Universidade do Estado do Rio de Janeiro
Faculdade de Ciências Econômicas

Resumo de econometria I (séries temporais)

Gabriel de Almeida Arruda

Disclaimer: Este trabalho foi feito com base nas aulas do Prof. Dr. Fernando Aiube, com as notas de aula do Kotze (2019) e com os livros: Box et al. (2015), Hamilton (1994) e **??**)

1 Introdução

Uma série de tempo é caracterizada por uma sequencia de observações tomada sequencialmente no tempo. Podemos chamar essa sequencia de observações $(y_1, y_2, y_3 ...)$ como sendo variáveis aleatórias, onde $y = f(t) + \varepsilon_t$, em outras palavras, cada observação é uma função do tempo mais um fator aleatório (ruido branco). Como esse y depende do tempo, vamos ter um valor diferente de y para cada valor de $t = 0, 1, 2, ..., T$; $y^* = f(t^*) + \varepsilon_{t^*}$. Então y_1 é o valor de y no período 1, y_2 é o valor de y no período 2 e assim vai. Uma série de tempo pode ser dividida entre um processo determinístico ou estocástico.

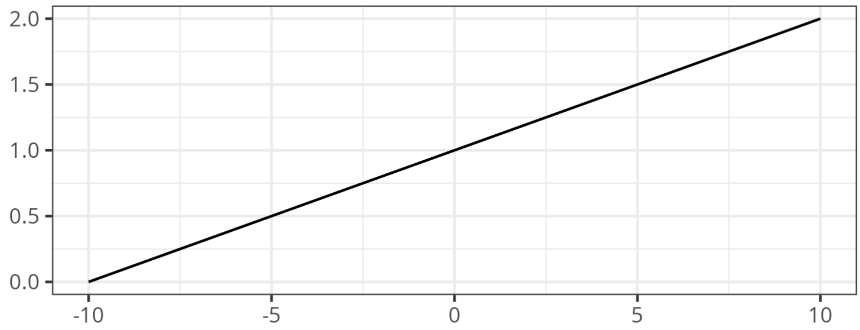
1.1 Processo determinístico

É um processo que não depende de um termo aleatório e sempre irá ter o mesmo resultado dado um valor inicial. O processo $T_t = 1 + 0,1t$ é um processo determinístico, pois não depende de nenhum fator aleatório, estando sempre acompanhado de uma constante (1) e um termo de tendencia determinística $(0, 1t)$.

Podemos observar essa função melhor, gerando o seguinte gráfico no **R**.

```
ggplot() +
  xlim(c(-10,10)) +
  geom_function(fun = function(x) {1+0.1*x}) +
  labs(y="")+
  theme_bw()
```

Figura 1 – Processo determinístico



1.2 Processo estocástico

O que diferencia um processo estocástico do determinístico é o fator aleatório, por mais que esse fator seja aleatório ele apresenta uma função de distribuição de probabilidade, por mais sabe-se qual é o valor inicial de uma série, não da para saber com exatidão o próximo valor, mas podemos atribuir probabilidades a diferentes valores.

1.2.1 Ruido Branco

Um ruido branco, é uma sequencia serial de variáveis aleatórias não correlacionadas com média zero e variância finita e constante, seria aquele erro (u) da primeira parte do curso. Assumimos duas condições importantes para o ruido branco: ele deve ser independente um do outro e deve apresentar uma distribuição normal[1]. Sendo escrito da seguinte forma

$$\varepsilon_t \sim i.i.d.\mathcal{N}(0,\ \sigma^2)$$

ou também

$$\varepsilon_t \sim RB(0,\ \sigma^2)$$

O ruido branco assume 3 propriedades propriedades:

- $\mathbb{E}[\varepsilon_t] = \mathbb{E}[\varepsilon_t \mid \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots] = 0$
- $\mathbb{E}[\varepsilon_t \varepsilon_{t-j}] = \text{cov}[\varepsilon_t \varepsilon_{t-j}] = 0$
- $\text{var}[\varepsilon_t] = \text{cov}[\varepsilon_t \varepsilon_t] = \sigma^2_{\varepsilon_t}$

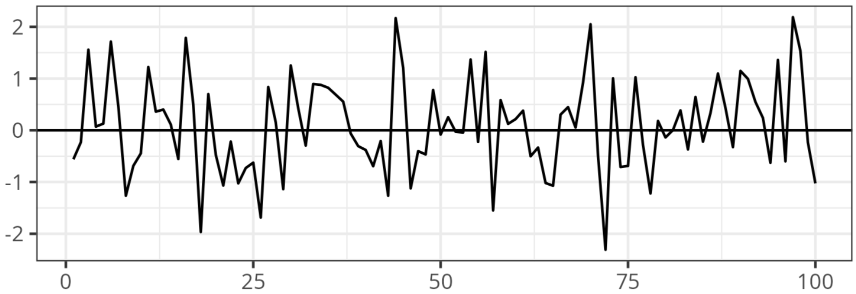
As duas primeiras propriedades diz respeito a impossibilidade de preditividade e a ausência de autocorrelação. A terceira diz respeito a homocedasticidade, a variância ser constante.

