Aula 2 - Econometria Financeira 1.

Pedro Albuquerque

Universidade de Brasília

26 de Março de 2018

UnB

O pressuposto principal da análise de séries temporais é assumir que o processo e suas propriedades (pelo menos alguma delas) não variem com o tempo.

Propriedades básicas: Anteriormente introduzimos o conceito de função de autocovariância de séries temporais, $\{x_t\}$ tal que:

1. $\gamma(h) = Cov(X_{t+h}, X_t)$.
2. $\rho(h) = \gamma(h)/\gamma(0)$

para $h = 0, \pm 1, \pm 2, \ldots.$

Essas métricas são úteis na avaliação de processos estacionários, por exemplo suponha $\{X_t\}$ uma série temporal gaussiana, isto é para qualquer coleção de inteiros $i_1, \ldots, i_n$ o vetor aleatório $(X_{i_1}, \ldots, X_{i_n})^\top$ possui distribuição normal multivariada.

Suponha ainda que observamos X_n, nesse caso qual seria a "melhor"previsão para X_{n+h} ?

Por definição, o "melhor"preditor será aquele que minizar o Erro Quadrático Médio:

$$\text{MSE}(\hat{\theta}) = \mathbb{E}_{\hat{\theta}}\left[(\hat{\theta} - \theta)^2\right] = \mathbb{V}_{\hat{\theta}}(\hat{\theta}) + \text{Bias}_{\hat{\theta}}(\hat{\theta}, \theta)^2$$

Para isso, seja $\mathbf{X}$ um vetor aleatório com distribuição Normal Multivariada com vetor de médias $\boldsymbol{\mu}$ e matriz de variâncias e covariâncias $\Sigma = \Sigma_{\mathbf{xx}}$, de maneira sumarizada $\mathbf{X} \sim N_n(\boldsymbol{\mu}, \Sigma_{\mathbf{xx}})$.

Nesse caso, a função densidade de probabilidade é dada por:

$$f_{\mathbf{X}}(\mathbf{x}) = (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}$$

Se $\mathbf{X} \sim N_n(\boldsymbol{\mu}, \Sigma_{\mathbf{xx}})$ podemos definir a normal multivariada padronizada $\mathbf{Z}$ fazendo:

$$\mathbf{Z} = \Sigma^{-\frac{1}{2}}(\mathbf{X}-\boldsymbol{\mu})$$

onde $\Sigma^{-\frac{1}{2}}$ é obtido pela inversa da decomposição na forma $\Sigma^{-1} = \mathbf{A}\mathbf{A}^{\top}$ onde $\mathbf{A} = \Sigma^{-\frac{1}{2}}$.

Nesse caso temos que $\mathbf{Z} \sim N_n(\mathbf{0}, \mathbf{I}_n)$. Suponha agora que possamos particionar nosso vetor aleatório na forma:

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}^{(1)} \\ \mathbf{X}^{(2)} \end{bmatrix} \tag{1}$$

tal que:

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}^{(1)} \\ \mathbf{X}^{(2)} \end{bmatrix} \sim N_n\left(\begin{bmatrix} \boldsymbol{\mu}^{(1)} \\ \boldsymbol{\mu}^{(2)} \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right) \tag{2}$$

Nesse formato temos $\boldsymbol{\mu}^{(i)} = \mathbb{E}(\mathbf{X}^{(i)})$ e $\Sigma_{ij} = \mathbb{E}[(\mathbf{X}^{(i)} - \boldsymbol{\mu}^{(i)})(\mathbf{X}^{(j)} - \boldsymbol{\mu}^{(j)})^{\top}]$.

Se $\mathbf{X}^{(1)}$ e $\mathbf{X}^{(2)}$ são independentes então $\Sigma_{12} = \Sigma_{21}^{\top} = \mathbf{0}$. Já a distribuição condicional de $\mathbf{X}^{(1)}|\mathbf{X}^{(2)} = \mathbf{x}^{(2)}$ é dada por:

$$\mathbf{X}^{(1)}|\mathbf{X}^{(2)} = \mathbf{x}^{(2)} \sim N_m(\boldsymbol{\mu}^{(1)}+\Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}^{(2)}-\boldsymbol{\mu}^{(2)}), \Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$$

Particularmente temos:

$$\mathbb{E}[\mathbf{X}^{(1)}|\mathbf{X}^{(2)} = \mathbf{x}^{(2)}] = \boldsymbol{\mu}^{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}^{(2)} - \boldsymbol{\mu}^{(2)})$$

