S_{t-k}^{-1})$$

$$V(\phi_{t-k}|D_t) = \frac{2}{n_{t-k}S_{t-k}^2}(1-\delta_{V,t-k+1}) + \frac{2}{n_t(-k+1)S_t^2(-k+1)}\delta_{V,t-k+1}$$

inicializada em $S_t(0) = S_t$ e $n_t(0) = n_t$.

2. Outro modelo utilizado para a modelagem da variância observacional é o da volatilidade estocástica, que é descrito por um processo AR(1) nas log volatilidades

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

23 Estimação de Hiperparâmetros

Introdução

Como foi visto, um MLD é definido pelas equações de observação e de sistema, isto é,

$$\begin{aligned}y_t &= \mathbf{F}'_t \boldsymbol{\theta}_t + v_t & v_t &\sim N(0, V_t) \\ \boldsymbol{\theta}_t &= \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t & \boldsymbol{\omega}_t &\sim N(0, \mathbf{W}_t)\end{aligned}$$

cujas quádrupla é definida por $M = \{\mathbf{F}_t, \mathbf{G}_t, V_t, \mathbf{W}_t\}$.

Se todos os elementos da quádrupla são conhecidos a inferência segue como no teorema 7.1. Se V é desconhecido, a inferência segue de acordo com o teorema ??.

Entretanto se há quantidades desconhecidas no vetor \mathbf{F}_t (ex.: previsão no VAR), na matriz \mathbf{G} (ex.: modelo de função de transferência), ou ainda, se os valores de \mathbf{W} ou δ não são conhecidos (comum na prática), o modelo é não linear e o teorema ?? não se aplica.

Estas quantidades são conhecidas como hiperparâmetros do modelo. Seja $\boldsymbol{\lambda}$ o vetor contendo todas as quantidades desconhecidas do modelo. Desta forma o modelo fica agora definido em função de $\boldsymbol{\lambda}$, isto é, $M(\boldsymbol{\lambda}) = \{\mathbf{F}_t(\boldsymbol{\lambda}), \mathbf{G}_t(\boldsymbol{\lambda}), V_t, \mathbf{W}_t(\boldsymbol{\lambda})\}$.

Deve ser notado que condicionado a um particular conjunto de valores das componentes de $\boldsymbol{\lambda}$ o modelo torna-se linear e a inferência segue normalmente.

Quando há presença de hiperparâmetros no modelo, estes devem ser integrados para que se possa fazer inferência sobre quantidades desconhecidas de interesse, $\boldsymbol{\theta}_t, y_{t+k}, \dots$.

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

Embora $p(\boldsymbol{\theta}_t|\boldsymbol{\lambda}, D_t)$ e $p(y_t|\boldsymbol{\lambda}, D_{t-1})$ sejam conhecidas, em geral, não conseguimos fazer a integração acima sem alguma aproximação.

Como fazer a inferência na presença de hiperparâmetros? Possíveis soluções são:

- (1) discretizar $\boldsymbol{\lambda}$: podemos assim fazer todas as contas pois integrais viram somas. Cada valor de $\boldsymbol{\lambda}$ define um modelo \Rightarrow temos modelos multiprocesso (cap. 12 do livro texto);
- (2) aplicar linearização;
- (3) fazer integração numérica;
- (4) simulação.

Veremos em detalhes algumas das técnicas citadas acima.

Linearização

Resolve não-linearidades em \mathbf{F} e/ou \mathbf{G} .

Considere o modelo com $\boldsymbol{\theta}_t = \lambda \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t$, não-linear.

Podemos redefinir os parâmetros como $\boldsymbol{\theta}_t = (\boldsymbol{\theta}_t, \lambda)$ evoluindo via

$$\boldsymbol{\theta}_t^* = g(\boldsymbol{\theta}_{t-1}) + \boldsymbol{\omega}_t^* \text{ onde}$$

$$\boldsymbol{\theta}^* = \begin{pmatrix} \boldsymbol{\theta}_t \\ \lambda \end{pmatrix}, g(\boldsymbol{\theta}_{t-1}^*) = \lambda \begin{pmatrix} \boldsymbol{\theta}_{t-1} \\ 1 \end{pmatrix} \text{ e } \boldsymbol{\omega}_t^* = \begin{pmatrix} \boldsymbol{\omega}_t \\ 0 \end{pmatrix}$$

Idéia é aproximar g por função linear usando aproximação de Taylor com truncamento no termo de 1a. ordem.

A aproximação é feita em torno da melhor estimativa de $\boldsymbol{\theta}_{t-1}$ no tempo $t - 1$, dada por \mathbf{m}_{t-1} .