Para visualizar isso, criei uma função no **R** que cria um conjunto de dados aleatórios usando a função **rnorm**, onde **n** é o tamanho da amostra que queremos. A função **tibble** cria um data frame (matriz), onde a coluna **x** é uma sequencia de 1 até n e a coluna **y** recebe nosso conjunto de dados aleatórios.

[1] Chamada também de *Distribuição normal gaussiana.*

```r
ruido.branco <- function(n){
  set.seed(123)
  dados <- tibble(x = 1:n, y=rnorm(n, mean = 0, sd = 1))

  grafico <- ggplot(df, aes(x,y)) +
    geom_line() +
    geom_hline(yintercept = 0)+
    labs(y="", x="")+
    theme_bw()

  return(grafico)
}

ruido.branco(n = 100)
```

Figura 2 – Ruido Branco



1.2.2 Passeio aleatório

Esse tempo passeio aleatório se diz pelo falo de que o valor de uma variável em um determinado período é igual ao seu valor no período passado mais um fator aleatório determinado por ε_t, sendo descrito por

$$y_t = y_{t-1} + \varepsilon_t$$

Vamos supor um valor inicial para y_t como sendo y_1, então:

$$y_1 = y_0 + \varepsilon_1$$

$$y_2 = y_1 + \varepsilon_2$$

$$y_2 = y_0 + \varepsilon_1 + \varepsilon_2$$

resolvendo isso recursivamente, temos

$$y_t = y_0 + \sum_{i=1}^{t} \varepsilon_i$$

Se $y_0 \sim 0$, podemos então dizer que o passeio aleatório é uma soma acumulada de ruídos brancos $y_t = \sum_{i=1}^{t} \varepsilon_i$

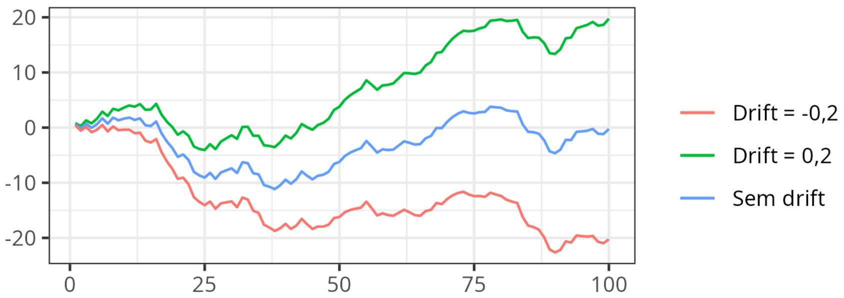
Pode acontecer o caso do passeio aleatório ter uma constante γ adicionada

$$y_1 = \gamma + y_0 + \varepsilon_1$$

logo quando $y_0 \sim 0$, vamos ter que

$$y_t = \gamma t + \sum_{i=1}^{t} \varepsilon_i$$

Figura 3 – Passeio Aleatório



1.3 Estacionariedade e Autocorrelação

Antes de entrar no assunto de autocorrelação, devemos ter em mente algumas propriedades referente ao primeiro momento (esperança) e segundo momento (variância) de um passeio aleatório. O primeiro momento de um passeio aleatório, pode ser calculado pela sua média

$$\begin{aligned}
\mathbb{E}[y_t] &= \mathbb{E}[y_0 + \sum_{i=1}^{t} \varepsilon_i] \\
&= \mathbb{E}[y_0] + \mathbb{E}[\sum_{i=1}^{t} \varepsilon_i] \\
&= \mathbb{E}[y_0] + 0 \\
&= y_0 = \mu
\end{aligned}$$

Nota-se que a esperança (média) de y_t não depende de t, sendo uma constante e é por isso que estou chamando de μ

O segundo momento é a variância

$$\begin{aligned}
\text{Var}[y_t] &= \mathbb{E}[y_t^2] = \mathbb{E}[(y_t - \mathbb{E}[y_t])^2] \\
&= \mathbb{E}[y_t^2] - \mathbb{E}[y_0]^2 \\
&= \mathbb{E}[(y_0 + \sum \varepsilon_i)^2] \\
&= \mathbb{E}[y_0^2 + 2y_0 \sum \varepsilon_i + \sum \varepsilon_i^2] - \mu^2 \\
&= \mathbb{E}[y_0^2] + 2y_0 \sum \mathbb{E}[\varepsilon_i] + \sum \mathbb{E}[\varepsilon_i^2] - \mu^2 \\
&= \mu^2 + 0 + \sum \sigma^2 - \mu^2 \\
&= t\sigma^2
\end{aligned}$$

Podemos observar que a variância de um processo aleatório não é constante, pois vai depender de t

Outra parada importante é definir a covariância para k lags

$$\begin{aligned}\text{Cov}[y_t, y_{t-k}] &= \mathbb{E}\{(y_t - \mathbb{E}[y_t])(y_{t-k} - \mathbb{E}[y_{t-k}])\} \\ &= \mathbb{E}[(y_t - \mu)(y_{t-k} - \mu)]\end{aligned}$$