Usando essa proposição para nossas séries temporais temos que a distribuição de X_{n+h} dado X_n é:

$$X_{n+h}|X_n \sim N(\mu + \rho(h)(x_n - \mu), \sigma^2(1 - \rho(h)^2))$$

onde μ e σ^2 são as médias e variâncias de X_t. Assim o mlehor estimador para a previsão é tal que:

$$\arg\min_{c \in \mathbb{R}} \mathbb{E}[X_{n+h} - c]^2 = \mathbb{E}[X_{n+h}]$$

para um preditor constante.

UnB

No caso de um preditor funcional temos:

$$\underset{m(X_n)\in\mathcal{F}}{\arg\min}\,\mathbb{E}[X_{n+h}-m(X_n)]^2=\sigma^2(1-\rho(h)^2)$$

Obs: O que acontece se $\rho(h)\to\pm 1$?

Outra métrica possível é considerar somente os funcionais lineares no espaço de funções $\mathcal{F}_\ell\subset\mathcal{F}$, na forma:

$$\underset{l(X_n)\in\mathcal{F}_\ell}{\arg\min}\,\mathbb{E}[X_{n+h}-(aX_n+b)]^2$$

o qual é mais fácil de ser obtido do que a minimização global do MSE.

Nesse exemplo o melhor preditor linear é dado por:

$$\ell(X_n) = \mu + \rho(h)(X_n - \mu)$$

o qual fornece MSE igual a:

$$\mathbb{E}[X_{n+h} - \ell(X_n)]^2 = \sigma^2(1 - \rho(h)^2)$$

As propriedades básicas da função de autocovariância são:

1. $\gamma(0) \geq 0$.
2. $|\gamma(h)| \leq \gamma(0) \forall h$
3. $\gamma(h)$ é uma função par, isto é, $\gamma(h) = \gamma(-h)$

A função de autocovariância possui outra propriedade fundamental dada por:

$$\sum_{i=1}^{n} \sum_{i=1}^{n} a_i \kappa(i-j) a_j \geq 0$$

para todo inteiro n e vetor real $\mathbf{a} = (a_1, \ldots, a_n)^\top$ para qualquer função real $\kappa(\cdot, \cdot)$.

Definição.

Uma série é dita ser estritamente estacionária se
$(X_1, \ldots, X_n)^\top \stackrel{d}{=} (X_{1+h}, \ldots, X_{n+h})^\top$.
Para todos os inteiros h e $n \geq 1$.
Obs: Dizemos que $\{X_t\}$ é fracamente estacionário se $\mathbb{E}(X_t^2) < \infty$ para todo t.

outra maneira mais simples de construir uma série $\{X_t\}$ estritamente estacionária é filtrar uma sequência iid de variáveis aleatórias.

Seja $\{Z_t\}$ uma sequência iid definimos:

$$X_t = g(Z_t, \ldots, Z_{t-1}, \ldots, Z_{t-q})$$

para alguma função real $g(\cdot, \ldots, \cdot)$. Então $\{X_t\}$ será estritamente estacionária também denominada de q-dependente ou q-correlacionada.

Exemplo 2.1

Dizemos que $\{X_t\}$ é um processo de médias móveis de ordem q (MA(q)) se:

$$X_t = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}$$

tal que $\{Z_t\} \sim WN(0, \sigma^2)$ e $\theta_1, \ldots, \theta_q$ constantes.

Proposição: Se $\{X_t\}$ é um processo estacionário q-correlacionado com média zero, então esse processo pode ser representado por um processo $MA(q)$.

A classe de modelos de séries temporais lineares, o qual inclui a família ARMA fornece um panorama geral para o estudo dos processos estacionários.

De fato, qualquer processo estacionário de segunda ordem ou é um processo linear ou pode ser transformado em processo linear. Isso é conhecido como **Decomposição de Wold**.

Definição:

A série temporal $\{X_t\}$ é um processo linear se possui a seguinte representação:

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$$

para todo t onde $\{Z_t\} \sim WN(0, \sigma^2)$ e $\{\psi_j\}$ é uma sequência de constantes com $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$.

