Previsão de Séries Temporais

Eduarda

Cronograma

CONTEÚDO	DIA
1. Apresentação	26/08 (Merlin)
2. Definição de Séries Temporais	09/09 (JP)
3. Machine Learning + Pré-processamento de TS	16/09 (Jorge)
4. Forecasting	**23/09 (Eduarda)**
5. Algoritmos 1 (Distância, Intervalo, Dicionário)	30/09 ()
6. Algoritmos 2 (Deep Learning, Comitês)	07/10 (Miller)
7. Aula prática usando Aeon	14/10 ()

Introdução

A **previsão** do valor de uma série temporal y no tempo t (y_t usando todas as observações passadas é denotada por $\hat{y}_{t|t-1}$ ou só $\hat{y}_t$.

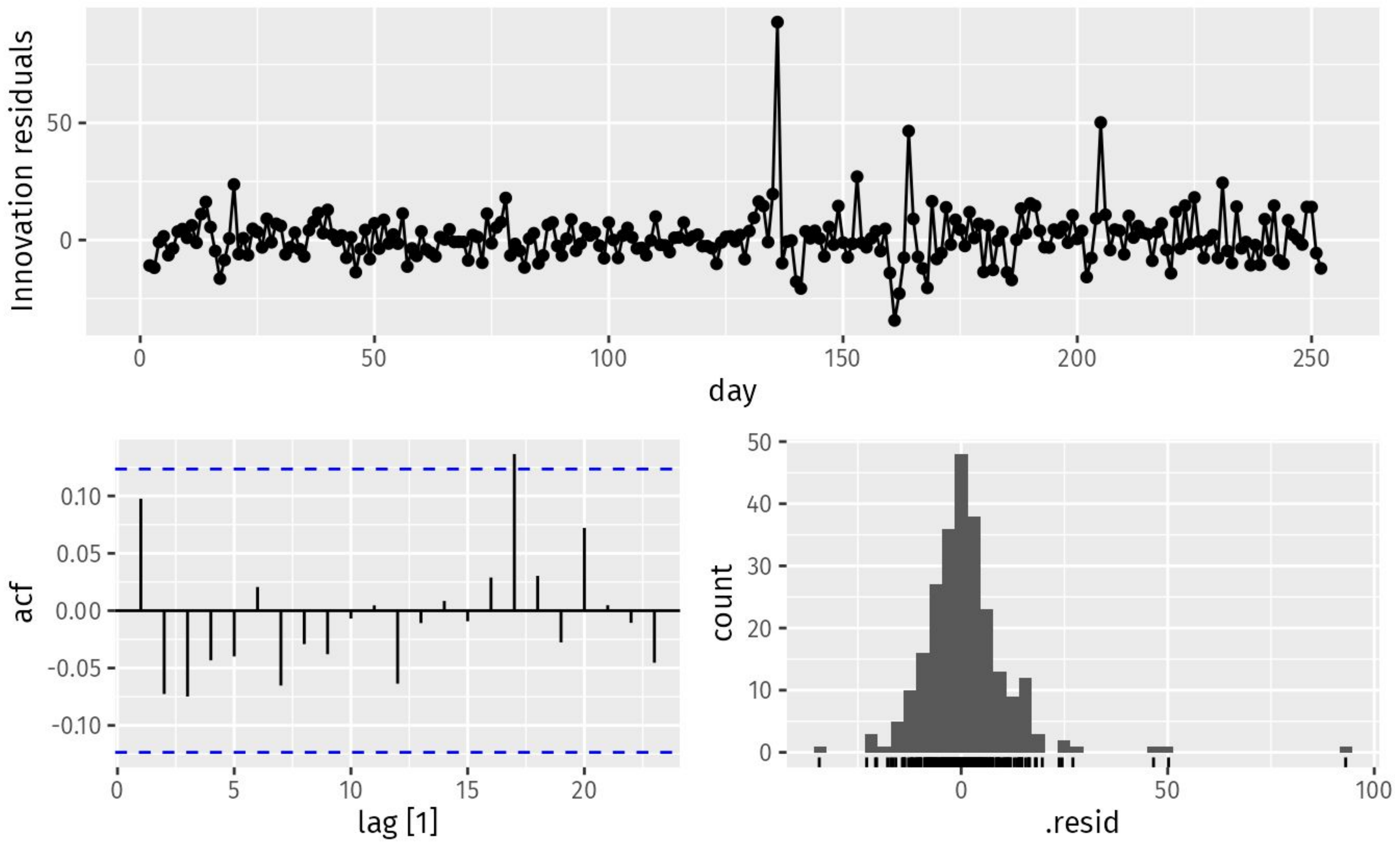
O **resíduo** é a diferença entre o valor observado e o valor previsto: $e_t = y_t - \hat{y}_t$.