Estacionariedade

A hipótese da estacionariedade é um caso particular do **processo estocástico**, no qual se assume que o processo está em um estado de equilíbrio. O caso dos processos estritamente estacionários (estacionariedade forte) é quando **todas** as suas propriedades (momentos) não são afetadas pelo tempo. Como essas condições são muito restritas, na grande maioria das vezes nos referimos a esse processo estocástico como **fracamente estacionário**, ou **covariância-estacionário**, ou **estacionário de segunda ordem**. Isso quer dizer que apenas a média e a variância deve ser constante e que a covariância deve depender apenas do numero de lags (k)

$$\begin{aligned}&\mathbb{E}[y_t] = \mu \\ &\text{Var}[y_t] = \mathbb{E}[(y_t - \mu)^2] = \sigma^2 \\ &\text{Cov}[y_t, y_{t+k}] = \mathbb{E}[(y_t - \mu)(y_{t-k} - \mu)] = \gamma_k\end{aligned}$$

Função de autocorrelação (FAC)

Ao assumir a hipótese da estacionaridade, vamos ter que a função de conjunta de probabilidade de y_{t_1} e y_{t_2} sera a mesma para todo o tempo t_1 e t_2, que será constante. Isso implica que a **covariância** entre y_t e y_{t+k} será separada apenas por k intervalos de tempo (lag). Assim, a *autocovariancia* ao lag k é definida por

$$\gamma_k = \text{Cov}[y_t, y_{t+k}] = \mathbb{E}[(y_t - \mu)(y_{t-k} - \mu)]$$

Da mesma forma, teremos que a **autocorrelação** é dada por

$$\begin{aligned}\rho_k &= \frac{\mathbb{E}[(y_t - \mu)(y_{t-k} - \mu)]}{\sqrt{\mathbb{E}[(y_t - \mu)^2]\,\mathbb{E}[(y_t - \mu)^2]}} \\ &= \frac{\mathbb{E}[(y_t - \mu)(y_{t-k} - \mu)]}{\sigma^2}\end{aligned}$$

como em um processo estacionário a variância é constante, vamos ter que $\gamma_0 = \sigma^2$. Podemos então dizer que a autocorrelação no lag k é

$$\rho_k = \frac{\gamma_k}{\gamma_0}$$

e isso implica que $\rho_0 = 1$ se $k = 0$.

A **Função de autocorrelação** é quando se plota um gráfico ρ_k contra o lag k

2 Modelos ARIMA

Antes de chegar no modelo ARIMA, vou passar pelo AR, dps o MA.

2.1 Modelos lineares estacionários

2.1.1 Modelo Autorregressivo (AR)

Um processo é dito autoregressivo quando ele carrega consigo msm uma certa persistência, quer dizer que o passado influencia o presente, quando apenas o período anterior influencia o atual vamos ter um AR(1)

$$y_t = \phi y_{t-1} + \varepsilon_t$$

onde ϕ é um coeficiente tbm chamado de **peso**, seria o peso da influencia de um período no outro. Um caso especial do processo *autoregressivo* é o AR(p)

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \varepsilon_t$$

onde todos os pesos ϕ são diferente de 0.

Podemos reescrever esse AR(p) usando o **operador lag**

$$(1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p) y_t = \varepsilon_t$$

ou então

$$\phi(L) y_t = \varepsilon_t$$

onde $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p$.

O processo *autoregressivo* deve respeitar algumas condições para ser estacionário. Seja o processo AR(1) para exemplificar

$$(1 - \phi L) y_t = \varepsilon_t$$

e para que esse processo seja estacionário, devemos ter que todas as raízes de $|\phi(L)| > 1$ e isso implica que o parâmetro ϕ do processo AR(1) deve ser $|\phi| < 1$ para ser estacionário. Como as raiz de $1 - \phi L = 0$ é $L = \phi^{-1}$, isso quer dizer que a raiz **está fora do circulo unitário**, já que $|\phi| < 1$.

O Mesmo raciocínio pode ser levado para o processo autorregressivo de ordens superiores, seja o caso do AR(p)

$$\phi(L) y_t = (1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p) y_t = \varepsilon_t$$

onde a condição de estacionaridade é que **TODAS AS RAÍZES** de $\phi(L) = 0$ deve ser menor que 1 em modulo, estar **fora do circulo unitário**, existe alguns testes para ver se uma série é estacionária ou não, como o do dick-fuller e KPSS, será abordado mais a frente.

Um processo autorregressivo é dito **inversível** quando podemos reescrever o **AR(1)** em forma de **MA(∞)**, isso é usado em alguns casos como o caso das funções de impulso e resposta que é usada quando queremos avaliar o quanto um choque em uma variável afeta a outra. Seja então o processo AR(1)

$$y_t = \phi y_{t-1} + \varepsilon_t$$

escrevendo na notação do operador lag

$$(1 - \phi L)y_t = \varepsilon_t$$
$$y_t = \frac{\varepsilon_t}{(1 - \phi L)}$$

podemos ver que essa expressão tem a cara de um somatório de PG infinita e a única condição para que isso seja verdade é a de que $\phi \leq 1$ para que ela seja convergente, podemos então dizer que no caso da **série ser estacionária podemos dizer que ela é inversível**, sendo ε_t o primeiro elemento e ϕL a razão. Então é a mesma coisa que dizer que isso é

$$y_t = \varepsilon_t + \phi L \varepsilon_t + \phi^2 L^2 \varepsilon_t + ...$$

ou ainda

$$y_t = \varepsilon_t + \phi \varepsilon_{t-1} + \phi^2 \varepsilon_{t-2} + ...$$
$$y_t = \sum_{i=0}^{\infty} \phi^i \varepsilon_{t-i}$$