Em termos do operador de defasagem $\mathcal{B}$ podemos reescrever como

$$X_t = \psi(\mathcal{B})Z_t \text{ onde } \psi(\mathcal{B}) = \sum_{j=-\infty}^{\infty} \psi_j \mathcal{B}^j.$$

O operador $\psi(\mathcal{B})$ pode ser visto como um **filtro linear**, o qual quando aplicado ao ruído branco $\{Z_t\}$ produz como resultado $\{X_t\}$.

Proposição:

Seja $\{Y_t\}$ uma série temporal estacionária com média zero e função de autocovariância γ_y. Se $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ então a série temporal:

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Y_{t-j} = \psi(\mathcal{B})Y_t$$

é estacionária com média zero e função de autocovariância:

$$\gamma_x(h) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_k \gamma(h+k-j)$$

UnB

Exemplo 2.2

Considere um modelo AR(1) na forma $X_t - \phi X_{t-1} = Z_t$ onde $\{Z_t\} \sim WN(0, \sigma^2)$ e $|\phi| < 1$ e Z_t não correlacionado com X_s, $s < t$. Seja o processo:

$$X_t = \sum_{j=0}^{\infty} \phi^j Z_{t-j}$$

é um caso geral para $AR(1)$. Nesse caso:

$$\gamma_X(h) = \sum_{j=0}^{\infty} \phi^j \phi^{j+h} = \frac{\sigma^2 \phi^h}{1 - \phi^2}$$

$h \geq 0$.

Note que ainda pelo exemplo anterior temos:

$$
\begin{aligned}
Y_t &= \phi Y_{t-1} + Z_{t-1} & (3)\\
&= Z_t + \phi Z_{t-1} + \phi^2 Y_{t-2} & (4)\\
&= \vdots & (5)\\
&= Z_t + \phi Z_{t-1} + \cdots + \phi^k Z_{t-k} + \phi^{k+1} Y_{t-k+1} & (6)
\end{aligned}
$$

Se $\{Y_t\}$ é um processo estacionário então $\mathbb{E}[Y_t^2]$ é finito e não depende de t, logo:

$$\mathbb{E}[Y_t - \sum_{j=0}^{k} \phi^j Z_{t-j}]^2 = \phi^{2k+2}\mathbb{E}[Y_{t-k-1}]^2 \to 0$$

quando $k \to \infty$. No caso de $|\phi| > 1$ a série não converge.

Seja $\{X_t\}$ um processo ARMA(1,1) então seu PGD é dado por:

$$X_t - \phi X_{t-1} = Z_t + \theta Z_{t-1}$$

onde $\{Z_t\} \sim WN(0, \sigma^2)$ e $\phi + \theta \neq 0$.

Usando o operador de defasagem podemos reescrever na forma:

$$\phi(\mathcal{B})X_t = \theta(\mathcal{B})Z_t$$

onde $\phi(\mathcal{B})$ e $\theta(\mathcal{B})$ são os filtros lineares: $\phi(\mathcal{B}) = 1 - \phi\mathcal{B}$ e $\theta(\mathcal{B}) = 1 + \theta\mathcal{B}$.

O primeiro passo é estudar os valores de ϕ e θ para os quais a solução é estacionária.

Se $|\phi| < 1$, seja $\chi(z)$ a expansão por séries de potência de:

$$\chi(z) = \sum_{j=0}^{\infty} \phi^j z^j$$

o qual possui coeficientes absolutamente somáveis.

No caso do processo ARMA(1,1) concluimos que $\chi(\mathcal{B})\phi(\mathcal{B}) = 1$. Aplicando $\chi(\mathcal{B})$ dos dois lados temos:

$$\phi(\mathcal{B})X_t = \theta(\mathcal{B})Z_t \tag{7}$$

$$\chi(\mathcal{B})\phi(\mathcal{B})X_t = \chi(\mathcal{B})\theta(\mathcal{B})Z_t \tag{8}$$

$$X_t = \chi(\mathcal{B})\theta(\mathcal{B})Z_t = \phi(\mathcal{B})Z_t \tag{9}$$

UnB

$$\psi(\mathcal{B}) = \sum_{k=-\infty}^{\infty} \psi_j \mathcal{B}^j = (1 + \phi\mathcal{B} + \phi^2\mathcal{B}^2 + \dots)(1 + \theta\mathcal{B})$$ multiplicando pelo lado direito temos:

$$\psi_0 = 1, \psi_j = (\phi + \theta)\phi^{j-1} \text{ para } j \geq 1$$

Concluimos que o processo $MA(\infty)$ é a única solução, isto é:

$$X_t = Z_t + (\phi + \theta)\sum_{j=1}^{\infty} \phi^{j-1} Z_{t-j}$$

Agora suponha que $|\phi| > 1$. Assim como anteriormente o primeiro passo é representar $1/\phi(z)$ como uma série de potências de z com coeficientes absolutamente somáveis pela expansão z^{-1}:

$$\chi(z) = -\sum_{j=1}^{\infty} \phi^{-j} z^{-1}$$

então podemos avaliar o mesmo argumento como no caso $|\phi| < 1$ para obter a solução única do processo estacionário.

