

EAE1223: ECONOMETRIA III

AULA 5 - METODOLOGIA DE BOX-JENKINS

Luis A. F. Alvarez

10 de abril de 2024

A METODOLOGIA

- A metodologia de Box-Jenkins consiste numa série de etapas para estimar um modelo **univariado** de **previsão**.
 - Ideia é estimar um modelo simples, embora flexível, aos dados.
- Trata-se da metodologia básica de previsão em séries de tempo.
 - Diversas metodologias modernas incorporam o “espírito” de Box-Jenkins.
 - *Benchmark* para avaliar outros modelos de previsão.

ETAPAS DA METODOLOGIA

- A metodologia consiste de quatro etapas:
 1. **Identificação**: nessa etapa, avaliamos os dados e identificamos quais modelos são candidatos plausíveis para reproduzir os dados.
 2. **Estimação**: nessa etapa, estimamos os modelos candidatos.
 3. **Diagnóstico**: nessa etapa, avaliamos quais dos modelos se saíram melhor, de acordo com alguns critérios.
 4. **Previsão**: por fim, realizamos a previsão de acordo com nosso modelo.
- Nesta aula, discutiremos cada uma dessas etapas.
- Começaremos revisando a classe de modelos estudadas na metodologia de Box-Jenkins.

Modelos considerados

MA(q)

- Dizemos que uma série de tempo $\{Y_t\}$ segue um MA(q) se:

$$Y_t = \alpha + \epsilon_t + \sum_{j=1}^q \theta_j \epsilon_{t-j}$$

onde $\{\epsilon_t\}$ é um ruído branco.

- Série hoje depende diretamente da realização atual e das últimas q realizações de um ruído branco.
- **Todo** processo de média móvel é fracamente estacionário.
 - De fato, $\mathbb{E}[Y_t] = \alpha$, $\mathbb{V}[Y_t] = \left(1 + \sum_{j=1}^q \theta_j^2\right) \sigma_\epsilon^2$,
 $\text{cov}(Y_t, Y_{t-s}) = (\theta_s + \sum_{j=s+1}^q \theta_j \theta_{j-s}) \sigma_\epsilon^2$ se $s \leq q$ e 0 do contrário.
Correlação morre após q períodos.
- Um processo MA(q) é dito **invertível** se pode ser escrito como:

$$Y_t = \omega + \sum_{j=1}^{\infty} \psi_j Y_{t-j} + \epsilon_t$$

(MA(q) pode ser representado como AR(∞)).

AR(p) ESTACIONÁRIO

- Dizemos que uma série de tempo $\{Y_t\}$ segue um AR(p) estacionário se ela se escreve como:

$$Y_t = \alpha + \sum_{j=1}^p \beta_j Y_{t-j} + \epsilon_t$$

onde $\{\epsilon_t\}$ é um ruído branco e os coeficientes β_j são tais que o processo resultante é fracamente estacionário.

- Série hoje depende diretamente das realizações passadas nos últimos p períodos, mais um ruído branco.
 - Mas persistência não é tão grande, de modo que a série é estacionária (não há raiz unitária).
- Recorde-se que um AR(p) é estacionário se, e somente se, ele se escreve como um MA(∞):

$$Y_t = \kappa + \epsilon_t + \sum_{j=1}^{\infty} \tau_j \epsilon_{t-j}$$

ARMA(p,q) ESTACIONÁRIO

- Dizemos que uma série de tempo $\{Y_t\}$ segue um ARMA(p,q) estacionário se:

$$Y_t = \alpha + \sum_{i=1}^p \beta_i Y_{t-i} + \epsilon_t + \sum_{j=1}^q \theta_j \epsilon_{t-j}$$

onde $\{\epsilon_t\}$ é um ruído branco, e os β_i são tais que o processo resultante é estacionário.

- Combinação dos dois modelos anteriores.
- Metodologia de Box-Jenkins visará a estimar modelos na classe ARMA(p,q).

E SE OS DADOS FOREM NÃO ESTACIONÁRIOS?

- Até agora, discutimos a modelagem supondo as séries estacionárias.
 - Como fazer a previsão em casos não estacionários?
- Se as séries apresentarem uma tendência estocástica, trabalhamos com os dados em primeira diferença $\{\Delta Y_t\}$.
- Conduzidas as etapas da metodologia Box-Jenkins com os dados diferenciados, e encontradas projeções para ΔY_t fora da amostra, recompomos as projeções em nível usando o fato de que $Y_{t+1} = Y_t + \Delta Y_{t+1}$.
 - Isto é, se temos T observações, projetamos $\widehat{Y}_{T+1} = Y_T + \widehat{\Delta Y}_{T+1}$, $Y_{T+2} = Y_T + \widehat{\Delta Y}_{T+1} + \widehat{\Delta Y}_{T+2}$, e assim por diante.

MODELOS ARIMA(P,D,Q)

- Em outras palavras, no caso de raiz unitária, a metodologia irá estimar um modelo $\text{ARIMA}(p,1,q)$, da forma:

$$\Delta Y_t = \alpha + \sum_{i=1}^p \beta_i \Delta Y_{t-i} + \epsilon_t + \sum_{j=1}^q \theta_j \epsilon_{t-j}$$

- De modo geral, se a série é $I(d)$, a modelagem considerará modelos $\text{ARIMA}(p,d,q)$:

$$\Phi(L)(1-L)^d Y_t = \alpha + \Psi(L)\epsilon_t,$$

para polinômios $\Phi(L)$ e $\Psi(L)$ de grau p e q , respectivamente.

PREVISÃO COM TENDÊNCIA DETERMINÍSTICA

- Se séries apresentam tendência determinística, conduzimos a metodologia com dados *detrended*. Calculadas as projeções *detrended*, recompomos projeções em nível usando a tendência estimada.
- Por exemplo, se estimamos uma tendência linear:

$$Y_t = \tilde{a} + \tilde{b}t + \tilde{\xi}_t$$

e ajustamos um modelo ARMA para $\tilde{\xi}_t$, a projeção para fora da amostra é:

$$\widehat{Y_{T+h}} = \tilde{a} + \tilde{b}(T+h) + \widehat{\tilde{\xi}_{T+h}},$$

onde $\widehat{\tilde{\xi}_{T+h}}$ é a projeção do ARMA para $T+h$ (veremos como calculá-la na etapa de previsão).