Um bom método de previsão deve gerar resíduos **sem correlação** e com **média nula**. Se isso não for obtido, ainda existe mais informação nas observações para o modelo extrair.

É útil que os resíduos tenham variância constante e distribuição normal, mas não é necessário.

Coeficientes de autocorrelação medem a correlação da série y com ela mesma deslocada no tempo. Juntos formam a função de autocorrelação (ACF).

$$r_k = \frac{\sum_{t=k+1}^{T} (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^{T} (y_t - \bar{y})^2}$$


Innovation residuals
50
0
0
50
100
150
200
250
day
acf
0.10
0.05
0.00
-0.05
-0.10
0
5
10
15
20
lag [1]
count
50
40
30
20
10
0
0
50
100
.resid

Ao usar a ACF para fazer testes de autocorrelação, em que cada coeficiente é submetido a um teste de hipótese, pode-se ter falsos positivos, principalmente considerando uma grande quantidade de coeficientes.

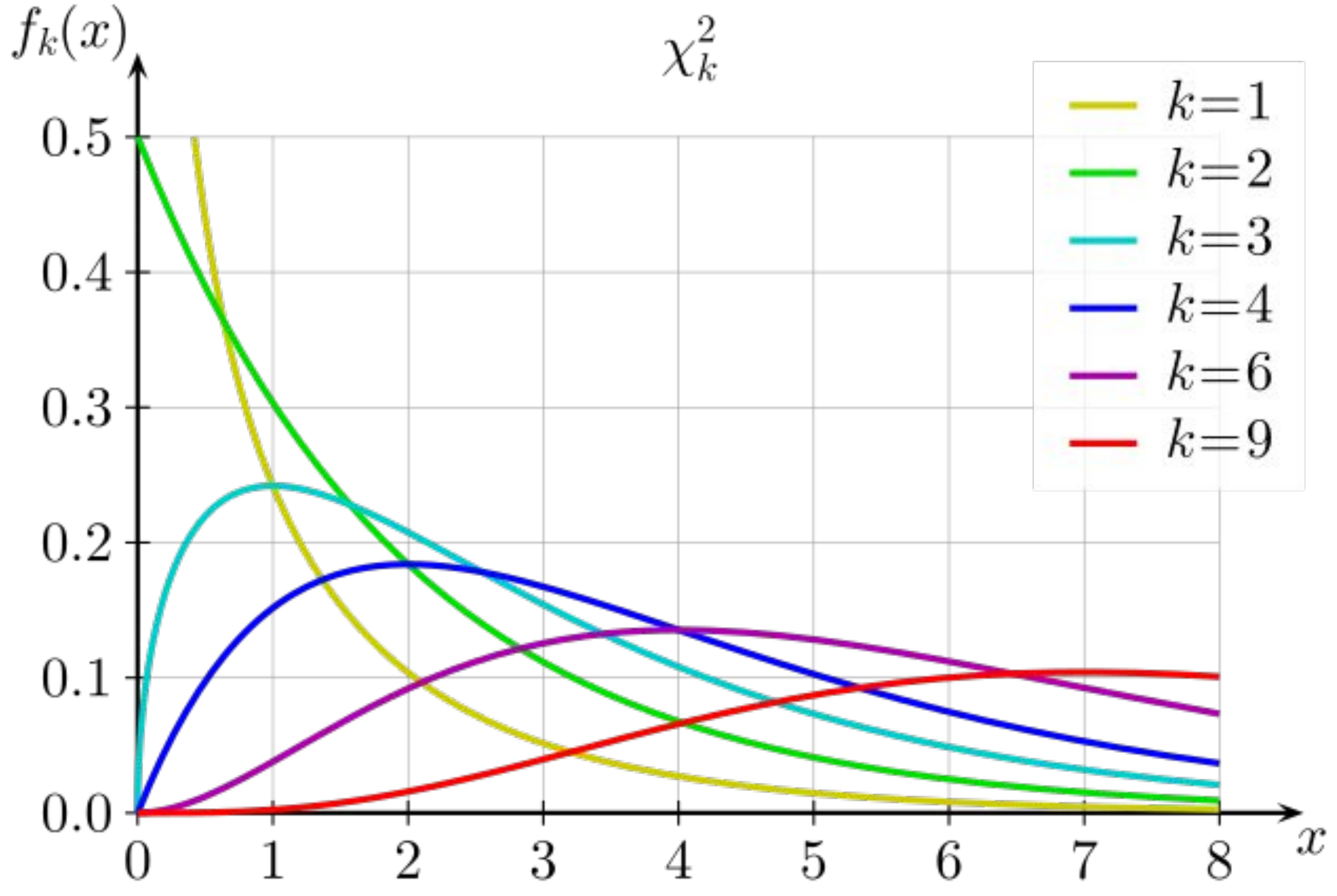
Teste de Box-Pierce:

$$Q = T \sum_{k=1}^{\ell} r_k^2$$

Teste de Ljung-Box:

$$Q^* = T(T+2) \sum_{k=1}^{\ell} (T-k)^{-1} r_k^2$$

Se as autocorrelações vieram de ruído branco, Q e Q^* tem distribuição χ^2 com ℓ graus de liberdade (ℓ é o lag máximo usado para calcular Q e Q^*). Por isso, valores muito altos de Q e Q^* indicam que os resíduos não são ruídos brancos.



Pode-se expressar previsões em um intervalo, ao invés de um ponto, que tem alguma probabilidade de conter o valor observado. Na maioria dos modelos, esse intervalo tem distribuição normal.

A média é a previsão pontual e o desvio padrão pode ser estimado por:

$$\hat{\sigma} = \sqrt{\frac{1}{T-K-M}\sum_{t=1}^{T} e_t^2}$$

K: número de parametros estimados no método de previsão

M: número de resíduos faltantes

Modelos de Regressão

Faz-se uma predição da série temporal *y* assumindo que tem uma relação linear com a série temporal *x*

Regressão linear **simples** (x é univariada): $y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$

Regressão linear **múltipla** (x é multivariada):

$$y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \cdots + \beta_k x_{k,t} + \varepsilon_t$$

Assume-se agora que os resíduos, além de terem média nula e não terem autocorrelação, não tem correlação com as variáveis preditoras (x).

Estimar os coeficientes por mínimos quadrados.

$$\sum_{t=1}^{T} \varepsilon_t^2$$

$$\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 x_{1,t} + \hat{\beta}_2 x_{2,t} + \cdots + \hat{\beta}_k x_{k,t}$$

Pode-se medir quão bem uma regressão se ajusta aos dados com o quadrado da correlação entre y e ŷ ou com o desvio padrão dos resíduos (também usado para gerar o intervalo da previsão)

$$R^2 = \frac{\sum(\hat{y}_t - \bar{y})^2}{\sum(y_t - \bar{y})^2}$$

$$\hat{\sigma}_e = \sqrt{\frac{1}{T-k-1}\sum_{t=1}^{T} e_t^2}$$

Modelos de Suavização Exponencial

Suavização exponencial simples

Melhor para observações sem tendência ou sazonalidade.

Média ponderada das observações anteriores com mais peso para observações mais recentes. Tem previsão constante:

$$\hat{y}_{T+h|T} = \hat{y}_{T+1|T}$$

Hiperparâmetro $0 \leq \alpha \leq 1$, quanto mais perto de 1 mais peso para observações recentes.