Seja $\chi(\mathcal{B}) = -\sum_{j=1}^{\infty} \phi^j \mathcal{B}^j$ e aplicando $\chi(\mathcal{B})$ dos dois lados obtemos:

$$X_t = \chi(\mathcal{B})\theta(\mathcal{B})Z_t = -\theta\phi^{-1}Z_t - (\theta + \phi)\sum_{j=1}^{\infty}\phi^{-j-1}Z_{t+j}$$

Se $\phi = \pm 1$ não existe solução estacionária. Consequentemente nã há um processo ARMA(1,1) com $\phi = \pm 1$, em resumo para um processo ARMA(1,1):

1. Uma solução estacionária ARMA(1,1) só existe se $\phi \neq \pm 1$.
2. Se $|\phi| < 1$ então a solução estacionária é unicamente dada por $X_t = Z_t + (\phi + \theta)\sum_{j=1}^{\infty}\phi^{j-1}Z_{t-j}$. Nesse caso dizemos que $\{X_t\}$ é causado por $\{Z_t\}$ uma vez que X_t pode ser expresso em termos de Z_s para $s \leq t$.

UnB

Finalmente, se $|\phi| > 1$ então a solução estacionária é dada por:

$$X_t = \chi(\mathcal{B})\theta(\mathcal{B})Z_t = -\theta\phi^{-1}Z_t - (\theta+\phi)\sum_{j=1}^{\infty}\phi^{-j-1}Z_{t+j}$$

Essa é uma solução **não causal** uma vez que X_t é uma função de Z_s com $s \geq t$.

Vamos mostrar agora que um processo ARMA(1,1):

$$X_t - \phi X_{t-1} = Z_t + \theta Z_{t-1}$$

com $\{Z_t\} \sim WN(0, \sigma^2)$ e $\phi + \theta \neq 0$ é invertível (e portanto estacionário) se $|\theta| < 1$.

Para demonstrar isso, seja $\zeta(z)$ a expansão por séries de potência de $1/\theta(z)$, isto é, $\sum_{j=0}^{\infty}(-\theta)^j z^j$ o qual possui coeficientes absolutamente somáveis.

Uma vez que $\psi(\mathcal{B}) = \alpha(\mathcal{B})\beta(\mathcal{B}) = \beta(\mathcal{B})\alpha(\mathcal{B})$ temos $\zeta(\mathcal{B})\theta(\mathcal{B}) = 1$. Aplicando $\zeta(\mathcal{B})$ temos:

$$\phi(\mathcal{B})X_t = \theta(\mathcal{B})Z_t \tag{10}$$

$$\zeta(\mathcal{B})\phi(\mathcal{B})X_t = \zeta(\mathcal{B})\theta(\mathcal{B})Z_t \tag{11}$$

$$Z_t = \zeta(\mathcal{B})\phi(\mathcal{B})X_t = \pi(\mathcal{B})X_t \tag{12}$$

onde

$$\pi(\mathcal{B}) = \sum_{j=0}^{\infty} \pi_j \mathcal{B}^j (1 - \theta\mathcal{B} + (-\theta)^2\mathcal{B} + \dots)(1 - \phi\mathcal{B})$$

Como $\psi_j = \sum_{k=-\infty}^{\infty} \alpha_k \beta_{j-k} = \sum_{k=-\infty}^{\infty} \beta_k \alpha_{j-k}$ temos:

$$Z_t = X_t - (\phi + \theta) \sum_{j=1}^{\infty} (-\theta)^{j-1} X_{t-j}$$

então, o processo ARMA(1,1) é invertível uma vez que Z_t pode ser expresso em termo dos valores presentes e passados de X_s com $s \leq t$.

