Transformação de Box-Cox
Modelos Sazonais

R. Ballini

Bibliografia Básica:

- Enders, W. *Applied Econometric Time Series*. Cap. 2-p.97–103.
- Bueno, R. L. S. *Econometria de Séries Temporais*. Cap. 3,p.94-104.
- Box, G.E., Jenkins, G. M., Reinsel, G.C. & Ljung, G. M.(2016). *Time series analysis: forecasting and control*. Cap. 9.
- Morettin, P. A. *Análise de Séries Temporais*. Cap. 10.

Transformação de Box-Cox

No caso de séries econômicas e financeiras poderá ser necessário aplicar à série original alguma transformação não-linear, como a logarítmica.
Box e Cox (1964) propuseram a seguinte transformação da forma:

$$Y_t^{(\lambda)} = \begin{cases} \frac{Y_t^{\lambda} - 1}{\lambda}, & \text{se } \lambda \neq 0 \\ log(Y_t), & \text{se } \lambda = 0 \end{cases}$$

chamada **transformação de Box-Cox** (1964). O parâmetro λ é estimado usando o método de máxima verossimilhança.

Razões para transformar os dados: estabilizar a variância e tornar o efeito sazonal aditivo.

Erros de Previsão

Calcula-se os erros de previsão $e_{t+h} = y_{t+h} - \hat{y}_{t+h}$:

Erro médio:

$$ME = \frac{\sum_{h=1}^{H} e_{t+h}}{H} =$$

Raiz do Erro médio quadrático :

$$RMSE = \sqrt{\frac{\sum_{h=1}^{H} e_{t+h}^2}{H}}$$

Erro médio absoluto :

$$MAE = \frac{\sum_{h=1}^{H} |e_{t+h}|}{H}$$

Erro médio absoluto percentual:

$$MAPE = \frac{1}{H} \sum_{h=1}^{H} \frac{|e_{t+h}|}{y_{t+h}}$$

Erros de Previsão

Coeficiente U de Theil avalia o desempenho da previsão em relação à previsão ingênua ou trivial:

$$UTHEIL = \frac{\sqrt{\sum_{h=1}^{H} e_{t+h}^2}}{\sqrt{\sum_{h=1}^{H} (y_{t+h} - y_{t+h-1})^2}}$$

Previsão ingênua ou trivial significa que a estimativa do valor futuro é igual ao valor atual.

O coeficiente U de Theil analisa a qualidade da previsão através dos valores:

- $U > 1$: significa que o erro do modelo é maior do que da previsão ingênua;
- $U < 1$: significa que o erro do modelo é menor que da previsão ingênua.

O coeficiente U de Theil menor do que 1 indica uma previsão melhor que a previsão ingênua; quanto mais próximo o mesmo for de zero, melhor será o resultado da previsão.

Sazonalidade

Definição:

Movimento intra-anual sistemático causado por variações climáticas, férias, feriados, etc.

Modelo Geral de Sazonalidade:

Se a sazonalidade independe da tendência:

$$Y_t = T_t + S_t + \epsilon_t$$

Se as amplitudes sazonais variam com a tendência:

$$Y_t = T_t \times S_t \times \epsilon_t$$

Sazonalidade

Ajustamento sazonal ou dessazonalização:

Identificar a presença do componente sazonal, estimá-lo e removê-lo, para analisar o comportamento de tendência de longo prazo.

Metodologias para Tratar Sazonalidade:

1. Modelos com dummies sazonais: usado quando o processo sazonal é puramente determinístico.

Sazonalidade

2. Filtros:

i. Operador diferença: $\Delta_s = (1 - L^s)$:

Se o valor de y_t depende de y_{t-12}, então tomar somente a primeira diferença não sazonal $\Delta y_t = y_t - y_{t-1}$ não induz à estacionariedade. Pode ser necessário tomar a primeira diferença sazonal (ou talvez a segunda diferença),

$$\Delta_{12} y_t = y_t - y_{t-12}$$

Sazonalidade

3. Modelos ARIMA com sazonalidade: há duas maneiras de lidar com séries sazonais:

i. Sazonalidade é removida por algum método de dessazonalização, e após os coeficientes ϕ e θ são estimados usando a metodologia Box-Jenkins;

ii. Tratar a sazonalidade no modelo ARIMA: modelos sazonais denominados **SARIMA**.