Suavização exponencial simples

$$
\begin{aligned}
\hat{y}_{T+1|T} &= \alpha y_T + \alpha(1-\alpha) y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + \cdots \\
&= \alpha y_T + (1-\alpha)\hat{y}_{T|T-1} \\
&= \sum_{j=0}^{T-1} \alpha(1-\alpha)^j y_{T-j} + (1-\alpha)^T \ell_0
\end{aligned}
$$

Tem-se também o hiperparâmetro $\ell_0 = \hat{y}_1$.

Suavização exponencial simples

Para estimar α e ℓ_0, resolve-se um problema de otimização na soma dos quadrados dos resíduos. Não há fórmulas que retornam os hiperparâmetros como nos modelos de regressão.

$$\sum_{t=1}^{T}(y_t - \hat{y}_{t|t-1})^2 = \sum_{t=1}^{T} e_t^2$$

Método de Holt

Estende a suavização exponencial simples para observações com **tendência**. Não tem mais previsão constante.

$$\hat{y}_{t+h|t} = \ell_t + hb_t$$
$$\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1})$$
$$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$$

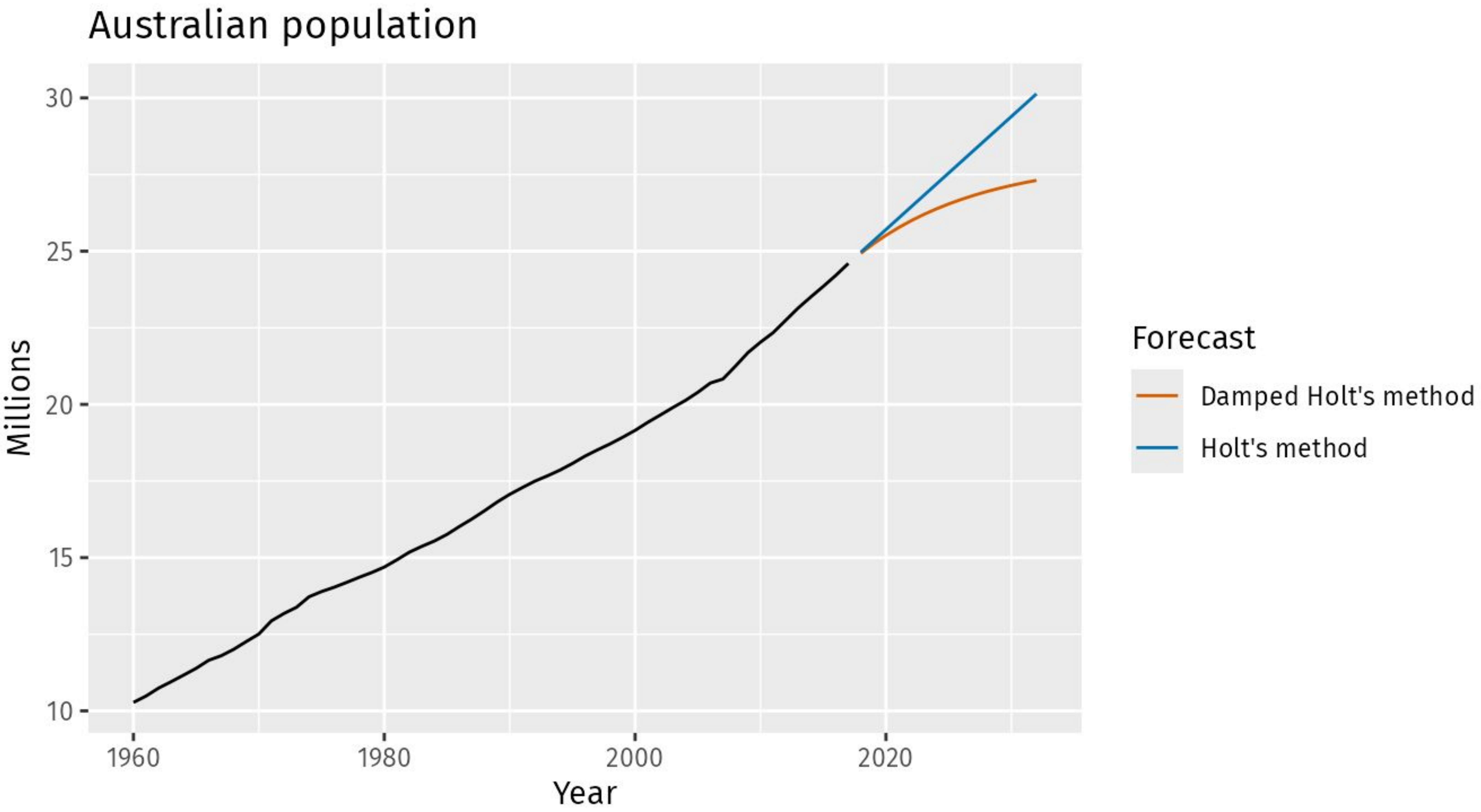
Hiperparâmetros $0 \leq \alpha, \ \beta^* \leq 1, \ \ell_0$ e b_0.