Um argumento similar pode ser usado para demonstrar a **não causalidade** quando $|\phi| > 1$ o processo ARMA(1,1) é não-invertível quando $|\theta| > 1$, uma vez que:

$$Z_t = -\phi\theta^{-1}X_t + (\theta + \phi)\sum_{j=1}^{\infty}(-\theta)^{-j-1}X_{t+j}$$

Assim temos que se $|\theta| < 1$ o processo ARMA(1,1) é invertível e Z_t é expresso em termos de X_s $s \leq t$.

Se $\theta > 1$ o processo ARMA(1,1) é não-invertível e Z_t é expresso em termos de X_s com $s \geq t$.

Note que $\gamma(\cdot)$ e $\rho(\cdot)$ são estimados como:

$$\hat{\gamma}(h) = n^{-1} \sum_{t=1}^{n-|h|} (X_{t+|h|} - \overline{X}_n)(X_t - \overline{X}_n)$$

e

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$$

Podemos construir a matriz de autocovariância amostral como:

$$\hat{\Gamma}_k = \begin{bmatrix} \hat{\gamma}(0) & \hat{\gamma}(1) & \dots & \hat{\gamma}(k-1) \\ & \hat{\gamma}(0) & \dots & \vdots \\ & & \ddots & \vdots \\ & & & \hat{\gamma}(0) \end{bmatrix} \tag{13}$$

De igual modo a matriz de autocorrelação é construída por:

$$\hat{R}_k = \frac{\hat{\Gamma}_k}{\hat{\gamma}(0)}$$

Box e Jenkins (1976) p. 33 sugere que se use $n \geq 0$ e $h \leq n/4$.

A distribuição amostral é dada por $\hat{\boldsymbol{\rho}}_k \sim N(\hat{\boldsymbol{\rho}}_k, n^{-1}W)$ com $\hat{\boldsymbol{\rho}} = (\hat{\rho}(1), \ldots, \hat{\rho}(k))$

e W é a matriz obtida pela fórmula de Bartlett:

$$W_{ij} = \sum_{k=1}^{\infty}\{\rho(k+i)+\rho(k-i)-2\rho(i)\rho(k)\}\{\rho(k+j)+\rho(k-j)-2\rho(j)\rho(k)\}$$

UnB

Exemplo 2.3

Considere a série elec.txt calcule a distribuição amostral da autocorrelação usando o R.

Agora vamos considerar o problema de previsão para X_{n+h} e $h \geq 0$ de uma série estacionária com média μ e função de autocovariância $\gamma(\cdot)$ em termos dos valores de $\{X_n, \ldots, X_1\}$.

O objetivo é encontrar a combinação linear de $1, X_n, X_{n-1}, \ldots, X_1$ para prever X_{n+h} com o menor MSE possível.

O melhor preditor linear em termos de $1, X_n, X_{n-1}, \ldots, X_1$ será denotado por $\mathcal{P}_n X_{n+h} = a_0 + a_1 X_n + \cdots + a_n X_1$ onde o processo é:

$$\arg\min_{\mathbf{a} \in \mathbb{R}^{n+1}} \mathbb{E}[(X_{n+h} - a_0 - a_1 X_n - \cdots - a_n X_1)^2]$$

Para resolver esse problema derivamos e igualamos a zero, obtendo assim o sistema:

$$\begin{cases} \mathbb{E}\left[X_{n+h} - a_0 - \sum_{i=1}^{n} a_i X_{n+1-i}\right] = 0 \\ \mathbb{E}\left[\left(X_{n+h} - a_0 - \sum_{i=1}^{n} a_i X_{n+1-i}\right) X_{n+1-j}\right] = 0 \end{cases} \tag{14}$$

para $j = 1, \ldots, n$. Essas equações podem ser escritas na forma $a_0 = \mu(1 - \sum_{i=1}^{n} a_i)$ e $\Gamma_n a_n = \gamma_n(h)$.

Nesse caso temos $\gamma_n(h) = (\gamma(h), \gamma(h+1), \ldots, \gamma(h+n-1))^\top$ e portanto $\mathcal{P}_n X_{n+h} = \mu + \sum_{i=1}^{n} a_i(X_{n+1-i} - \mu)$, onde $\mathbf{a}_n$ satisfa $\Gamma_n \mathbf{a}_n = \gamma_n(h)$.

Exemplo 2.4

Simule um processo AR(1) com 1000 obeservações e faça a melhor previsão linear para as últimas observações usando como base as 950 primeiras.