- Estimação de ARMA para dado *detrended* é equivalente a estimar um modelo ARMA(p,q) com tendência determinística.

$$\Phi(L)Y_t = \alpha + \gamma t + \Psi(L)\epsilon_t.$$

“FILOSOFIA” DA METODOLOGIA DE BOX-JENKINS

- A restrição a modelos ARMA(p,q) pode ser entendida a partir do **teorema de decomposição de Wold**.
- Segundo esse teorema, **qualquer processo fracamente estacionário** pode ser representado pela soma de um MA(∞), acrescido de uma função determinística dos Y no passados:

$$Y_t = \epsilon_t + \sum_{l=0}^{\infty} \psi_l \epsilon_{t-l} + \kappa_t,$$

onde κ_t é função “aproximadamente” linear de Y_{t-1} , Y_{t-2} ...

- Ideia de Box-Jenkins é aproximar essa representação por um ARMA(p,q), com p e q **pequenos**.
 - Ideia é que aproximação **parcimoniosa**, por ser menos ruidosa, tende a funcionar melhor que modelos muito complexos.

Identificação

IDENTIFICAÇÃO DE UM ARMA(p,q)

- A etapa de identificação da metodologia Box-Jenkins consiste em encontrar quais modelos da classe ARMA(p,q) melhor caracterizam a série de interesse.
- A identificação consiste em analisar a função de autocorrelação (FAC) e a função de autocorrelação parcial (FACP) da série estacionária.

FUNÇÃO DE AUTOCORRELAÇÃO SERIAL (FAC)

- A **função de autocorrelação (FAC)** de uma série $\{Y_t\}_t$ estacionária é o mapa que associa, a cada número $k > 0$, a autocorrelação de ordem k , i.e. $\gamma_k = \text{cor}(Y_t, Y_{t-k})$.
- Note que essa função está bem definida para processos estacionários, visto que $\text{cor}(Y_t, Y_{t-k})$ não depende de t .

$$\text{cor}(Y_t, Y_{t-k}) = \frac{\text{cov}(Y_t, Y_{t-k})}{\text{sd}(Y_t) \text{sd}(Y_{t-k})} = \frac{\text{cov}(Y_t, Y_{t-k})}{\mathbb{V}(Y_t)}$$

FUNÇÃO DE AUTOCORRELAÇÃO SERIAL PARCIAL (FACP)

- A função de autocorrelação parcial (FACP) de uma série $\{Y_t\}_t$ estacionária é o mapa que associa, a cada número $k > 0$, a correlação θ_k entre Y_t e Y_{t-k} , controlando por $Y_{t-1}, Y_{t-2}, \dots, Y_{t-k+1}$.
- A autocorrelação parcial de ordem k (θ_k) é dada pelo coeficiente τ_k associado a Y_{t-k} no modelo preditivo linear:

$$\begin{aligned} Y_t &= \tau_0 + \tau_1 Y_{t-1} + \tau_2 Y_{t-2} \dots + \tau_k Y_{t-k} + \nu_t \\ \mathbb{E}[\nu_t] &= 0, \mathbb{E}[\nu_t Y_{t-j}] = 0, \quad j = 1, \dots, k \end{aligned} \tag{1}$$

- Note que função está bem definida para processos estacionários, visto que coeficientes do melhor preditor linear em (1) não dependem de t .

FAC E FACP ESTIMADAS

- Na prática, não observamos a FAC nem a FACP de um processo, mas podemos estimá-las usando as realizações da série de interesse.
 - Estimamos a FAC calculando as autocorrelações nos dados.

$$\hat{\gamma}_k = \frac{\sum_{t=1}^{T-k} (Y_t - \bar{y})(Y_{t+k} - \bar{y})}{\sum_{t=1}^T (Y_t - \bar{y})^2},$$

onde $\bar{y} = T^{-1} \sum_{t=1}^T Y_t$.

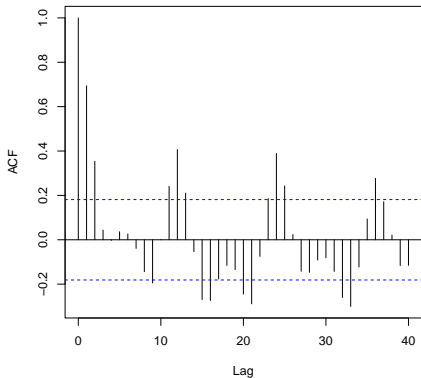
- Estimamos a FACP ajustando o modelo (1) aos dados.

INFERÊNCIA SOBRE FAC E FACP POPULACIONAIS

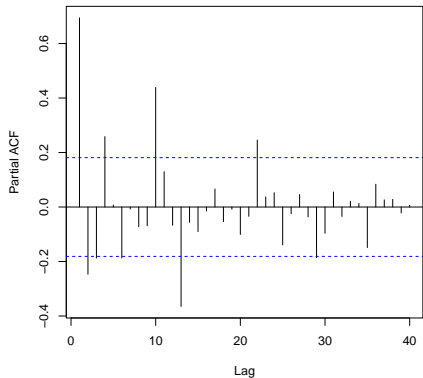
- Como observamos apenas algumas realizações do processo, gostaríamos de testar hipóteses sobre a FAC e FACP populacionais.
- Sob algumas condições mais estridentes, o intervalo $[-2/\sqrt{T}, 2/\sqrt{T}]$ é uma região de aceitação aproximadamente válida para o teste da nula $\gamma_k = 0$ ($\theta_k = 0$ na FACP) contra a alternativa bilateral ao nível de significância de 5%, com base na estatística $\hat{\gamma}_k$ ($\hat{\theta}_k$).
 - São esses intervalos que são apresentados, no R, quando computamos a FAC e FACP.
 - Para autocorrelações (parciais) estimadas que excedem esses limites, rejeitamos a hipótese nula de não autocorrelação (parcial) a essa ordem.

FAC E FACP AMOSTRAIS DE Δ DESEMPREGO

```
Series as.numeric(diff(desemprego_brasil))
```



```
Series as.numeric(diff(desemprego_brasil))
```



INFERÊNCIA CONJUNTA SOBRE A FAC

- Para testar a nula conjunta de que $\gamma_1 = \gamma_2 = \dots = \gamma_s = 0$, onde s é pequeno relativamente a T , podemos usar a **estatística de Ljung-Box**:

$$\hat{Q} = T(T+2) \sum_{r=1}^s \frac{\hat{\gamma}_r^2}{T-r}.$$

- Com T grande, sob a nula, \hat{Q} segue uma qui-quadrado com s graus de liberdade. Valores altos da estatística são evidência contra a nula, i.e. evidência de que ao menos uma autocorrelação entre as testadas é diferente de zero.