Autocorrelação: De acordo com Box et al. (2015, p. 58), temos que o caso de um AR(1) satisfaz a seguinte condição

$$\rho_k = \phi_1 \rho_{k-1}$$

então para $k = 1$

$$\rho_1 = \phi_1$$

e resolvendo recursivamente

$$\rho_2 = \phi_1 \rho_1 = \phi_1^2$$
$$\rho_3 = \phi_1 \rho_2 = \phi_1 \phi_1^2 = \phi_1^3$$
$$\therefore \rho_k = \phi_1^k$$

Da mesma maneira, podemos ver que a autocorrelação de um **AR(2)** satisfaz a seguinte condição

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}$$

então para $k = 1$

$$\rho_1 = \phi_1 + \phi_2 \rho_{-1}$$

como a correlação é simétrica, podemos trocar $\rho_{-1} \to \rho_1$

$$\rho_1 = \phi_1 + \phi_2\rho_1$$

$$\rho_1 = \frac{\phi_1}{1-\phi_2}$$

e resolvendo recursivamente

$$\rho_2 = \phi_1\rho_1 + \phi_2\rho_0$$

$$\rho_2 = \frac{\phi_1^2}{1-\phi_2} + \phi_2$$

$$\rho_3 = \phi_1\rho_2 + \phi_2\rho_1$$

$$\rho_3 = \phi_1\left[\frac{\phi_1^2}{1-\phi_2} + \phi_2\right] + \frac{\phi_1\phi_2}{1-\phi_2}$$

$$\vdots$$

*** terminar essa parte mais tarde****

Variância: Uma maneira fácil de se calcular a variância de um processo AR(1) é usando o operador da variância **Var[.]**:

$$\begin{aligned}
\text{Var}[y_t] &= \text{Var}[\phi y_{t-1}] + \text{Var}[\varepsilon_t] \\
\text{Var}[y_t] &= \phi^2\text{Var}[y_t] + \text{Var}[\varepsilon_t] \\
\text{Var}[y_t] &= \phi^2\text{Var}[y_t] + \sigma_\varepsilon^2 \\
(1-\phi^2)\text{Var}[y_t] &= \sigma_\varepsilon^2 \\
\sigma_y^2 &= \frac{\sigma_\varepsilon^2}{(1-\phi^2)}
\end{aligned}$$

se vc observar o valor de $\phi > 1$ a variância é negativa, oq seria um absurdo; se $\phi = 1$ a variância vai para infinito e é por isso que quando o processo é não estacionário ele apresenta variância infinita.

EXEMPLO: Seja a modelo $y_t = -0,3y_{t-1} + 0,6y_{t-2} + \varepsilon_t$, verifique se é estacionário.

Podemos reescrever esse modelo usando o operador lag

$$(1 + 0,3L - 0,6L^2)y_t = \varepsilon_t$$

temos então aqui nesse exemplo que $\phi(L) = 1 + 0,3L - 0,6L^2$ e como a condição para que o processo seja estacionário as raízes de $\phi(L) = 0$ tem que ser maior que 1 em módulo. Vamos ter que

$$\begin{aligned}
L &= \frac{-0,3 \pm \sqrt{0,3^2 - 4(-0,6)}}{2(-0,6)} \\
&= \frac{-0,3 \pm 1,578}{-1,2} \;\Rightarrow\; L = -1,065 \text{ e } 1,565
\end{aligned}$$

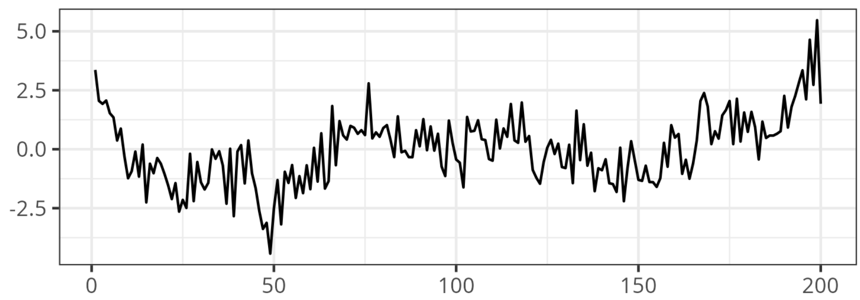
sendo todas as raízes maiores que um em modulo, então o processo é estacionário.

Simulando no R

Podemos simular um processo autorregressivo no R da seguinte maneira usando o modelo do exemplo anterior da seguinte maneira

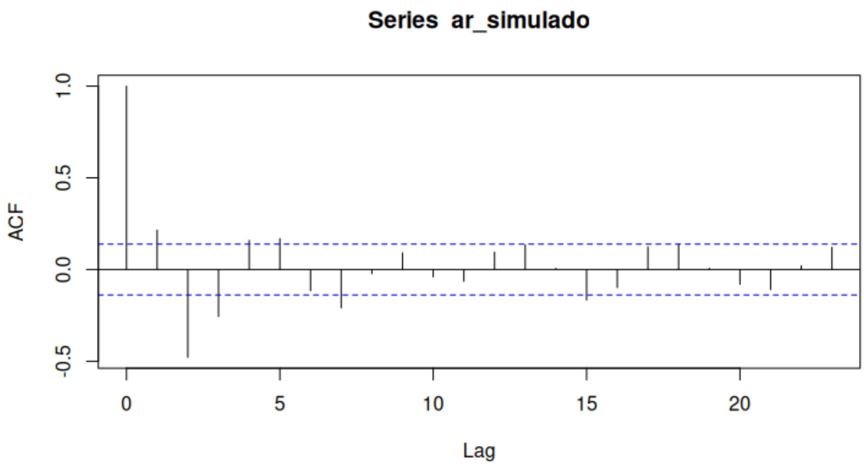
```
set.seed(123)

ar_simulado <- arima.sim(model = list(order = c(2, 0, 0),
                         ar = c(0.3, -0.6)), n = 200)

plot(ar_simulado)
```