Faça isso usando o R.

Para o operador de previsão temos as seguintes propriedades:

1. $\mathcal{P}_n X_{n+h} = \mu + \sum_{i=1}^{n} a_i (X_{n+1-i} - \mu)$.
2. $\mathbb{E}[X_{n+h} - \mathcal{P}_n X_{n+h}]^2 = \gamma(0) - \mathbf{a}_n^\top \gamma_n(h)$.
3. $\mathbb{E}(X_{n+h} - \mathcal{P}_n X_{n+h}) = 0$.
4. $\mathbb{E}[(X_{n+h} - \mathcal{P}_n X_{n+h}) X_j] = 0, j = 1, \ldots, n$.

UnB

Exemplo 2.5

Considere uma série estacionária AR(1) com P.G.D dado por $X_t = \phi X_{t-1} + Z_t$ onde $|\phi| < 1$ e $Z_t \sim RB(0, \sigma^2)$ encontre o melhor preditor linear e seu erro-quadrático médio.

Suponha que Y e $W_n, \ldots, W_1$ sejam quaisquer variáveis aleatórias com segundo momento finito e médias $\mu = \mathbb{E}(Y), \mu_i = \mathbb{W}_i$ e covariâncias dadas por $Cov(Y, Y), Cov(Y, W_i)$ e $Cov(W_i, W_j)$ postuladas serem conhecidas.

Podemos escrever na forma matricial $\mathbf{W} = (W_n, \ldots, W_1)^\top$ com vetor de médias $\boldsymbol{\mu}_{\mathbf{W}} = (\mu_n, \ldots, \mu_1)^\top$ e covariãncias:

$$\gamma = Cov(Y, \mathbf{W}) = [Cov(Y, W_n), \ldots, Cov(Y, W_1)]^\top$$

com matriz:

$$\Gamma = Cov(\mathbf{W}, \mathbf{W})$$

Usando a mesma argumentação é possível mostrar que o melhor preditor linear é dado por:

$$\mathcal{P}(Y|\mathbf{W}) = \mu_Y + \mathbf{a}^\top(\mathbf{W} - \boldsymbol{\mu}_{\mathbf{W}})$$

onde $\mathbf{a} = (a_1, \ldots, a_n)^\top$ é ma solução do sistema:

$$\Gamma\mathbf{a} = \gamma$$

Exemplo 2.6

Considere uma série estacionária AR(1) com P.G.D dado por $X_t = \phi X_{t-1} + Z_t$ onde $|\phi| < 1$ e $Z_t \sim RB(0, \sigma^2)$. Suponha que observamos X_1, X_3 mas não X_2 encontre o melhor preditor linear para X_2 e seu erro-quadrático médio.

Para $\mathbf{W} = (W_n, \ldots, W_1)^\top$ e Y com segundo momento finito, vimos como calcular o melhor preditor linear de Y em termos de $W_n, \ldots, W_1$.

A função $\mathcal{P}(\cdot|\mathbf{W})$ a qual converte Y em $\mathcal{P}(Y|\mathbf{W})$ é denominado **operador de previsão**, cujas propriedades são dadas a seguir. Suponha que $\mathbb{E}(U^2) < \infty$, $\mathbb{E}(V^2) < \infty$, $\Gamma = Cov(\mathbf{W}, \mathbf{W})$ e $\boldsymbol{\beta}, \alpha_1, \ldots, \alpha_n$ são constantes então:

1. $\mathcal{P}(U|\mathbf{W}) = \mathbb{E}(U) + \mathbf{a}^\top(\mathbf{W} - \mathbb{E}(W))$ onde $\Gamma\mathbf{a} = Cov(U, \mathbf{W})$.
2. $\mathbb{E}[(U - \mathcal{P}(U|\mathbf{W}))\mathbf{W}] = \mathbf{0}$ e $\mathbb{E}[U - \mathcal{P}(U|\mathbf{W})] = 0$.
3. $\mathbb{E}[(U - \mathcal{P}(U|\mathbf{W}))^2] = \mathbb{V}(U) - \mathbf{a}^\top Cov(U, \mathbf{W})$.
4. $\mathcal{P}(\alpha_1 U + \alpha_2 V + \beta|\mathbf{W}) = \alpha_1 \mathcal{P}(U|\mathbf{W}) + \alpha_2 \mathcal{P}(V|\mathbf{W}) + \beta$.
5. $\mathcal{P}(\sum_{i=1}^n \alpha_i W_i + \beta|\mathbf{W}) = \sum_{i=1}^n \alpha_i W_i + \beta$.
6. $\mathcal{P}(U|\mathbf{W}) = \mathbb{E}(U)$ se $Cov(U, \mathbf{W}) = \mathbf{0}$.
7. $\mathcal{P}(U|\mathbf{W}) = \mathcal{P}(\mathcal{P}(U|\mathbf{W}, \mathbf{V})|\mathbf{W})$ se $\mathbf{V}$ é um vetor aleatótio tal que $\mathbb{E}(\mathbf{V}\mathbf{V}^\top)/, \infty$