FAC E FACP DE UM AR(p) ESTACIONÁRIO

- Como vimos em aula anterior, a FAC de um AR(1) estacionário é dada por:

$$\gamma_k = \text{cor}(Y_t, Y_{t-k}) = \beta_1^k, \quad |\beta_1| < 1$$

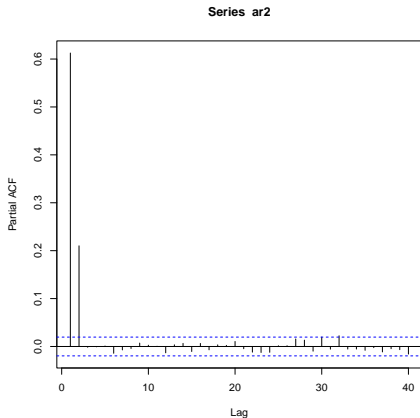
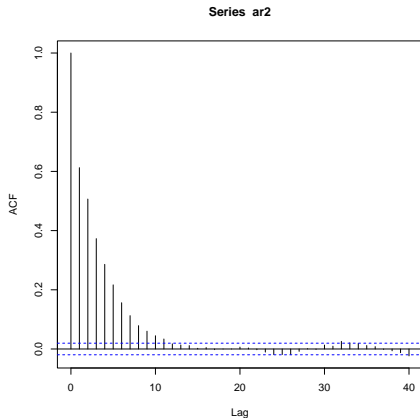
isto é, a FAC apresenta decaimento geométrico em direção a zero.

- De modo geral, a FAC de um AR(p) estacionário apresenta **decaimento em direção a zero**, visto que um AR(p) estacionário pode ser escrito como um MA(∞), i.e.

$$Y_t = \mu + \epsilon_t + \sum_{j=1}^{\infty} \omega_j \epsilon_{t-j}$$

- E a FACP de um AR(p) estacionário? Pela definição da FACP de ordem k como o coeficiente de Y_{t-k} na regressão populacional de Y_t em $Y_{t-1}, Y_{t-2}, \dots, Y_{t-k}$, esperamos que a FACP seja **truncada em p** , visto que o processo só depende diretamente das p primeiras defasagens.

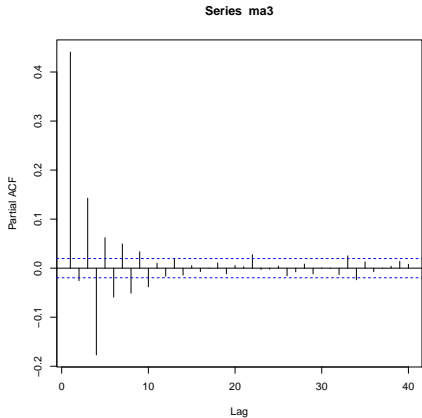
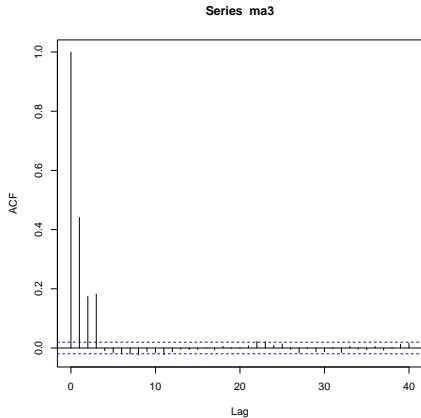
EXEMPLO: FAC E FACP ESTIMADAS DE UM AR(2) ($T=10.000$)



FAC E FACP DE UM MA(Q) INVERTÍVEL

- Como vimos em aula anterior, a FAC de um MA(q) é **truncada** em q , visto que a correlação morre após q períodos.
- E a FACP? Se o processo MA for invertível, vimos que ele pode ser escrito como um $AR(\infty)$. Dessa representação, fica claro que a FACP de um MA(q) apresenta **decaimento em direção a zero**.

EXEMPLO: FAC E FACP ESTIMADAS DE UM MA(3) ($T=10.000$)

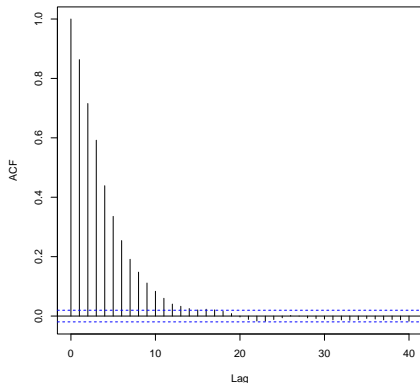


FAC E FACP DE UM ARMA(p,q) ESTACIONÁRIO E INVERTÍVEL

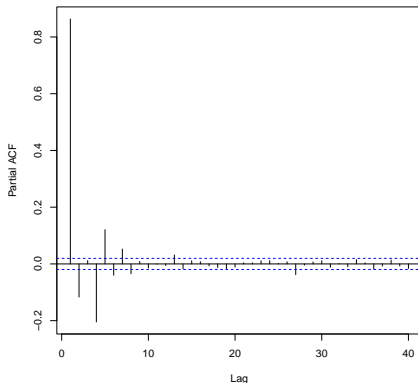
- Generalizando a discussão anterior, um ARMA(p,q) estacionário cuja parte MA é invertível pode ser representado tanto como um $AR(\infty)$ como um $MA(\infty)$. Nesse caso, tanto a FAC como a FACP apresentam decaimento.
- Nesses casos, costuma-se considerar a ordem máxima q_{\max} em que a FAC torna-se pouco significativa e a ordem p_{\max} em que a FACP torna-se pouco significativa e considerar todos os ARMA(p,q), $0 \leq p \leq p_{\max}$ e $0 \leq q \leq q_{\max}$ como candidatos.

EXEMPLO: FAC E FACP ESTIMADAS DE UM ARMA(2,3) (T=10.000)

Series arma23



Series arma23



RESUMO

Modelo	FAC	FACP
AR(p) estacionário	decai	truncada em p
MA(q) invertível	truncada em q	decai
ARMA(p,q) estacionário e invertível	decai (esp. após q)	decai (esp. após p)

Estimação

ESTIMAÇÃO CONDICIONAL VS. INCONDICIONAL

- Para a estimação de modelos $\text{ARMA}(p,q)$, há duas abordagens de estimação.
 - Na abordagem **condicional**, não utilizamos a informação acerca da distribuição das primeiras observações na estimação.
 - Na abordagem **incondicional**, fazemos hipóteses adicionais sobre a distribuição do ruído branco, que nos permitem incorporar a distribuição das primeiras observações na análise.
- Computacionalmente, a abordagem condicional é mais simples, embora menos eficiente que a segunda.
- Embora a abordagem incondicional aparente requerer mais hipóteses, visto que especificamos a distribuição do ruído branco, a estimação é robusta a violações dessa hipótese quando o número de observações é grande (“pseudo” máxima verossimilhança).

