IA e ML Aplicados a Finanças

Prof. Leandro Maciel

AULA 5: Previsão de Séries Temporais

Agenda

1. Séries temporais

- **Série temporal** $\rightarrow$ dados indexados no tempo;
- Séries temporais fin/eco: preços, PIB, inflação, desemprego, etc;

Econometria de Séries Temporais (Estado-da-Arte):

- Desenvolver modelos para inferência e previsão;
- Previsão como mecanismo de redução de **incertezas**;
- Instrumento de apoio na tomada de decisão.

Modelos de equações de diferenças (base teórica):

- Variável explicada por valores passados, pelo tempo e outras variáveis;
- Hipótese de dependência temporal (passado influencia futuro);
- Modelo matemático com **componente estocástico** (modelo estatístico);
- **Processo estocástico** $\rightarrow$ v.a. indexada no tempo;
- Estrutura de mapeamento de séries temporais por eq. de diferenças:

$$y_t = f(y_{t-1}, y_{t-2}, \ldots, y_{t-p}, t, x_{1,t-1}, \ldots, x_{2,t-1}, \ldots, \epsilon_t, \epsilon_{t-1}, \ldots)$$

- Previsão da inflação como aplicação empírica:
  - variável crucial em economia e finanças;
  - importante para condução de políticas fiscal e macroeconômica;
  - relevante na precificação de ativos/derivativos.
- Variável de difícil predição (aleatoriedades, dinâmica complexa);
- Análise inicial: visualização e estatísticas.

- Índice Nacional de Preços ao Consumidor Amplo (IPCA):
  - o IPCA mede a inflação de um conjunto de bens e serviços comercializados no varejo, referentes ao consumo pessoal das famílias, cujo rendimento varia entre 1 e 40 salários mínimos, visando uma cobertura de 90% das famílias pertencentes as áreas urbanas de abrangência do Sistema Nacional de Índices de Preços ao Consumidor (SNIPC), qualquer que seja a fonte de rendimentos. É calculado a partir dos resultados dos índices regionais, utilizando-se a média aritmética ponderada e cuja a variável de ponderação é o Rendimento Familiar Monetário Disponível, tendo como fonte de informação a Pesquisa de Orçamentos Familiares - POF. O IPCA tem sido utilizado pelo Banco Central do Brasil como principal parâmetro de monitoramento do sistema de metas de inflação desde a implementação do sistema no ano de 1999;
  - frequência mensal;
  - taxa de variação do IPCA (% a.m.) $\rightarrow$ inflação.

Modelos ARIMA $\rightarrow$ estado-da-arte para previsão de séries temporais;

- *Autoregressive Integrated Moving Average*;
- Técnica estatística baseada em equações de diferenças;
- Robusta para séries temporais com diferentes dinâmicas;
- Condição para identificação: séries temporais estacionárias;
- Restrição para garantir estabilidade do modelo.

2. Estacionariedade

- Notação: $\{y_t\}, t = 1, 2, \ldots \rightarrow$ série temporal ou processo estocástico;

- Série y_t, com média e variância finitos, é **covariância estacionária** se:

$E(y_t) = E(y_{t-s}) = \mu \rightarrow$ **média**

$E[(y_t - \mu)^2] = E[(y_{t-s} - \mu)^2] = \sigma_y^2 \rightarrow$ **variância**

$E[(y_t - \mu)(y_{t-s} - \mu)] = E[(y_{t-j} - \mu)(y_{t-j-s} - \mu)] = \gamma_s \rightarrow$ **covariância**

$\mu, \sigma_y^2, \gamma_s \rightarrow$ constantes e $\forall$ t e $t - s$.

2. Estacionariedade

- **Estacionariedade**: momentos estatísticos constantes;
- Séries com tendência são não estacionárias (média muda com o tempo);
- Identificação: testes de estacionariedade (ou **raiz unitária**);
- Testa-se $H_0 : a = 1$ em $y_t = ay_{t-1} + \epsilon_t$, com $\{\epsilon_t\} \sim RB(0, \sigma^2)$;
- Raiz unitária $\rightarrow a = 1$ processo explosivo (não estacionário) - simular?
- $\Delta y_t = y_t - y_{t-1} = \rho y_{t-1} + \epsilon_t$ onde $\rho = a - 1 \rightarrow H_0 : \rho = 0$ (usual).

Teste Dickey-Fuller Aumentado (ADF) (Dickey e Fuller, 1979):

- $H_0 : \rho = 0 \rightarrow$ tem raiz unitária $\rightarrow$ **série é não estacionária**;
- Valores críticos dependem do modelo e do tamanho da amostra.

$$\Delta y_t = \rho y_{t-1} + \sum_{i=2}^{p} \beta_i \Delta y_{t-i+1} + \epsilon_t \rightarrow \text{sem intercepto e sem tendência}$$

$$\Delta y_t = a_0 + \rho y_{t-1} + \sum_{i=2}^{p} \beta_i \Delta y_{t-i+1} + \epsilon_t \rightarrow \text{com intercepto}$$

$$\Delta y_t = a_0 + \rho y_{t-1} + a_2 t + \sum_{i=2}^{p} \beta_i \Delta y_{t-i+1} + \epsilon_t \rightarrow \text{com intercepto e tendência}$$

2. Estacionariedade

- Como tornar uma série não estacionária em estacionária?
- Se $\{y_t\}$ é não estacionária, usamos **diferenciação**:

$$\Delta y_{t-1} = y_t - y_{t-1}$$

- Se $\{\Delta y_t\}$ é estacionária, $\{y_t\}$ é I(1): **integrada de ordem 1**;
- C.c., diferenciamos sucessivamente até tornar série estacionária;
- Se $\{y_t\}$ é estacionária, então $\{y_t\}$ é I(0): estacionária em nível.