Figura 4 – AR(2)



E podemos usar a função *acf()* do próprio **R** para rodar o comando **acf(ar-simulado)** para plotar a função de autocorrelação

Figura 5 – Autocorrelação



2.1.2 Modelo de Média móvel (MA)

O modelo de média móvel é um extensão do processo de ruido branco, aqui os parâmetros de peso é o simbolo θ. O modelo é escrito da seguinte forma

$$y_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + ... + \theta_q\varepsilon_{t-q}$$

O motivo pelo qual se usa esse tipo de modelo, é pq se parar para pensar em economia estamos diante de diversos choques ao longo do tempo, seja por uma greve, desastres naturais, decisões do governo, etc... E esses choque normalmente não terão apenas um efeito imediato na série, muita das vezes apresentam um efeito contemporâneo, demorando a ser dissipado e tendo efeito ao longo do tempo em vários períodos consecutivos.

Seja o modelo de MA(q) com $\mu = 0$ com esperança

$$\begin{aligned}\mathbb{E}[y_t] &= \mathbb{E}[\varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + ... + \theta_q\varepsilon_{t-q}] \\ &= 0\end{aligned}$$

e variância

$$\begin{aligned}\text{Var}[y_t] &= \text{Var}[\varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + ... + \theta_q\varepsilon_{t-q}] \\ &= \text{Var}[\varepsilon_t] + \text{Var}[\theta_1\varepsilon_{t-1}] + \text{Var}[\theta_2\varepsilon_{t-2}] + ... + \text{Var}[\theta_q\varepsilon_{t-q}] \\ &= \sigma_\varepsilon^2 + \theta_1^2\sigma_\varepsilon^2 + \theta_2^2\sigma_\varepsilon^2 + ... + \theta_q^2\sigma_\varepsilon^2 \\ &= (1 + \theta_1^2 + \theta_2^2 + ... + \theta_q^2)\sigma_\varepsilon^2\end{aligned}$$

nota-se então que a variância de um modelo MA(2) é igual a $\text{Var}[y_t] = (1+\theta_1^2+\theta_2^2)\sigma_\varepsilon^2$

Invertibilidade: As condições de invertibilidade do modelo é igual as condições de um modelo AR(p) ser estacionário, isto quer dizer que $|\theta| < 1$ ou então que as raízes de $\theta(L) = 0$ sejam maiores que 1 em módulo

Função de autocorrelação: Vamos ter a seguinte formula para o caso geral de um modelo MA(q),

$$\rho_k = \begin{cases} \dfrac{-\theta_k + \theta_1\theta_{k+1} + \cdots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \cdots + \theta_q^2} & k = 1, 2, \ldots, q \\ 0 & k > q \end{cases}$$

onde temos que $\gamma_0 = (1 + \theta_1^2 + \cdots + \theta_q^2)\sigma_\varepsilon^2$ e que $\gamma_k = (-\theta_k + \theta_1\theta_{k+1} + \cdots + \theta_{q-k}\theta_q)\sigma_\varepsilon^2$

Uma coisa a se notar nessa equação é que quando tivermos que o numero de lags da função de autocorrelação for maior que a ordem do modelo, a FAC será igual a zero. Seja um modelo MA(2), então $\rho_3, ..., \rho_k = 0$.

Essa equação pode ser aplicada para modelos MA(1) também

$$\rho_k = \begin{cases} \dfrac{-\theta_1}{1 + \theta_1^2} & k = 1 \\ 0 & k > 1 \end{cases}$$

de forma análoga, para um MA(2) vamos ter

$$\rho_1 = \frac{-\theta_1 + \theta_1\theta_2}{1 + \theta_1^2 + \theta_2^2}$$

$$\rho_2 = \frac{-\theta_2}{1 + \theta_1^2 + \theta_2^2}$$

2.2 Modelos lineares não-estacionários

A maioria das séries de tempo apresenta não estacionariedade isto quer dizer que em um modelo ARMA, se as raízes de $\phi(L) = 0$ estiver fora do circulo unitário ela é dita estacionaria, mas quando estão fora do circulo unitário ela terá um comportamento explosivo, sendo não estacionaria. Um modelo ARMA sem sazonalidade pode ser escrito da seguinte forma

$$\phi(L)y_t = \theta(L)\varepsilon_t$$

se $\phi(L)$ for um operador autorregressivo não estacionário, tal que d raízes de $\phi(L) = 0$ são unitárias e o resto está fora do circulo unitário, esse modelo pode ser escrito como

$$\phi(L)(1 - L)^d y_t = \theta(L)\varepsilon_t$$

onde $\phi(L)$ se torna um operador autorregressivo estacionário, no caso aqui foi feita a separação de todas as raízes que estão fora do circulo unitário, das raízes que estão dentro do circulo unitário. Sabendo que a diferenciação torna as raízes estacionaria, podemos usar o operador $\nabla^d = (1 - L)^d$ para $d \geq 1$; aqui ∇ é chamado de operador de diferenciação.