ESTIMAÇÃO CONDICIONAL DO AR(p)

- Podemos estimar os parâmetros de um AR(p) através de mínimos quadrados ordinários:

$$(\hat{\alpha}, \hat{\beta}_1, \dots, \hat{\beta}_p) \in \operatorname{argmin}_{a, b_1, \dots, b_p} \frac{1}{T-p} \sum_{t=p+1}^T (Y_t - a - b_1 Y_{t-1} - \dots - b_p Y_{t-p})^2$$

- Note que não tentamos prever as p primeiras observações, para as quais não temos todas as defasagens.
- Estimadores $(\hat{\alpha}, \hat{\beta}_1, \dots, \hat{\beta}_p)$ coincidem com o estimador de máxima verossimilhança que usa a distribuição de $(Y_{p+1}, Y_{p+2} \dots Y_T)$ **condicional** a (Y_1, Y_2, \dots, Y_p) , tomando o ruído branco como $\epsilon_{it} \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$.

ESTIMAÇÃO CONDICIONAL DO MA(1)

- Para a estimação de um modelo MA(1) com base numa série $\{Y_t\}_{t=1}^T$, precisamos de um chute inicial para o ruído branco no período $t = 0$.
- Chamemos esse chute de $\tilde{\epsilon}_0$ (padrão é tomar $\tilde{\epsilon}_0 = 0$).
- Dado o chute inicial, e dado um valor candidato c para o parâmetro θ_1 , e um valor candidato a para o intercepto α , podemos imputar o ruído branco em $t = 1$:

$$\tilde{\epsilon}_1(a, c) = Y_1 - a - c\tilde{\epsilon}_0.$$

- Procedendo recursivamente, obtemos, para todo $t \geq 2$.

$$\tilde{\epsilon}_t(a, c) = Y_t - a - c\tilde{\epsilon}_{t-1}(a, c)$$

- MA(1) pode ser estimado como:

$$(\hat{\alpha}, \hat{\theta}_1) \in \operatorname{argmin}_{a, c} \frac{1}{T} \sum_{t=1}^T (Y_t - a - c\tilde{\epsilon}_{t-1}(a, c))^2$$

ESTIMAÇÃO CONDICIONAL DO MA(1)

- Note que $(a, c) \mapsto a + c\tilde{\epsilon}_{t-1}(a, c)$ varia não linearmente com (a, c) .
 - Estimação se dá através de algoritmos para mínimos quadrados não lineares (não há expressão fechada para o mínimo).
- Observe também que a primeira observação não contribui à estimação de c , visto que $(Y_t - a - c\tilde{\epsilon}_0)^2 = (Y_t - a)^2$.
- Se o MA(1) é **invertível**, então, com T grande, efeito do chute inicial sobre a função objetivo desaparece.
 - Contribuição do chute inicial ao erro de previsão em t é da ordem de c^t , que desaparece quando $|c| < 1$.

ESTIMAÇÃO CONDICIONAL DO ARMA(p,q)

- Estendendo a discussão anterior, a estimação condicional de um ARMA(p,q) é dada pela minimização de:

$$\sum_{t=p+1}^T (Y_t - a - b_1 Y_{t-1} \dots - b_p Y_{t-p} - c_1 \tilde{\epsilon}_{t-1}(a, \mathbf{b}, \mathbf{c}) \dots - c_q \tilde{\epsilon}_{t-q}(a, \mathbf{b}, \mathbf{c}))^2$$

onde $\tilde{\epsilon}_t(a, \mathbf{b}, \mathbf{c})$ são definidos recursivamente, para cada valor candidato dos parâmetros $(a, \mathbf{b}, \mathbf{c})$ e chutes iniciais dos erros

$$\tilde{\epsilon}_{p-q+1} = \tilde{\epsilon}_{p-q+2} = \dots = \tilde{\epsilon}_p = 0 .$$

- Perdemos p observações pois não observamos os valores de Y anteriores a $t = 1$. Além disso, para as observações de $p + 1$ a $p + q$, não temos informação completa para inferir todos os θ_j .

ESTIMAÇÃO INCONDICIONAL DO ARMA(p,q)

- Na estimação condicional, perdemos informação nas $p + q$ primeiras observações.
- Se fizermos uma hipótese distributiva sobre os ruídos brancos, somos capazes de caracterizar a distribuição das $p+q$ primeiras observações.
 - Por exemplo, se $\epsilon_t \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$ e Y_t seguir um AR(1) estacionário, então $Y_1 \sim N(\alpha/(1 - \alpha), \sigma^2/(1 - \rho^2))$.
- A estimação **incondicional** de um ARMA(p,q) se dá pela máxima verossimilhança que usa a distribuição conjunta de $\{Y_1, \dots, Y_T\}$, sob a hipótese auxiliar de que $\epsilon_t \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$.
 - Método padrão na função arima do R.
- Se ruído branco de fato é Gaussiano, estimador é **eficiente**: dentre todos os estimadores de um ARMA(p,q) Gaussiano, estimador é o de menor variância.
- Mesmo que o ruído branco não seja Gaussiano, estimador ainda é **consistente** para os parâmetros de um ARMA(p,q).

Diagnóstico

DIAGNÓSTICO

- Estimados os modelos candidatos, procedemos à etapa de diagnóstico. A ideia é avaliar os modelos conforme algumas métricas:
 1. Critérios de informação.
 2. Parcimônia.
 3. Não autocorrelação dos erros.
 4. Estabilidade e invertibilidade das partes AR e MA.
 5. Convergência numérica dos estimadores.
 6. Normalidade dos erros.

CRITÉRIOS DE INFORMAÇÃO

- A princípio, gostaríamos de uma métrica que indicasse quanto da variabilidade do processo é explicada pelo modelo.
 - Quantidade não observada; precisa ser estimada.
- Intuitivamente, um estimador dessa quantidade poderia ser dado pela soma dos quadrados dos resíduos (SSR) do modelo estimado.
- O problema dessa métrica é que modelos mais complexos **necessariamente** apresentam SSR menor.
 - Maior flexibilidade leva a melhor ajuste dentro da amostra.
 - Isso não significa que o modelo necessariamente explique bem a variação verdadeira do processo (em particular, **fora da amostra**).
- Se fôssemos escolher o modelo pelo menor SSR, sempre escolheríamos o modelo mais complexo, incorrendo num problema conhecido como sobreajuste (*overfitting*).
 - Modelo funcionará, em geral, muito mal fora da amostra, pois estimadores dos parâmetros apresentam alta variância.

CRITÉRIOS DE INFORMAÇÃO (CONT.)

- A ideia de **um critério de informação** é **penalizar** a SSR pelo número de parâmetros estimados.
 - A penalização pode ser vista como uma forma de corrigir o viés do SSR em estimar a capacidade preditiva de um modelo.
- Os critérios de informação mais utilizados são o AIC e o BIC. Para um ARMA(p, q) com intercepto, eles são dados por:

$$AIC = T \ln(SSR) + 2(p + q + 1)$$

$$BIC = T \ln(SSR) + \ln(T)(p + q + 1)$$

- Quanto menor o critério de informação, melhor.
- Para $T > 7$, BIC escolhe modelos com menos parâmetros.
- Se a estimação do modelo é condicional, importante ajustar amostra para que mesmo número de observações sejam usadas no cálculo dos critérios em todos os modelos comparados (isso já vale, por construção, na estimação incondicional).