Como $g(\boldsymbol{\theta}_{t-1}) = g(\mathbf{m}_{t-1}) + \mathbf{G}_t(\boldsymbol{\theta}_{t-1} - \mathbf{m}_{t-1}) + \text{termos de ordem superior}$ onde $\mathbf{G}_t = \left. \frac{\partial g(\boldsymbol{\theta}_{t-1})}{\partial \boldsymbol{\theta}_{t-1}} \right|_{\boldsymbol{\theta}_{t-1} = \mathbf{m}_{t-1}}$ temos que

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

No exemplo, $\mathbf{G}_t = \left. \begin{pmatrix} \lambda \mathbf{1} & \boldsymbol{\theta}_{t-1} \\ 0 & 1 \end{pmatrix} \right|_{\boldsymbol{\theta}_{t-1} = \mathbf{m}_{t-1}}$.

Daí, obtém-se $\boldsymbol{\theta}_t | D_{t-1} \sim N(\mathbf{a}_t, \mathbf{R}_t)$ onde

$$\mathbf{a}_t = \mathbf{h}_t + \mathbf{G}_t \mathbf{m}_{t-1} \text{ e}$$

$$\mathbf{R}_t = \mathbf{G}_t \mathbf{C}_{t-1} \mathbf{G}_t' + \mathbf{W}_t$$

Integração numérica

Inferência Bayesiana exata nem sempre é possível.

Integração numérica é baseada em preceitos determinísticos.

Difícil uso em espaços paramétricos maiores ($p > 10$).

1) Aproximação da posteriori (π) pela normal

Obtida a partir da expansão de Taylor do $\log \pi$ em torno da moda.

MLG: moda obtida a partir de algoritmo MQRI (mínimos quadrados reponderados iterativos)

a cada iteração: mod. de regressão normal é construído.

2) Aproximação de Laplace

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

Aproximam-se integrandos por formas quadráticas no log.
Erros de ordem $O(n^{-1})$ se anulam \rightarrow erro fica $O(n^{-2})$.

Densidades marginais podem ser obtidas mas custo computacional é alto.

3) Aproximação via quadratura Gaussiana

Aproxima integrandos da forma $\exp(-x^2/2)h(x)$

Se h for uma função polinomial \rightarrow aproximação é exata.

Aproximações baseadas na normal: reparametrizações podem ser úteis.

Métodos baseados em simulação estocástica

Inferência baseada em amostras da posteriori $\pi(\boldsymbol{\theta})$ onde $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)$.

Amostra é sempre substituto parcial da informação de densidade.

Métodos analíticos: erros diminuem com o aumento do número de observações, que não depende do analista.

Métodos baseados em simulação: erros dependem do número de valores gerados, que só depende do analista.

Considere amostra de $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_n$ de π .

Amostra da marginal de $\theta_i : \theta_{i1}, \dots, \theta_{in}$

Muito mais simples que obtenção analítica da marginal

Amostra de $t(\boldsymbol{\theta}) : t_1 = t(\boldsymbol{\theta}_1), \dots, t_n = t(\boldsymbol{\theta}_n)$

Muito mais simples que obtenção da posteriori de $t(\boldsymbol{\theta})$

De posse de uma amostra é possível obter:

- (i) estimativas pontuais,
- (ii) intervalos de confiança,
- (iii) reconstrução de densidades marginais.

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

Simulação via reamostragem

Normalmente, complicado fazer geração da posteriori π .

Idéia: usar técnicas de reamostragem tomando uma densidade auxiliar q (que pode ser a priori de θ).

Essa densidade é conhecida como densidade de importância.

Usando o método da rejeição, deve existir c conhecida, tal que $\pi(\theta) \leq cq(\theta)$, $\forall \theta$. Assim a amostragem é feita:

- (i) gerando-se um valor θ da densidade $q(\theta)$;
- (ii) aceitando-se esse valor com probabilidade $w(\theta) = \frac{\pi(\theta)}{cq(\theta)}$
Se q =priori, $w(\theta) = l(\theta)/l_{max}$.

Usando o método da reamostragem ponderada (SIR)

- (i) gera-se amostra de $\theta_1, \dots, \theta_n$ da densidade $q(\theta)$;
- (ii) reamostram-se os valores da amostra com probabilidades

$$h_i = \frac{w(\theta_i)}{\sum_{j=1}^n w(\theta_j)}, \quad i = 1, 2, \dots, n. \quad (52)$$

$$\text{Se } q\text{=priori, } h_i = \frac{l(\theta_i)}{\sum_{j=1}^n l(\theta_j)}$$

Problemas:

- a) M.Rej.: reamostra menor que a amostra original;
- b) conflito entre priori e verossimilhança;
- c) maximização da verossimilhança, no método da rejeição;
- d) No método SIR, amostra de π é aproximada.