$$\phi(L)\nabla^d y_t = \theta(L)\varepsilon_t$$

esse processo sera conhecido como integração, o I do ARIMA(p, d, q), onde a ordem de integração desse processo é dado pelo valor de d. E se chamarmos $w_t = \nabla^d y_t$, o processo passa a ser escrito da seguinte forma

$$\phi(L)w_t = \theta(L)\varepsilon_t$$

Abaixo podemos ver alguns exemplos de processo ARIMA:

1. O modelo ARIMA(0,1,1):

$$\begin{aligned} \nabla y_t &= \varepsilon_t - \theta_1\varepsilon_{t-1} \\ &= (1 - \theta_1 L)\varepsilon \end{aligned}$$

onde $p = 0,\ d = 1,\ q = 1,\ \phi(L) = 1, \theta(L) = (1 - \theta_1)\varepsilon$

2. O modelo ARIMA(1,1,1):

$$\nabla y_t - \phi_1 \nabla y_{t-1} = \varepsilon_t - \theta_1 \varepsilon_{t-1}$$
$$(1 - \phi_1 L)\nabla y_t = (1 - \theta_1 L)\varepsilon$$

onde $p = 1,\ d = 1,\ q = 1,\ \phi(L) = (1 - \phi_1 L)\nabla y_t,\ \theta(L) = (1 - \theta_1 L)\varepsilon$

EXEMPLO: Seja a modelo $y_t = 0,1y_{t-1} + 0,9y_{t-2} + \varepsilon_t$, verifique se é estacionário.

Reescrevendo usando o operador lag

$$(1 - 0,1L - 0,9L^2)y_t = \varepsilon_t$$

uma dica aqui é sempre isolar o que estiver ao quadrado tirando do parenteses

$$(0,9)(1,1111 - 0,1111L - L^2)y_t = \varepsilon_t$$

usando a formula quadrática, vamos encontrar que as raízes de $\phi(L)$ são -1,1111 e 1; como uma das raízes é igual a 1, esse processo é não estacionário. Reescrevendo, temos

$$(0,9)(L + 1,1111)(L - 1)y_t = \varepsilon_t$$
$$(1 + 0,9L)(L - 1)y_t = \varepsilon_t$$
$$(1 + 0,9L)(1 - L)y_t = \varepsilon_t$$
$$(1 + 0,9L)\nabla y_t = \varepsilon_t$$

assim sendo um modelo ARIMA(1,1,1) pois temos ∇y_t que corresponde a integração de ordem 1 (I(1)). Também pode ser reescrito da seguinte maneira

$$()1 + 0,9L)w_t = \varepsilon_t$$

mas note, que como aqui não temos o operador de diferenciação, este modelo é um ARMA(1,1).

3 Identificação do Modelo

Uma das formas de se identificar qual o melhor modelo usar é por meio dos critérios de informação de Akaike (AIC) ou o critério de informação Schwartz (*Bayesian information Criterion* - BIC). A ideia por trás do critério de informação é tentar mensurar a qualidade de um modelo de acordo com a informação que ele carrega e também eles presam pela simplicidade, quer dizer que quanto mais parâmetros e mais lags for adicionado, mais penalizado será o modelo de acordo com o critério de informação prezando sempre por um modelo mais parcimonioso.

O ideal é que o AIC e BIC sejam menor possível (vale notar que pode ser negativo). Quanto melhor for o ajuste do modelo, mais AIC e BIC irá se aproximar de $-\infty$. O critério de informação pode ser usado para comparar dois modelos; Um modelo A pode ser dito melhor que o B, se o critério de informação (AIC ou BIC) do A for menor que do B, lembrando que para essa comparação, o ideal é que os modelos tenham o mesmo tamanho de amostra. São calculados da seguinte maneira

$$\text{AIC} = \frac{-2ln(L)}{T} + \frac{2n}{T}$$
$$\text{BIC} = \frac{-2ln(L)}{T} + \frac{nln(T)}{T}$$

onde $n = p+q+1$ que é soma dos parâmetros do modelo; T é o tamanho da amostra; L é o valor máximo da função de máxima verossimilhança.

Note que a diferença entre eles é que no BIC vamos ter $ln(T)$ ao invés do 2 multiplicando o n e como $ln(T)$ sempre será maior que 2, quer dizer que o BIC é mais parcimonioso que o AIC; o custo marginal de se adicionar regressores é maior do BIC do que no AIC.