- Os critérios de informação induzem **parcimônia** no modelo escolhido, ajudando a evitar o problema de sobreajuste.
- Ainda sob essa lógica, é costumeiro verificar quais coeficientes são estatisticamente significantes na especificação: se houver muitos coeficientes insignificantes, talvez valha trabalhar com um modelo mais simples.

NÃO AUTOCORRELAÇÃO DOS ERROS

- Os modelos ARMA discutidos supõem que os erros são ruído branco.
 - Dessa forma, esperaríamos que os *resíduos* de nosso modelo fossem aproximadamente não autocorrelacionados.
 - Se houver correlação nos resíduos, ainda há informação nos dados que a parte sistemática do modelo não está capturando.
- Teste da hipótese nula de que as autocorrelações dos erros de ordens 1 até s são zero, contra a alternativa de que ao menos uma é diferente de zero, podem ser conduzidos usando a estatística de [Ljung-Box](#):

$$Q = T(T+2) \sum_{j=1}^s \hat{\gamma}_j^2 / (T-j),$$

onde $\hat{\gamma}_j$ é autocorrelação de ordem j estimada com base nos resíduos.

- Sob a nula, distribuição de Q é aproximadamente uma χ^2 com $s - (p + q)$ graus de liberdade.
- Quanto maior Q , maior a evidência contra a nula. Assim, região crítica do teste é da forma $Q > c$ onde c é o quantil apropriado da distribuição χ^2 .

ESTABILIDADE E INVERTIBILIDADE

- Recorde-se que a análise de identificação dos modelos ARMA(p,q) pressupõe que os processos sejam estacionários e invertíveis. Assim, é costumeiro verificar se os coeficientes estimados de fato nos levam a processos estacionários e invertíveis.
 - Se isso não ocorrer, devemos suspeitar de nossas estimativas.
- Podemos checar a estacionariedade e invertibilidade do ARMA(p,q) resolvendo, respectivamente, as equações de grau p e q dos polinômios estimados, e avaliando se as raízes se encontram todas fora do círculo

CONVERGÊNCIA NUMÉRICA

- Os estimadores mais usados de modelos ARMA são não lineares e não possuem solução fechada.
 - Por esse motivo, pacotes estatísticos usam algoritmos de otimização para estimar o modelo.
- É importante checar se os algoritmos de otimização de fato convergiram para um mínimo.
 - Se esse não é o caso, devemos descartar as estimativas.

NORMALIDADE DOS ERROS

- Recorde-se que, se os ruídos brancos forem Gaussianos, o estimador de máxima verossimilhança do ARMA(p,q) é eficiente.
 - Ainda assim, mesmo que os ruídos brancos não sejam Gaussianos, o estimador é consistente.
- Nesse sentido, é costumeiro testar a hipótese de normalidade dos erros de um modelo ARMA.
- Isso é feito verificando a assimetria e curtose dos *resíduos* do modelo, e quanto elas distam do esperado em uma distribuição normal.
- Sob a nula de normalidade, a estatística de teste de Jarque-Bera possui distribuição χ^2 com **dois** graus de liberdade.
 - Valores grandes da estatística são evidência contra a nula.

Previsão

PREVISÃO UM PASSO À FRENTE

- Estimado um ARMA(p,q) com base num conjunto de dados $\{Y_t\}_{t=1}^T$, como podemos calcular uma previsão para Y_{T+1} ?
- Recorde-se que, se o processo é descrito por um ARMA(p,q), então:

$$Y_{T+1} = \alpha + \sum_{j=1}^p \beta_j Y_{T+1-j} + \epsilon_{T+1} + \sum_{j=1}^q \theta_j \epsilon_{T+1-j},$$

- Estimação de um ARMA nos dá estimativas para os parâmetros, além de estimativas dos ruídos brancos na janela de estimação, isto é, $\hat{\epsilon}_j$, $j \leq T$.
- Único componente “desconhecido” é ϵ_{T+1} .
 - Mas, como se trata de um ruído branco, um processo sem memória, o melhor a se fazer é $\hat{\epsilon}_{T+1} = \mathbb{E}[\epsilon_{T+1}] = 0$.
- Assim, a **previsão um passo à frente**, com base num ARMA(p,q), é dada por:

$$\hat{Y}_{t+1} = \hat{\alpha} + \sum_{j=1}^p \hat{\beta}_j Y_{T+1-j} + \sum_{j=1}^q \hat{\theta}_j \hat{\epsilon}_{T+1-j}$$

PREVISÃO DOIS PASSOS À FRENTE

- Também podemos estar interessados em prever o que ocorrerá dois passos adiante, i.e. Y_{T+2} .
- Neste caso, temos que:

$$Y_{T+2} = \alpha + \sum_{j=1}^T \beta_j Y_{T+2-j} + \epsilon_{T+2} + \sum_{j=1}^q \theta_j \epsilon_{T+2-j}.$$

- Com base na informação até T , “desconhecemos” ϵ_{T+1} , ϵ_{T+2} e Y_{T+1} .
 - Vamos fazer $\hat{\epsilon}_{T+2} = \hat{\epsilon}_{T+1} = 0$.
 - Vamos usar nossa previsão de \hat{Y}_{T+1} para imputar Y_{T+1} .
- Assim, a **previsão dois passo à frente**, com base num ARMA(p,q), é dada por:

$$\hat{Y}_{t+1} = \hat{\alpha} + \hat{\beta}_1 \hat{Y}_{T+1} + \sum_{j=2}^p \hat{\beta}_j Y_{T+1-j} + \sum_{j=2}^q \hat{\theta}_j \hat{\epsilon}_{T+1-j}$$

PREVISÃO h PASSOS À FRENTE

- Procedendo recursivamente, podemos definir, para qualquer $h \in \mathbb{N}$, a projeção h passos à frente, \hat{Y}_{t+h} , em que usamos $\hat{\epsilon}_{T+j} = 0$ para $j > 0$ e as previsões \hat{Y}_{T+j} , $h - 1 \geq j \geq 1$, na imputação dos termos desconhecidos.
- Observe que para $h > \max(p, q)$, a projeção não usa nenhuma informação nova, somente projeções de horizontes anteriores.
- De fato, quando $h \rightarrow \infty$, as projeções de um ARMA(p, q) estacionário convergem para a média incondicional estimada do processo, i.e.:

$$\lim_{h \rightarrow \infty} \hat{Y}_{T+h} = \frac{\hat{\alpha}}{1 - \hat{\beta}_1 - \hat{\beta}_2 - \dots - \hat{\beta}_p}$$