Exemplo 2.7

Seja a série temporal $\{Y_t\}$ com P.G.D AR(1) com média μ. Seja $\{X_t = Y_t - \mu\}$ um processo AR(1) com média zero mostre que:

$$Y_t - \mu = \phi(Y_{t-1} - \mu) + Z_t$$

Encontre o melhor preditor linear para $\mathcal{P}_n Y_{n+h}$

Uma vez que $\mathcal{P}(Y|\mathbf{W}) = \mu_Y + \mathbf{a}^\top(\mathbf{W} - \mu_{\mathbf{W}})$ e $\Gamma\mathbf{a} = \gamma$ podemos construir um algoritmo para estimar os parâmetros dos models de séries temporais lineares, como:

$$\mathcal{P}_n X_{n+1} = \boldsymbol{\phi}_n^\top X_n = \phi_{n1} X_n + \cdots + \phi_{nn} X_1 \tag{15}$$

onde $\boldsymbol{\phi}_n = \Gamma_n^{-1}\gamma_n$ tal que $\boldsymbol{\gamma}_n(\gamma(1), \ldots, \gamma(n))^\top$ com erro quadrático médio dado por:

$$\nu_n = \mathbb{E}[X_{n+1} - \mathcal{P}_n X_{n+1}]^2 = \gamma(0) - \boldsymbol{\phi}_n^\top \gamma_n$$

Os coeficientes $\phi_{n1}, \ldots, \phi_{nn}$ podem ser computados recursivamente por meio dos passos:

1. $\phi_{nn} = [\gamma(n) - \sum_{j=1}^{n-1} \phi_{n-1,j}\gamma(n-j)]\nu_{n-1}^{-1}$.
2. $$\begin{bmatrix} \phi_{n1} \\ \vdots \\ \phi_{n,n-1} \end{bmatrix} = \begin{bmatrix} \phi_{n-1,1} \\ \vdots \\ \phi_{n-1,n-1} \end{bmatrix} - \phi_{nn} \begin{bmatrix} \phi_{n-1,n-1} \\ \vdots \\ \phi_{n-1,1} \end{bmatrix} \tag{16}$$

tal que $\nu_n = \nu_{n-1}[1 - \phi_{nn}^2]$ onde $\phi_{11} = \gamma(1)/\gamma(0)$ com $\nu_0 = \gamma(0)$.

Exemplo 2.8

Usando a série simulada encontre os coeficientes de previsão usando o Algoritmo de Durbin-Levison no R.

Outro algoritmo importante que pode ser aplicado a qualquer série temporal com segundo momento finito sendo estacionária ou não, é o **algoritmo de inovação**.

Suponha que $\{X_t\}$ é uma série com média zero e $\mathbb{E}(|X_t|^2) < \infty$ para cada t, seja $\mathbb{E}(X_iX_j) = \kappa(i,j)$ podemos introduzir a seguinte variável:

$$\hat{X}_n = \begin{cases} 0, \text{ se } n = 1 \\ \mathcal{P}_{n-1}X_n, \text{ se } n = 2, 3, \ldots \end{cases} \quad (17)$$

e $\kappa_n = \mathbb{E}[X_{n+1} - \mathcal{P}_nX_{n+1}]^2$.