4 Testes do modelo

4.1 Testes de raiz unitária - Dick-Fuller

O teste de Dick-fuller é o teste mais usado para detectar a presença de raiz unitária, na qual a hipótese nula é se a série é um passeio aleatório, contra a hipótese alternativa se a série é estacionária. Para mostrar o teste, vamos usar um AR(1),

$$y_t = \phi y_{t-1} + \varepsilon_t, \quad \text{onde} \quad \varepsilon_t \sim i.i.d.\mathcal{N}(0,\, \sigma^2)$$

onde já sabemos que se $\phi = 1$ o processo é um passeio aleatório e se $\phi < 1$ ele é um processo estacionário. Aqui vamos temos que esse ϕ é um parâmetro estimado por MQO, por isso precisamos aplicar um teste a esse parâmetro para saber o quão confiável é esse coeficiente que foi estimado, para isso, Dick-Fuller aplicaram um teste de regressão que é derivado desse modelo AR(1), subtraindo y_{t-1} de ambos os lados

$$y_t - y_{t-1} = \phi y_{t-1} - y_{t-1} + \varepsilon_t$$
$$\nabla y_t = (\phi - 1)y_{t-1} + \varepsilon_t$$

substituindo $\delta = \phi - 1$, temos

$$\nabla y_t = \delta y_{t-1} + \varepsilon_t$$

Usando essa equação, pode-se investigar que se $\phi = 1 \Rightarrow \delta = 0$ e isso implica que $\nabla y_t = \varepsilon_t$ sendo um passeio aleatório; caso contrário, se $\phi < 1 \Rightarrow \delta \neq 0$ sendo um processo estacionário. Então a ideia desse teste de DF é olhar para o γ, onde

$$H_0 \, : \, \delta = 0$$

se a hipótese nula não for rejeitada (satisfeita ou aceita), vamos ter que y_t é integrado de ordem 1, tal que $y_t \sim I(1)$, pois será estacionário após uma defasagem. A hipótese alternativa tem a seguinte forma

$$H_a \ : \ \delta \neq 0$$

então se a hipótese nula for rejeitada, implica que $y_t \sim I(1)$, sendo a série estacionária. O interessante aqui é que com essa regressão de teste, conseguimos aplicar uma estatística de teste (τ) em δ, podendo assim comparar essa estatística do modelo com valores já calculados por Dick-Fuller, chamados de τ_c (tau calculado), assim

- Se $\tau < \tau_c \Rightarrow$ rejeita H_0 (não estacionário)
- Se $\tau > \tau_c \Rightarrow$ **NÃO** rejeita H_0 (estacionário)

esse valor de teste é um simples teste t, onde

$$\hat{t}_{DF} = \frac{\hat{\delta}}{SE(\hat{\delta})} = \frac{\phi - 1}{SE(\phi)}$$

SE corresponde ao *standard error* (erro padrão)

Esse teste de DF pode ser modificado para caso a série tenha um drift, adicionando uma constante e um termo de tendencia, assumindo a seguinte forma

$$\nabla y_t = \beta_1 + \beta_2 t + \delta y_{t-1} + \varepsilon_t$$

a hipótese nula permanece a mesma, para saber se essa constante e o termo de tendencia são significantes, temos que olhar a estatística de teste t de cada beta, ou olhar para seus *p-valor*.

Esse teste de Dick-Fuller é bem simples e não permite testar a persistência do processo, pois os resíduos podem ser autocorrelacionados. Assim, foi desenvolvido o teste de **Dick-Fuller Aumentado** (ADF) para controlar essa autocorrelação nos resíduos e possibilitando que se teste valores de lag maiores (AR(.) de ordem superior). Vamos considerar a regressão de teste para um AR(1)

$$\nabla y_t = \beta_1 + \beta_2 t + \delta y_{t-1} + \varepsilon_t$$

se a gente achar que se ∇y_t segue um AR(2), a regressão de teste será

$$\nabla y_t = \beta_1 + \beta_2 t + \delta y_{t-1} + \gamma \nabla y_{t-1} + \varepsilon_t$$

generalizando isso para um processo AR(p), temos

$$\nabla y_t = \beta_1 + \beta_2 t + \delta y_{t-1} + \sum_{i=0}^{m} \gamma i \nabla y_{t-i} + \varepsilon_t$$

aqui a hipótese é a mesma, sendo $H_0 : \delta = 0$ não estacionário e $H_a : \delta < 0$ estacionário. A significância dos γ pode ser testada através de um teste t.

4.2 Teste de autocorrelação - **Ljung–Box** e **Box-Pierce**

o teste de Ljung–Box pode ser definido como

- H_0 : os erros são independentemente distribuídos
- H_A : os erros **NÃO** são independentemente distribuídos; existe correlação serial

Não tem muito mais oq comentar nesse teste, apenas mostrar que a estatística de teste do Ljung–Box é

$$LB(k) = N(N+2)\sum_{j=1}^{k}(N-j)^{-1}\hat{\rho}_j^2(\hat{\varepsilon})$$

e que o teste Q (Box-Pierce) é

$$Q(k) = N\sum_{j=1}^{k}\hat{\rho}_j^2(\hat{\varepsilon})$$

e que ρ_j é a correlação dos resíduos ao lag j e N é o tamanho da amostra. Como aqui as estatísticas de teste se distribuem como uma qui-quadrado com $k-p-q$. Rejeita-se H_0 se

$$Q, LB > \tau$$

sendo $\tau = \chi^2_\alpha(k-p-q)$, em que α é o nível de significância.