Método de Holt amortecido

A tendência geralmente não se mantém indefinidamente no futuro, por isso incluí-se um parâmetro para amortizar a tendência.

$$\begin{aligned}
\hat{y}_{t+h|t} &= \ell_t + (\phi + \phi^2 + \cdots + \phi^h)b_t \\
\ell_t &= \alpha y_t + (1-\alpha)(\ell_{t-1} + \phi b_{t-1}) \\
b_t &= \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}
\end{aligned}$$

Hiperparâmetros $0 \leq \alpha,\ \beta^* \leq 1,\ \ell_0, b_0$ e ϕ.


Australian population
30
25
20
15
10
Millions
1960
1980
2000
2020
Year
Forecast
Damped Holt's method
Holt's method

Método de Holt-Winters

Estende a o método de Holt para observações com **sazonalidade**.

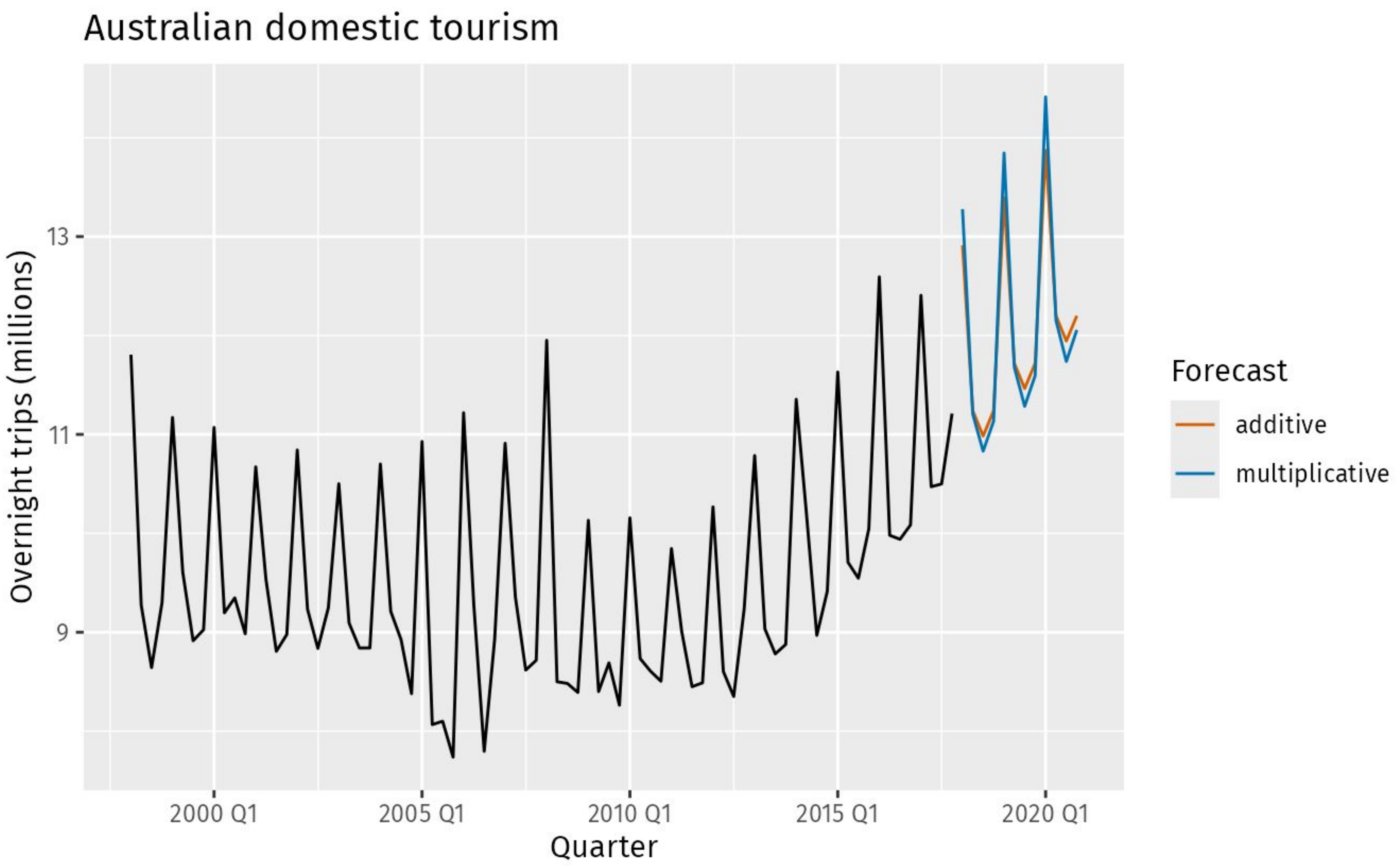
Aditivo:

$$
\begin{aligned}
\hat{y}_{t+h|t} &= \ell_t + hb_t + s_{t+h-m(k+1)} \\
\ell_t &= \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\
b_t &= \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1} \\
s_t &= \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m},
\end{aligned}
$$

Multiplicativo:

$$
\begin{aligned}
\hat{y}_{t+h|t} &= (\ell_t + hb_t)s_{t+h-m(k+1)} \\
\ell_t &= \alpha\frac{y_t}{s_{t-m}} + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\
b_t &= \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1} \\
s_t &= \gamma\frac{y_t}{(\ell_{t-1} + b_{t-1})} + (1-\gamma)s_{t-m}.
\end{aligned}
$$

Hiperparâmetros $0 \leq \alpha,\ \beta^*, \gamma \leq 1,\ \ \ell_0, b_0$, e s_0.


Australian domestic tourism
Overnight trips (millions)
13
11
9
2000 Q1
2005 Q1
2010 Q1
2015 Q1
2020 Q1
Quarter
Forecast
additive
multiplicative

Classificação de métodos de suavização exponencial

Trend Component	Seasonal Component		
	N	A	M
	(None)	(Additive)	(Multiplicative)
N (None)	(N,N)	(N,A)	(N,M)
A (Additive)	(A,N)	(A,A)	(A,M)
A_d (Additive damped)	(A_d,N)	(A_d,A)	(A_d,M)

Modelos ARIMA

Série **estacionária** é aquela em que as propriedade estatísticas não dependem do momento de observação. Os modelos a seguir geralmente são restritos a séries estacionárias.

Transformações (como a logaritmíca) podem estabilizar a variância da série e **diferenciações** podem estabilizar a média da série, eliminando tendência e sazonalidade.

1° ordem ordinária: $y'_t = y_t - y_{t-1}$

2° ordem ordinária: $y''_t = y'_t - y'_{t-1}$

1° ordem sazonal: $y'_t = y_t - y_{t-m}$

Modelo autoregressivo ou AR(p)

Previsão é uma combinação linear dos valores passados (regressão com a série x sendo a série y deslocada).

AR(p): $y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t$

p é a ordem do modelo e ε□ é ruído branco.

Modelo de médias móveis ou MA(p)

Ao invés de valores passados da série, usa erros de previsões passadas. Não é exatamente uma regressão por que os valores dos erros não foram observados.

MA(q): $$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}$$

q é a ordem do modelo e ε_t é ruído branco.

Modelo ARIMA não sazonal

Combina diferenciação com os modelos autoregressivo e de médias móveis (ARIMA = AutoRegressive Integrated Moving Average).