- Equivalentemente, as projeções de um ARIMA($p, 1, q$) convergem para projeções em que a variação é constante:

$$\lim_{h \rightarrow \infty} \Delta \hat{Y}_{T+h} = \frac{\hat{\alpha}}{1 - \hat{\beta}_1 - \hat{\beta}_2 - \dots - \hat{\beta}_p}$$

INTERVALO DE PREDIÇÃO

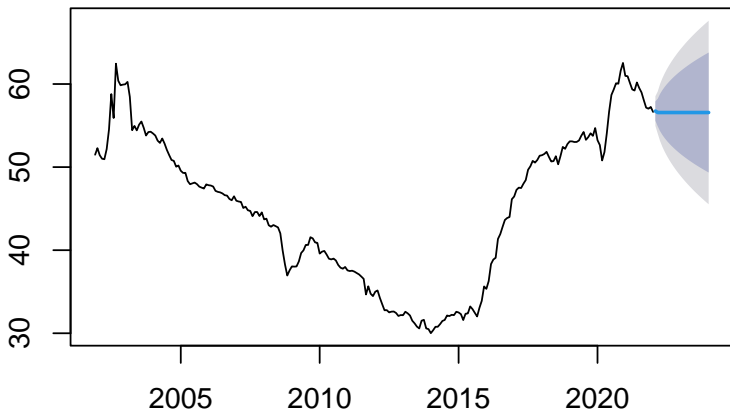
- Um intervalo de predição para Y_{T+h} com confiança γ é um par de funções dos dados, (L, U) , com a propriedade:

$$\mathbb{P}[L \leq Y_{T+h} \leq U] \geq \gamma.$$

- Intervalo de predição contém a realização de Y_{T+h} em ao menos $100\gamma\%$ dos casos, sobre todas as realizações possíveis da incerteza econômica.
- É possível construir um intervalo de predição, com base nas predições de um ARMA(p,q) Gaussiano.
 - Intervalo de predição leva em conta incerteza Gaussiana acerca dos ruídos brancos, ϵ_{T+j} , $j > 0$.
- À medida que h cresce, comprimento do intervalo de confiança cresce, visto que incerteza torna-se cada vez maior.

PROJEÇÕES: ENDIVIDAMENTO PÚBLICO

Forecasts from ARIMA(0,1,2)



Comparação de modelos

USANDO PROJEÇÕES PARA COMPARAR MODELOS

- Se temos uma série longa, podemos realizar um procedimento para validar e comparar modelos, sem incorrer em sobreajuste.
- Ideia é repartir a amostra em dois subconjuntos de períodos, $\mathcal{E} = \{1, 2, \dots, T_0\}$ e $\mathcal{V} = \{T_0 + 1, T_0 + 2, \dots, T\}$.
- No conjunto \mathcal{E} , realizamos todas as etapas para identificar, estimar e diagnosticar modelos.
- Definidos um conjunto \mathcal{M} de modelos estimados com os dados em \mathcal{E} , podemos calcular, para cada horizonte $h \in \mathbb{N}$, previsões h passos à frente, para os períodos em \mathcal{V} .
 - Isto é, para cada modelo $m \in \mathcal{M}$ estimado com os dados de \mathcal{E} , e $t + h \in \mathcal{V}$, definimos como $\hat{Y}_{t+h|t}(m)$ a previsão que se faz com o modelo estimado m , para o período $t + h$, **com as informações até t** .
- Nesse caso, podemos definir o erro quadrático médio da previsão fora da amostra, no horizonte h , como:

$$\text{MSE}(m; h) = \frac{1}{T - T_0 - (h - 1)} \sum_{t=T_0}^{T-h} (Y_t - \hat{Y}_{t+h|t}(m))^2$$

CORRIDA DE CAVALOS

- A medida $\text{MSE}(m; h)$ é livre da influência de sobreajuste, visto que o modelo foi estimado sem recorrer aos dados em \mathcal{V} .
- Podemos comparar os modelos em termos de qual tem o menor $\text{MSE}(m; h)$, para cada horizonte $h \in \mathbb{N}$.
 - A esse procedimento damos o nome de corrida de cavalos (*horseracing*).
 - O resultado dessa comparação é a escolha de um modelo “ótimo”, para cada horizonte.

TESTE DE DIEBOLD-MARIANO

- A comparação $\text{MSE}(m; h)$ para diferentes modelos está sujeita a contingências do período de validação \mathcal{V} .
 - Modelo pode ser bom para a janela \mathcal{V} , mas ideia é que modelo seja bom em prever o futuro.
- Sob estacionariedade do erro de previsão, Diebold e Mariano introduziram um teste da hipótese nula:

$$H_0 : \mathbb{E}[(Y_t - \hat{Y}_{t+h|t}(m))^2] = \mathbb{E}[(Y_t - \hat{Y}_{t+h|t}(m'))^2],$$

para modelos $m, m' \in \mathcal{M}$, e horizonte $h \in \mathbb{N}$.

- Teste é implementado através da estatística t :

$$\hat{t} = \frac{\frac{1}{T-T_0-(h-1)} \sum_{t=T_0}^{T-h} [(Y_t - \hat{Y}_{t+h|t}(m))^2 - (Y_t - \hat{Y}_{t+h|t}(m'))^2]}{\hat{\sigma}},$$

onde $\hat{\sigma}$ é um erro padrão HAC.

- Implementação computacional: construir série $\hat{e}_t = (Y_t - \hat{Y}_{t+h|t}(m))^2 - (Y_t - \hat{Y}_{t+h|t}(m'))^2$, $t = T_0 \dots T - h$. Rodar regressão de \hat{e}_t num intercepto, e conduzir o teste t no intercepto usando `vcovHAC`.

TESTE DE DIEBOLD-MARIANO (CONT.)

- Validade dos erros padrão HAC requer janela de validação grande.
 - Por outro lado, hipótese estacionariedade do erro de previsão requer T_0 também grande, e bem maior que janela de validação (Diebold, 2015).
- Metodologia de Diebold-Mariano pode ser estendida para analisar os determinantes da qualidade preditiva em uma classe de modelos.
- Ideia é considerar modelos lineares:

$$(Y_t - \hat{Y}_{t+h|t}(m))^2 = \gamma' h(m) + u_{m,t}, \quad m \in \mathcal{M}, t = T_0, \dots, T - h$$

onde $h(m) = (h_1(m), h_2(m), \dots, h_J(m))$ são J características do modelo m (e.g. indicador da presença de um componente MA).

