Aula Extra

Gilberto Pereira Sassi

IME-UFBA

Determinação de ordem

Critério de informação

i. *ACF* e *PACF* têm decaimento exponencial e/ou senoidal amortecido para modelo $ARMA(p, q)$
ii. *ACF* e *PACF* não são adequadas para determinar p e q

Definição:

Para $ARMA(p, q)$, considere $\hat{L}$ o valor máximo da verossimilhança. Então:

- Critério da Informação de Akaike (AIC): $AIC = -2\ln(\hat{L}) + 2(p + q + 1)$;
- Critério da Informação Bayesiana (BIC): $BIC = -2\ln(\hat{L}) + ln(n)$, em que é o tamanho da série temporal;
- Critério da Informação de Hannan-Quin (HQIC): $HQIC = -2\ln(\hat{L}) + \log(\log(n))$, em que é o tamanho da série temporal.

Observações:

1. AIC, BIC e HQIC mensuram a quantidade de informação não capturada pelo modelo;
2. AIC tende a escolher valores maiores de (p, q), BIC tende a escolher valores menores de (p, q), e HQIC é um meio termo;
3. `statsmodel.tsa.stattools.arma_order_select(x, max_ar, max_ma, ic=['aic', 'bic', 'hqic'], trend='c')`: escolhe valor adequado para (p, q);
4. Escolher d (modelo ARIMA) diferenciando até alcançar a estacionaridade.

Análise de Diagnóstico do Modelo

Análise de Diagnóstico

Queremos responder:

- modelo estimado é adequado aos dados?
- modelo explica o processo subjacente dos dados?

- $\hat{\varphi}$ e $\hat{\theta}$ são significativamente diferentes de zero
- resíduos tem média zero, são normais e não correlacionados;
  - resíduos: $X_j - \hat{X}_j(\hat{\beta})$, em que $\hat{\beta} = \begin{pmatrix} \hat{\varphi}_1 & \cdots & \hat{\varphi}_p & \hat{\theta}_1 & \cdots & \hat{\theta}_q \end{pmatrix}$
  - resíduos padronizados: $\dfrac{X_j - \hat{X}_j(\hat{\beta})}{\nu_{j-1}}$, em que $\nu_{j-1} = E\left[X_j - \hat{X}_j(\hat{\beta})\right]^2$

Análise de Resíduo

- Correlograma do resíduo: espera-se que $r_0 = 1$ e $r_k = 0, \forall k > 0$
- Teste de Ljung-Box: $H_0 : Corr(\epsilon_t, \epsilon_{t+m}) = 0$ com estatística de teste $Q_{LB}(m) = n(n+2)\sum_{k=1}^{m} \frac{r_k}{n-1}$ com $Q_{LB}(m) \sim \chi_{m-p-q}$ sob H_0
- QQplot e teste de normalidade (Jarque-Bera)
  - em alguns contextos, usamos o modelo apesar do resíduo não ser normal
- Correlograma do resíduo quadrado: espera-se que $r_0 = 1$ e $r_k = 0, \forall k > 0$

Previsão (Forecasting)

Previsão

Definição:

Seja $\{X_t\}$ uma série temporal com média zero e $\{x_{1:n}\}$ uma realização de $\{X_t\}$.

1. Previsão intra-amostra, previsão ex-post, valores ajustados: predição de x_t em que $1 \leq t \leq n$;
2. Previsão fora da mostra (out-of-sample forecasting), previsão ex-ante: previsão de valores futuro além de n, chamado de *origem da previsão*.

---

a. **Notação matemática:** $X_{n+h}, \hat{X}(h), P_n X_{n+h}, X_{n+h}^n$.
b. $\hat{X}_n(h)$ é chamada de *Previsão h passos à frente* e h é chamado de *horizonte de previsão*.

Observação:

- *Erro de previsão:* $e_n(h) = X_{n+h} - \hat{X}_n(h)$;
- *Erro Quadrático Médio:* $EQM = E[e_n(h)]^2$;
- Pode-se provar que a previsão um passo à frente que minimiza EQM é: $\hat{X}_n(1) = E(X_{n+1} \mid X_1, \ldots, X_n), \forall n \geq 1$;
- Existe $\{\phi_{nj} \mid 1 \leq j \leq n\}$ tal que $\hat{X}_n(1) = \sum_{j=1}^{n} \phi_{n,j} X_{n-j+1} = P_n X_{n+1}$, em que $P_n X_{n+1}$ é o melhor preditor linear.

- Para $ARMA(p,q)$ causal, invertível e média zero. A previsão um passo à frente é dada por:

$$\hat{X}_{n+1} = \begin{cases} 0, & n = 0, \\ \sum_{j=1}^{n} \theta_{nj}(X_{n+1-j} - \hat{X}_{n+1-j}), & 1 \leq n < \max(p,q), \\ \sum_{i=1}^{p} \varphi_i X_{n+1-i} + \sum_{j=1}^{q} \theta_{nj}(X_{n+1-j} - \hat{X}_{n+1-j}), & n \geq \max(p,q), \end{cases}$$

em que θ_{nj} são encontradas pelo Algoritmo de Inovações.

- Para $n > 1$ e para uma previsão $h > 1$ passos à frente, então

$$\begin{aligned} \hat{X}_{n+h} &= \hat{X}_n(h) \\ &= \sum_{i=1}^{p} \hat{X}_n(h-i) + \sum_{j=h}^{q} \theta_{n+h-1}(X_{n+h-j} - \hat{X}_n(h-j)). \end{aligned}$$

Exemplos

Exemplo 4.1 (Índice NAO desde janeiro de 1950).

Índice média mensal do Índice *North Atlantic Oscillation* (NAO) desde janeiro de 1950.

> *O índice NAO é importante para a pesquisa meteorológica e baseia-se na diferença de pressão atmosférica ao nível do mar entre a alta subtropical (dos Açores) e a baixa subpolar.*

Arquivo: `dados/brutos/nao.xlsx`.

Exemplo 4.2 (Simulating and Building an $ARMA(2,2)$ Model).

Simule e depois ajuste o modelo $ARMA(2,2)$ dado por:

$$X_t = 0.8X_{t-1} - 0.6X_{t-2} + 0.7\epsilon_{t-1} + 0.4\epsilon_{t-2} + \epsilon_t, \quad \epsilon_t \sim N(0,1).$$

Exemplo 4.3 (Global Annual Mean Surface Air Temperature Changes Series from 1880 to 1985).

Dados publicados por Hansen and Lebedeff (1987) que investiga aquecimento global.

Arquivo: `dados/brutos/global_temperature_1880_1985.csv`.

Exemplo 4.4 (Letras do Tesouro do Governo dos EUA de janeiro de 1950 a junho de 1988).

Taxa de Juros das Letras do Tesouro do Governo dos EUA de janeiro de 1950 a junho de 1988.

Arquivo: `dados/brutos/usbill_month.csv`.