Modelos SARIMA

O modelo sazonal multiplicativo geral, denotado por $SARIMA(p,d,q) \times (P;D;Q)_s$, é escrito como:

$$\phi(L)\Phi(L^s)\Delta^d\Delta_S^D y_t = \theta(L)\Theta(L^s)\epsilon_t$$

Na etapa de identificação da ordem dos filtros de um modelo SARIMA aplicamos as mesmas ferramentas que aplicamos num modelo ARIMA não sazonal, isto é, a FAC e FACP amostrais.

Modelos SARIMA

A determinação de p, P, q e Q é feita com base na FAC e FACP da série estacionária, isto é, na série resultante da aplicação das diferenças consecutivas e/ou sazonais. Logo, o primeiro passo é determinar as ordens d e D.

Por exemplo, se a FAC de Δy_t revelar valores altos nas ordens múltiplas de s e que declinam lentamente, é preciso aplicar a diferença sazonal ($\Delta^d \Delta_s^D y_t$).

Modelos SARIMA

Quando a série é gerada por um processo puramente sazonal (isto é, p = q = 0), a FAC e FACP têm um comportamento análogo ao dos modelos não sazonais, porém, os valores não nulos ocorrem apenas nas ordens múltiplas de *s*.

Na prática, os coeficientes de autocorrelações de ordens baixas ajudam a identificar p e q, e os coeficientes de ordens múltiplas de *s* ajudam a definir P e Q.

A identificação das ordens dos filtros sazonais é feita por tentativa até que se tenha pelo menos um "bom" modelo, isto é, um modelo que gere resíduos com comportamento próximo a $RB \sim (0; \sigma^2)$.

Modelos SARIMA

As outras etapas de estimação, verificação e previsão da aplicação da metodologia Box-Jenkins à modelos sazonais são igualmente as realizadas para os modelos não sazonais.

Para análise da série no R:

- Faça o gráfico da série, usando a função `plot()`
- Gráfico `monthplot()` ajuda a detectar visualmente a presença de sazonalidade na ST;
- Faça a decomposição da série, a partir da função `decompose()`
- A série é não estacionária na parte não sazonal?
- A série é não estacionária na parte sazonal?

Testando a estacionariedade da parte não sazonal

Há, basicamente, quatro maneiras de observar se a ST em estudo é ou não estacionária:

- Análise gráfica;
- Comparar a média e a variância em diferentes períodos de tempo da ST;
- Observar a FAC (Função de Autocorrelação);
- Testes de Raiz Unitária.

Testando a estacionariedade da parte sazonal

Caso nos lags sazonais a FAC apresenta decaimento lento, é uma indicação de que a série temporais seja não estacionária na parte sazonal.

Há teste de raiz unitária para série temporal não sazonal: HEGY

Outra possibilidade aplicar diferença na série já diferenciada na parte sazonal e aplicar ADF.

Modelando a ST

Etapas para modelagem:

1. **Especificação**: classe geral de estruturas SARIMA(p,d,q)(P,D,Q) é analisada;
2. **Identificação**: com base na FAC e FACP amostrais e outros critérios.
3. **Estimação**: os parâmetros do modelo identificado são estimados e testados estatisticamente sobre sua significância.

Modelando a ST

Etapas para modelagem:

4. **Diagnóstico**: faz-se uma análise dos resíduos (devem ser ruído branco) e testes de verificação (Ljung-Box) para ver se é adequado o modelo sugerido. Em seguida, verificar os modelos que apresentam menores valores para os critérios AIC e BIC. Caso haja problemas no diagnóstico, volta-se à identificação.

5. **Previsão**: Verificar quais modelos têm os menores erros