4.3 Teste de heterocedasticidade - **ARCH-LM**

O teste **ARCH lagrange multiplier** ou **ARCH-LM** é usado para detectar heterocedasticidade e é bem simples na verdade, ele teste se os erros da regressão são variantes no tempo sem de fato estimar um parâmetro *ARCH*. A estimação desse teste pode seguir a seguinte metodologia em dois passos

1. Estimar via MQO a equação da regressão mais adequada ou usar um modelo ARMA e assumir que $\hat{\varepsilon}_t^2$ seja o erro quadrático ajustado
2. Regredir os erros quadráticos contra uma constante e com q lags desses valores $\hat{\varepsilon}_{t-1}^2, \hat{\varepsilon}_{t-2}^2, ..., \hat{\varepsilon}_{t-q}^2$, estimando a seguinte regressão

$$\hat{\varepsilon}_t^2 = \alpha_0 + \alpha_1\hat{\varepsilon}_{t-1}^2 + \alpha_2\hat{\varepsilon}_{t-2}^2 + ... + \alpha_q\hat{\varepsilon}_{t-q}^2$$

Se não tiver efeito ARCH ou GARCH, os valores estimados de α_1 **até** α_q **tem que ser zero**. Essa regressão vai apresentar pouco poder explicatório, isso quer dizer que o R^2 será pequeno. Usando um tamanho de amostra T, a estatística de teste TR^2 (vale ressaltar que isso é um produto) converge para uma distribuição χ^2 com q graus de liberdade. Por isso vamos ter que tipo de teste entra na categoria de teste assintóticos, que é quando depende do tamanho da amostra, se a amostra for muito pequena o valor do teste (TR^2) será superestimado, mas para amostras grandes ele funciona bem. A hipótese nula aqui é $H_0 : \varepsilon_t$ sem efeito ARCH (**sem heterocedasticidade**); $H_a : \varepsilon_t$ tem efeito ARCH (**tem heterocedasticidade**). A hipótese nula é rejeitada se $TR^2 > \tau$

4.4 Teste de normalidade - **Jarque-Bera**

Dificilmente dados empíricos passam no teste de normalidade dos dados, o mais famoso teste para testar se os dados estão normalmente distribuídos é o teste de **Jarque-Bera**. *** CONTINUAR ***

5 Modelos de volatilidade

5.1 Modelos lineares

5.1.1 ARCH

Seja o modelo

$$y_t = \mathbb{E}[y_t \,|\, I_{t-1}] + \nu_t$$

onde $\nu_t \sim i.i.d\,(0,\, h_t)$ representa o erro do modelo, com média zero e uma variância h_t que é condicional ao conjunto de informações passadas $I_{t-1} = \{y_1, x_1, ..., y_{t-1}, x_{t-1}\}$ e ν_t pode ser separado em duas partes, uma é a parte estocástica $\varepsilon_t \sim i.i.d\,(0,\, 1)$ e a outra é o desvio padrão condicional que depende do tempo $h^{\frac{1}{2}}$. Logo, a variância de ν_t é

$$\mathbb{E}[\nu_t^2 \mid I_{t-1}] = h_t$$

Assumindo que a média é 0, podemos escrever o modelo ARCH básico da seguinte forma

$$y_t = h_t^{\frac{1}{2}} \varepsilon_t$$

como h é a variância condicional ao conjunto de informação, ele é definido como

$$h_t = w + \alpha y_{t-1}^2$$

onde ele depende de uma constante e do quadrado da do que aconteceu no passado.

Note que, como está ao quadrado a informação passada, o modelo ARCH que é da classe dos lineares, **atribuem peso igual para volatilidade tanto negativa**, quanto positiva.A ideia central desse tipo de modelo, é conseguir separar a volatilidade condicional do ruido branco e modelar essa volatilidade condicional h_t

Modelo AR-ARCH: Como aqui nos modelos de volatilidade nós estamos considerando o y_t como uma série de retornos, pode acontecer de que os retornos sejam correlacionados, ainda que fracamente, sendo então adicionado um AR(1) ao modelo

$$y_t = \phi y_{t-1} h_t^{\frac{1}{2}} \varepsilon_t$$
$$h_t = w + \alpha y_{t-1}^2$$

em que ϕ mede a correlação entre os retornos no tempo t e $t-1$

5.1.2 GARCH

Esse modelo é uma extensão do modelo ARCH, aqui ele inclui dentro da variância sua própria defasagem. Podemos escrever o modelo GARCH(1,1) da seguinte maneira

$$
\begin{aligned}
y_t &= \phi y_{t-1} h_t^{\frac{1}{2}} \varepsilon_t \\
h_t &= w + \alpha y_{t-1}^2 + \beta h_{t-1}
\end{aligned}
$$

e um modelo GARCH(p,q) é dado por

$$h_t = w + \sum_{i=1}^{q} \alpha_i y_{t-i}^2 + \sum_{i=1}^{p} \beta h_{t-p}$$

Referências

BOX, G. E. et al. **Time series analysis: forecasting and control**. [S.l.]: John Wiley & Sons, 2015.

ENDERS, W. **Applied Econometric Times Series**. Wiley, 2014. (Wiley Series in Probability and Statistics). ISBN 9781118918616. Disponível em: <https://books.google.com.br/books?id=lmr9oQEACAAJ>.

HAMILTON, J. D. **Time series analysis**. [S.l.]: Princeton university press, 1994.

KOTZE, K. **Time Series Analusis**. 2019. Disponível em: <hhttps://www.economodel.com/time-series-analysis-2019>.