- Usando erros padrão HAC em painel, é possível testar quais características de um modelo ensejam melhor qualidade preditiva.
 - Teste t da nula $\gamma_j = 0$ contra alternativa $\gamma_j < 0$, onde γ_j é o coeficiente associado $h_j(m)$.

Modelagem Sazonal

SAZONALIDADE E PREVISÃO

- Até aqui, não discorremos sobre o efeito da sazonalidade de uma série nas previsões.
- Nesse caso, há duas estratégias a se seguir.
 1. Se o objetivo é prever a série livre de seu componente sazonal, podemos realizar a metodologia de Box-Jenkins na série dessazonalizada.
 2. Por outro lado, se o objetivo é prever a série original, devemos incorporar a sazonalidade na análise.
- A maneira mais simples de incorporar a sazonalidade em nossa análise seria incluindo *dummies de período* entre os componentes determinísticos do processo.
 - Argumento Xreg na função arima do R.
- No entanto, essa abordagem supõe sazonalidade **não estocástica**: efeito das variações sazonais é sempre o mesmo.
- Box e Jenkins desenvolveram uma metodologia para previsão com **sazonalidade estocástica**.

ARIMA(P, D, Q) $_h$

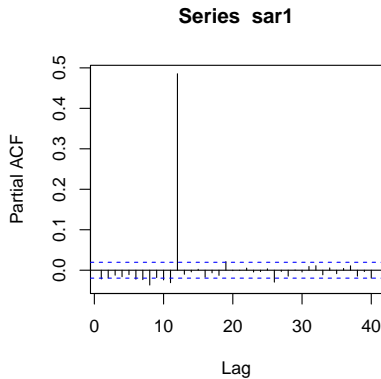
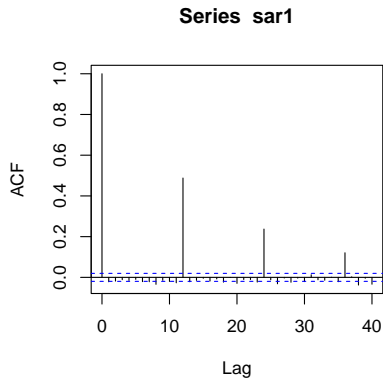
- Uma série de tempo com sazonalidade de frequência h segue um ARIMA(P, D, Q) $_h$ se pode ser descrita como:

$$(1 - \gamma_1 L^h - \gamma_2 L^{2h} \dots - \gamma_P L^{Ph})(1 - L^h)^D Y_t = (1 - \pi_1 L^h - \pi_2 L^{2h} \dots - \pi_Q L^{Qh}) \epsilon_t,$$

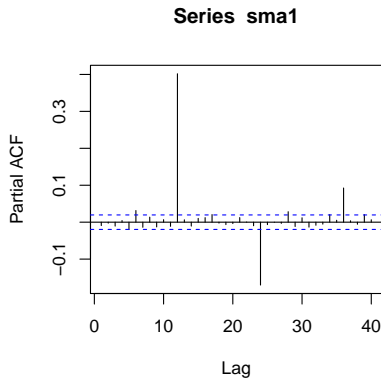
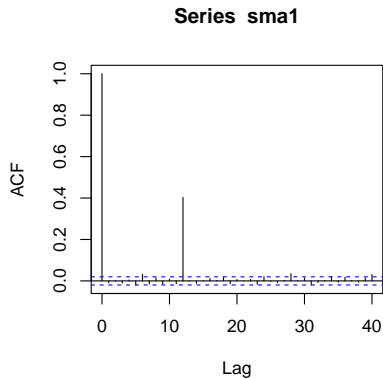
onde ϵ_t é ruído branco.

- ARIMA sazonal é modelo em que defasagens ocorrem a cada h períodos.
- Identificação da ordem P e Q é feita observando-se a FAC e FACP nas ordens $h, 2h, 3h \dots$
- Modelo permite também a presença de **raízes unitárias sazonais**.
 - Diremos que uma série de tempo tem uma raiz unitária sazonal se sua FAC nas ordens $h, 2h, 3h \dots$ **decai muito lentamente/não decai**.
 - Nesse caso, identificação de P e Q deve ser feita observando-se a FAC e FACP do processo em primeira diferença sazonal, $\Delta_h Y_t = Y_t - Y_{t-h}$.

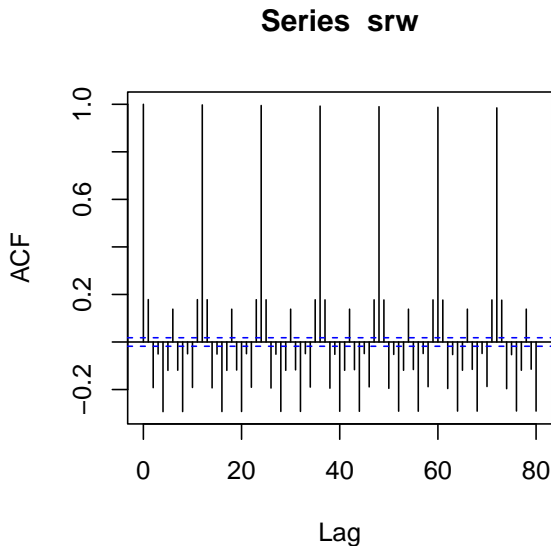
FAC E FACP DE UM $AR(1)_{12}$



FAC E FACP DE UM $MA(1)_{12}$



FAC DE PROCESSO COM RAIZ UNITÁRIA SAZONAL ($h = 12$)



IDENTIFICAÇÃO DA ORDEM (P, D, Q) NO $ARIMA(P, D, Q)_h$

- Série apresenta raiz unitária sazonal ($D = 1$), se FAC decai muito lentamente/não decai nas ordens $h, 2h, \dots$
- Na série livre de raiz unitária sazonal, (Y_t se $D = 0$ ou $\Delta_h Y_t$ se $D = 1$):

Modelo	FAC em $h, 2h, 3h \dots$	FACP em $h, 2h, 3h \dots$
$AR(P)_h$ estacionário	decai	truncada em P
$MA(Q)_h$ invertível	truncada em Q	decai
$ARMA(P, Q)_h$ estacionário e invertível	decai (esp. após Q)	decai (esp. após P)