ARIMA(p,d,q):

$$y'_t = c + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

d é a ordem da diferenciação.

Modelo ARIMA não sazonal

As previsões a longo prazo tendem a…

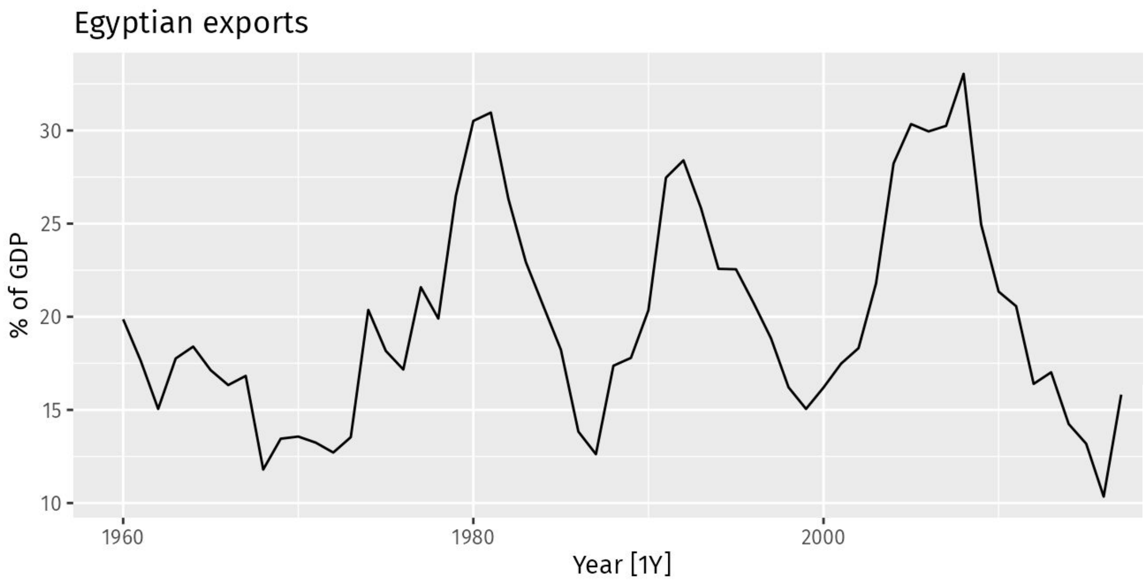
	d = 0	d = 1	d = 2
c = 0	0	constante não nula	uma reta
c != 0	média das observações	uma reta	uma curva quadrática

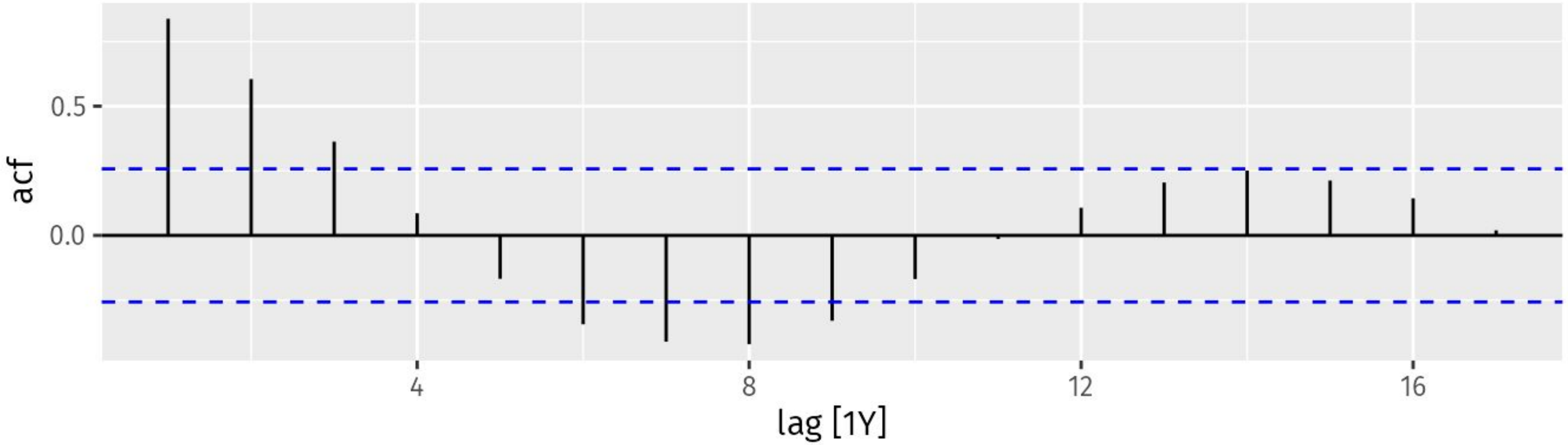
Quanto maior o valor de d mais rápido cresce o intervalo de previsão.