Método SIR em 2 passos

Sejam $q(\boldsymbol{\theta})$ uma densidade de importância,
 $p(\boldsymbol{\theta})$ uma densidade a priori
 $l(\boldsymbol{\theta})$ a verossimilhança e
 $\pi(\boldsymbol{\theta})$ a posteriori,

1. como primeira aproximação use o método SIR como descrito anteriormente; isto é a partir de q (ex.: priori) obtenha uma amostra da posteriori usando os pesos de 52;
2. a partir desta primeira amostra defina os parâmetros de uma nova função de importância q^* . No PRV, esses parâmetros são calculados de acordo com os valores mínimo e máximo da amostra da primeira aproximação;
3. obtenha agora uma amostra da função de importância com parâmetros definidos de acordo com o passo 2. Reamostragem com probabilidades,

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

onde

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

Aproximação via MCMC (Markov Chain Monte Carlo)

1) constrói-se cadeia de Markov cuja

- transição é feita de acordo com condicionais

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

\Rightarrow distribuição de equilíbrio é $\pi(\boldsymbol{\theta}, \boldsymbol{\lambda})$.

2) Gera-se uma trajetória dessa cadeia

$$\begin{array}{ccccccc} \text{Iteração } 0 & \rightarrow & \cdots & \rightarrow & \text{Iteração } j & \rightarrow & \text{Iteração } j + 1 & \rightarrow & \cdots \\ (\boldsymbol{\theta}^{(0)}, \boldsymbol{\lambda}^{(0)}) & & & & (\boldsymbol{\theta}^{(j)} \dots \boldsymbol{\lambda}^{(j)}) & & (\boldsymbol{\theta}^{(j+1)} \dots \boldsymbol{\lambda}^{(j+1)}) & & \end{array}$$

Esse algoritmo é conhecido como amostrador de Gibbs.

Valores gerados após equilíbrio constituem amostra de $\pi(\boldsymbol{\theta}, \boldsymbol{\lambda})$.

Exemplo: considere o modelo linear dinâmico com:

$\mathbf{F}_t, \mathbf{G}_t$ conhecidos e

$\boldsymbol{\theta}' = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_n)$, V e bfW desconhecidos

Precisamos obter as distribuições condicionais de $(\boldsymbol{\theta}, V|\mathbf{W})$ e $(\mathbf{W}|\boldsymbol{\theta}, V)$.

Pode se mostrar que:

(i) $(\boldsymbol{\theta}, V|\mathbf{W})$ é NG \rightarrow fácil de gerar

(ii) $(\mathbf{W}|\boldsymbol{\theta})$ é WI se a priori \mathbf{W} for WI \rightarrow fácil de gerar

Quando não se sabe gerar de $\pi(\boldsymbol{\lambda}|\boldsymbol{\theta})$

- gera-se $\boldsymbol{\lambda}^{(n)}$ de uma transição proposta $q(\boldsymbol{\lambda}|\boldsymbol{\lambda}^{(v)}, \boldsymbol{\theta})$
Ex.: $\boldsymbol{\lambda}|\boldsymbol{\lambda}^{(v)} \sim N(\boldsymbol{\lambda}^{(v)}, c\mathbf{I})$
- aceita-se $\boldsymbol{\lambda}^{(n)}$ com probabilidade $\alpha(\boldsymbol{\lambda}^{(v)}, \boldsymbol{\lambda}^{(n)})$ onde

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

Esse algoritmo é conhecido como algoritmo de Metropolis-Hastings.

Valores gerados após equilíbrio também constituem amostra de $\pi(\boldsymbol{\theta}, \boldsymbol{\lambda})$.

Exemplo: modelo linear dinâmico não-normal com:

$\mathbf{F}_t, \mathbf{G}_t$ conhecidos e

$\boldsymbol{\theta}' = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_n)$ e \mathbf{W} desconhecidos

Pode se mostrar que:

(i) $(\mathbf{W}|\boldsymbol{\theta})$ é WI se a priori \mathbf{W} for WI \rightarrow fácil de gerar

(ii) $(\boldsymbol{\theta}|\mathbf{W})$ não é nenhuma distr. conhecida \rightarrow não é fácil de gerar

Proposta baseada no algoritmo de MQRI ou em passeios aleatórios

24 Aplicação

Uso do software PRVWIN.

Exercícios com a série do produto da indústria geral do Brasil.

1. Inclusão do bloco de ciclo no modelo da série;
2. Estimação dos hiperparâmetros via SIR (em dados simulados e também na série do produto da indústria geral);
3. Exemplo de dados sobre memorização de propaganda (dados não-normais).