MODELOS SARIMA

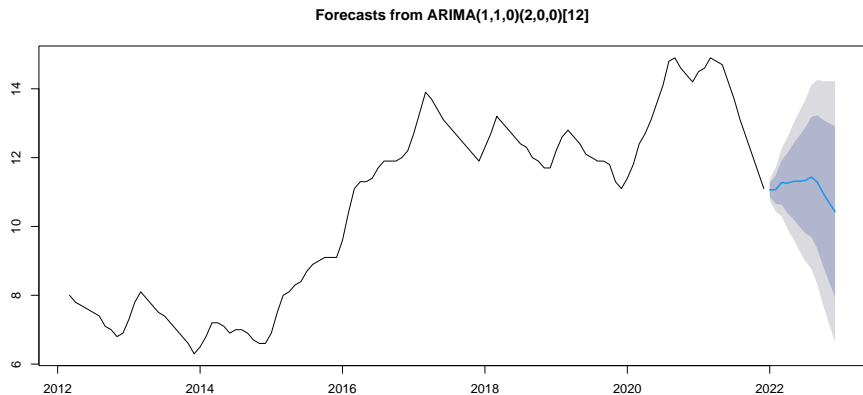
- Dizemos que uma série de tempo segue um processo $ARIMA(p, d, q) \times (P, D, Q)_h$, se:

$$\Phi_P(L^h)\Phi_p(L)(1 - L^h)^D(1 - L)^d Y_t = \Theta_Q(L^h)\Theta_q(L)\epsilon_t,$$

para polinômios de graus p (Φ_p), q (Θ_q), P (Φ_P), Q (Θ_Q), e ruído branco ϵ_t .

- Classe de modelos combina os ARIMA tradicionais com a modelagem sazonal (donde vem o nome SARIMA).
- Identificação das partes não sazonal e sazonal é feita separadamente, conforme vimos em cada modelo correspondente.

DESEMPREGO: MODELAGEM SAZONAL



Análise de intervenção

ANÁLISE DE INTERVENÇÃO

- Há, na Estatística, uma série de extensões à modelagem ARIMA para realizar o que se convencionou chamar “análise de intervenção”.
 - Ideia é utilizar os modelos que vimos para se avaliar os *efeitos* dinâmicos de uma política.
- No entanto, essas abordagens não incorporam a linguagem da inferência causal, seja através da definição de efeitos contrafactuais (modelos estruturais econométricos); seja através do modelo de resultados potenciais de Rubin (cuja origem se dá na Estatística).
 - Isso dificulta a interpretação das hipóteses de identificação necessárias ao uso dos métodos.
- Vamos estudar uma metodologia recente de inferência causal, usando os modelos ARIMA, em que as hipóteses são explicitadas em termos de resultados potenciais (Menchetti, Cipollini e Mealli, 2022).
 - Complemento útil para a análise dos efeitos agregados de políticas, quando não há unidades de controle não expostas disponíveis, ou a teoria econômica não nos fornece informação adicional acerca dos determinantes fundamentais do processo (veremos isso mais à frente).

AMBIENTE

- Considere um processo estocástico $\{Y_t\}_{t \in \mathcal{T}}$ de interesse.
- Suponha que há uma política de interesse, que é implementada a partir de um período T^* .
 - Tratamos T^* como variável aleatória, na medida em que T^* é incerto sobre realizações repetidas da incerteza econômica ($T^* = \infty$ se a política não é adotada em uma realização possível).
- Definimos **resultados potenciais**, $Y_t(1)$ e $Y_t(0)$, que expressam o que ocorre com o processo, sob a presença ou não da política.
- Efeito causal da política em t é dado por $\alpha_t = Y_t(1) - Y_t(0)$.
- **Problema fundamental da inferência causal**: nunca observamos $Y_t(0)$ e $Y_t(1)$ simultaneamente, visto que:

$$Y_t = \begin{cases} Y_t(0) & \text{se } t < T^* \\ Y_t(1) & \text{se } t \geq T^* \end{cases}$$

IDENTIFICAÇÃO DE EFEITOS CAUSAIS NA ABORDAGEM C-ARIMA

- Para utilizar a abordagem ARIMA em inferência causal, requereremos duas hipóteses.

Hipótese 1: MODELO ARIMA PARA $Y_t(0)$

O resultado potencial não tratado segue um modelo (S)ARIMA, onde $\{\epsilon_t\}$ são as inovações (ruídos brancos).

Hipótese 2: sobre a regra de decisão do tratamento

ϵ_t é independente de T^* , para todo $t \geq T^*$.

- Hipótese 2 essencialmente requer que decisão do tratamento dependa somente dos $Y_t(0)$ anteriores à decisão de tratamento, i.e. de $\{Y_t(0) : t < T^*\}$; ou de outros fatores independentes de $\{Y_t(0)\}_{t \in \mathcal{T}}$
 - Hipótese exclui a possibilidade de que decisão dependa de características X_t capazes de prever a inovação ϵ_t no futuro.
 - Se planejador usa modelagem Box-Jenkins para decidir tratamento, hipótese é satisfeita.

ESTIMAÇÃO E INFERÊNCIA

- Sob as duas hipóteses anteriores, é possível estimar os efeitos causais da política da seguinte maneira.
 - Aplicar a metodologia de Box-Jenkins para estimar um modelo (S)ARIMA, com dados até $T^* - 1$.
 - Usando o modelo estimado e os dados até $T^* - 1$, calcular, para $h > 0$, as previsões fora da amostra, h passos à frente: $\hat{Y}_{T^*-1+h|T^*-1}$.
 - Estimar o efeito do tratamento, no h -ésimo período após o início do tratamento como:

$$\hat{\alpha}_{T^*-1+h} = Y_{T^*-1+h} - \hat{Y}_{T^*-1+h|T^*-1}.$$

- Sob as hipóteses 1 e 2, e se a janela pré-tratamento é grande, estimador $\hat{\alpha}_{T^*-1+h}$ é aproximadamente não viciado para $\mathbb{E}[\alpha_{T^*-1+h}]$.
- Se computarmos intervalos de predição para Y com dados até $T^* - 1$, $[L_{T^*-1+h|T^*-1}, U_{T^*-1+h|T^*-1}]$, então $[Y_{T^*-1+h} - U_{T^*-1+h|T^*-1}, Y_{T^*-1+h} - L_{T^*-1+h|T^*-1}]$ é intervalo de predição válido para α_{T^*-1+h} .

REFERÊNCIAS



Diebold, Francis X. (2015). “Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold–Mariano Tests”. Em: *Journal of Business & Economic Statistics* 33.1, pp. 1–1. DOI: 10.1080/07350015.2014.983236. eprint: <https://doi.org/10.1080/07350015.2014.983236>. URL: <https://doi.org/10.1080/07350015.2014.983236>.



Menchetti, Fiammetta, Fabrizio Cipollini e Fabrizia Mealli (set. de 2022). “Combining counterfactual outcomes and ARIMA models for policy evaluation”. Em: *The Econometrics Journal* 26.1, pp. 1–24. ISSN: 1368-4221. DOI: 10.1093/ectj/utac024. eprint: <https://academic.oup.com/ectj/article-pdf/26/1/1/48597685/utac024.pdf>. URL: <https://doi.org/10.1093/ectj/utac024>.