3. Modelos ARIMA

- Um modelo ARIMA(p, d, q) pode ser representado como:

$$\Delta^d y_t = \omega + \alpha_1 \Delta^d y_{t-1} + \ldots + \alpha_p \Delta^d y_{t-p} + \beta_1 \epsilon_{t-1} + \ldots + \beta_q \epsilon_{t-q} + \epsilon_t$$

- $\{\omega, \alpha_i, \beta_j\} \rightarrow$ parâmetros, e $\epsilon_t \sim RB(0, \sigma^2)$;
- $d \rightarrow$ ordem de integração. Ex.: $\Delta y_t = y_t - y_{t-1}$, $\Delta^2 y_t = \Delta y_t - \Delta y_{t-1}$ ...
- Se a série é estacionária em nível - I(0) $\rightarrow d = 0$ e $\Delta^0 y_t = y_t$, então:

$$y_t = \omega + \alpha_1 y_{t-1} + \ldots + \alpha_p y_{t-p} + \beta_1 \epsilon_{t-1} + \ldots + \beta_q \epsilon_{t-q} + \epsilon_t$$

- Como **identificar** o modelo? Determinar p, d e q?
- $d \rightarrow$ testes de raiz unitária;
- p e $q \rightarrow$ visualização funções autocorrelação e autocorrelação parcial;
- Critérios de informação Akaike (AIC) e Bayesiano (BIC) (**parcimonia**):

$$\text{AIC} = -2\ \ln(L)/T + 2n/T \quad \text{e} \quad \text{BIC} = -2\ln(L)/T + n\ \ln(T)/T$$

- T tamanho da amostra, n no. de parâmetros e L log de verossimilhança.

Procedimento Box & Jenkins:

- 1. Identificação $\rightarrow$ testes raiz unitária e funções de autocorrelação;
- 2. Estimação $\rightarrow$ várias estruturas estimadas (máxima verossimilhança);
- 3. Diagnóstico $\rightarrow$ checar se resíduos são RB.

Para a série do IPCA?

Deixamos uma amostra para **teste** (últimos 24 meses)...

- Modelo ARIMA(1,0,1) para IPCA:

$$y_t = 0.4941 + 0.5411 \cdot y_{t-1} + 0.1684 \cdot \epsilon_{t-1}, \quad BIC = 122.29$$

- Modelo ARIMA(2,0,1) para IPCA:

$$y_t = 0.4938 + 0.2966 \cdot y_{t-1} + 0.1642 \cdot y_{t-2} + 0.1684 \cdot \epsilon_{t-1}, \quad BIC = 126.96$$

- Função ‘auto.arima’ $\rightarrow$ seleção automática;
- (!) não avalia adequação do modelo;
- Diagnóstico: os resíduos são um ruído branco?

- Dado $y_t = \hat{\omega} + \hat{\alpha}_1 y_{t-1}$, a **previsão um passo à frente** é:

$$\hat{y}_{t+1} = \hat{\omega} + \hat{\alpha}_1 y_t$$

- **Dois passos à frente**:

$$\hat{y}_{t+2} = \hat{\omega} + \hat{\alpha}_1 \hat{y}_{t+1} = \hat{\omega} + \hat{\alpha}_1(\hat{\omega} + \hat{\alpha}_1 y_t)$$

- h **passos à frente**:

$$\hat{y}_{t+h} = \hat{\omega}(1 + \hat{\alpha}_1 + \hat{\alpha}_1^2 + \ldots + \hat{\alpha}_1^{h-1}) + \hat{\alpha}^h y_t$$

- Precisamos definir métricas de erro para avaliar as previsões;
- Raiz do erro quadrático médio (RMSE):

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{N}}$$

y_i valor real, $\hat{y}_i$ valor previsto, N número de previsões;

- Menor valor $\rightarrow$ maior acurácia;
- Há inúmeras outras medidas de qualidade de ajuste.

Problema

Determine o melhor modelo ARIMA para a série do IPCA. Além disso, verifique se a inclusão de outras variáveis auxilia na previsão do IPCA quando inclusas como regressor exógeno no modelo ARIMA. Use o help do R para a função "Arima" para consultar a inclusão de regressores exógenos em modelos ARIMA.

- Próxima aula...
  - modelos não-lineares;
  - Redes Neurais Artificias...

Cowperwait, P. S. P. & Metcalfe, A. V. **Introductory Time Series With R**. New York: Springer, 2009. Capítulos 6 e 7.

Prof. Leandro Maciel

leandromaciel@usp.br