UnB

As inovações ou erros de previsão de 1 passo são definidas como:

$$U_n = X_n - \hat{X}_n$$

Na forma matricial temos $\mathbf{U}_n = (U_1, \ldots, U_n)^\top$ e $\mathbf{X}_n = (X_1, \ldots, X_n)^\top$. Podemos então escrever:

$$\mathbf{U}_n = \mathbf{A}_n \mathbf{X}_n$$

onde $\mathbf{A}_n$ possui a forma:

$$\mathbf{A}_n = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ a_{11} & 1 & 0 & \dots & 0 \\ a_{22} & a_{21} & 1 & \dots & 0 \\ \vdots & \dots & \ddots & \vdots & \\ a_{n-1,n-1} & a_{n-1,n-2} & a_{n-1,n-3} & \dots & 1 \end{bmatrix} \quad (18)$$

Se $\{X_t\}$ é estacionário $a_{ij} = -a_j$ com a_j obtido por $\Gamma_n \mathbf{a}_n = \gamma_n(h)$ com $h = 1$.

Isso implica que $\mathbf{A}_n$ é não singular com inversa dada por:

$$\mathbf{C}_n = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ \theta_{11} & 1 & 0 & \dots & 0 \\ \theta_{22} & a_{21} & 1 & \dots & 0 \\ \vdots & \dots & \ddots & \vdots & \\ \theta_{n-1,n-1} & \theta_{n-1,n-2} & \theta_{n-1,n-3} & \dots & 1 \end{bmatrix} \tag{19}$$

O vetor de previsões de 1 passo $\hat{\mathbf{X}}_n = (X_1, \mathcal{P}_1 X_2, \dots, \mathcal{P}_{n-1} X_n)^\top$ pode ser expresso como:

$$\hat{\mathbf{X}}_n = \mathbf{X}_n - \mathbf{U}_n = \mathbf{C}_n \mathbf{U}_n - \mathbf{U}_n = \Theta_n (\mathbf{X}_n - \hat{\mathbf{X}}_n) \tag{20}$$

A equação 20 pode ser reescrita como:

$$\hat{X}_{n+1} = \begin{cases} 0 \text{ se } n = 0 \\ \sum_{j=1}^{n} \theta_{nj}(X_{n+1-j} - \hat{X}_{n+1-j}) \text{ para } n = 1, 2, \ldots \end{cases} \quad (21)$$

Para os quais as previsões 1 passo podem ser computadas como $\hat{X}_1, \hat{X}_2, \ldots$ recursivamente usando os coeficientes θ_{ij}. O Algoritmo é definido pelos passos:

1. $\nu_0 = \kappa(1, 1)$
2. $\theta_{n,n-k} = \nu_k^{-1}\left(\kappa(n+1, k+1) - \sum_{j=0}^{k-1} \theta_{k,k-j}\theta_{n,n-j}\nu_j\right)$ para $0 \leq k \leq n$

e $\nu_n = \kappa(n+1, n+1) - \sum_{j=0}^{n-1} \theta_{n,n-j}^2 \nu_j$

Exemplo 2.8

Considere uma série $\{X_t\}$ definida como:

$$X_t = Z_t + \theta Z_{t-1}$$

onde $\{Z_t\} \sim WN(0, \sigma^2)$. Vimos anteriormente que $\kappa(i,j) = 0$ para $|i-j| > 1$, $\kappa(i,i) = \sigma^2(1+\theta^2)$ e $\kappa(i,i+1) = \theta\sigma^2$. Aplique o algoritmo de inovação.

Para a previsão h passos a frente podemos usar o resultado:

$$\mathcal{P}_n(X_{n+k} - \mathcal{P}_{n+k-1}X_{n+k}) = 0 \text{ para } k \geq 1$$

Como $\mathbb{E}(\text{Erro} \times \text{Variável de previsão}) = 0$ e sabendo que:

$$\mathbb{E}[(X_{n+k} - \mathcal{P}X_{n+k} - 0)X_{n+j-1}] = 0 \text{ para } j = 1, \ldots, n$$

temos:

$$\mathcal{P}_n X_{n+h} = \mathcal{P}_n \mathcal{P}_{n+h-1} X_{n+h} = \mathcal{P}_n \hat{X}_{n+h}$$

o qual é igual a:

$$\mathcal{P}_n \left(\sum_{j=1}^{n+h-1} \theta_{n+h-1,j} (X_{n+h-j} - \hat{X}_{n+h-j}) \right)$$

Usando a linearidade do operador $\mathcal{P}_n$ temos:

$$\mathcal{P}_n X_{n+h} = \sum_{j=h}^{n+h-1} \theta_{n+h-1,j}(X_{n+h-j} - \hat{X}_{n+h-j})$$

onde os coeficientes θ_{nj} são determinados pelo algoritmo de inovação.

Cálculo recursico para a previsão h-passos.