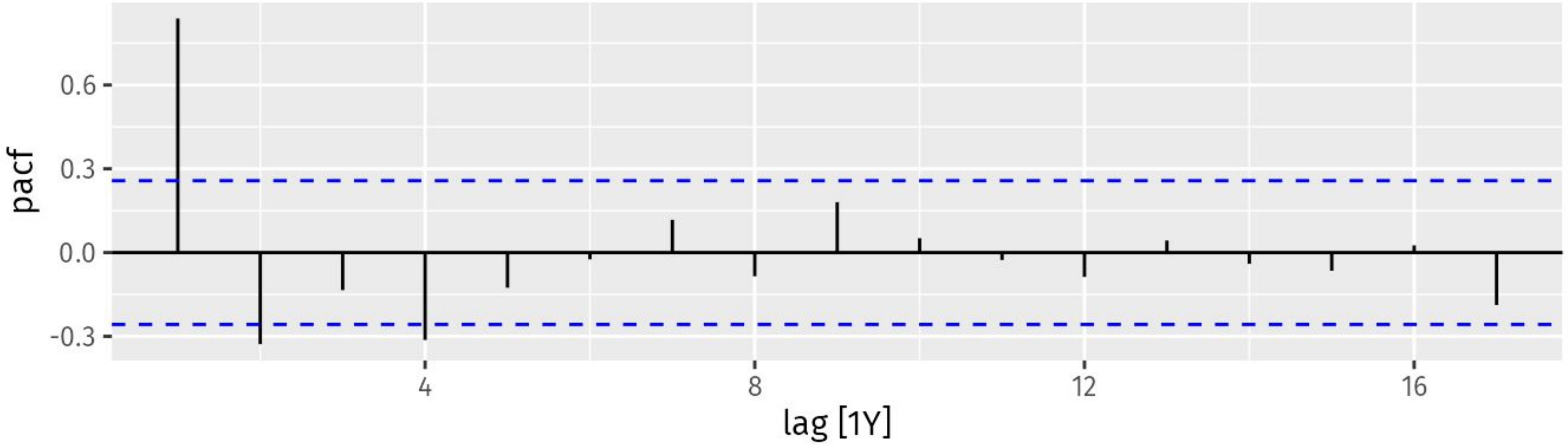
Modelo ARIMA não sazonal

Pode-se usar os gráficos ACF e PACF (Partial AutoCorrelation Function) para escolher p ou q.

Autocorrelação parcial mede a autocorrelação de y_t e y_{t-k} removendo os efeitos dos deslocamentos 1, 2, … , k-1. O k-ésimo coeficiente de autocorrelação parcial é igual a ϕ_k no modelo AR(k).


Egyptian exports
% of GDP
Year [1Y]
10
15
20
25
30
1960
1980
2000


acf
0.5
0.0
4
8
12
16
lag [1Y]


pacf
0.6
0.3
0.0
-0.3
4
8
12
16
lag [1Y]

Modelo ARIMA não sazonal

É adequado usar ARIMA(p,d,0) se:

- ACF é exponencialmente decadente ou senoidal;
- PACF tem um pico em p e não mais depois de p;

É adequado usar ARIMA(0,d,q) se:

- PACF é exponencialmente decadente ou senoidal;
- ACF tem um pico em q e não mais depois de q;

No exemplo do slide anterior, pode-se usar ARIMA(4,d,0).

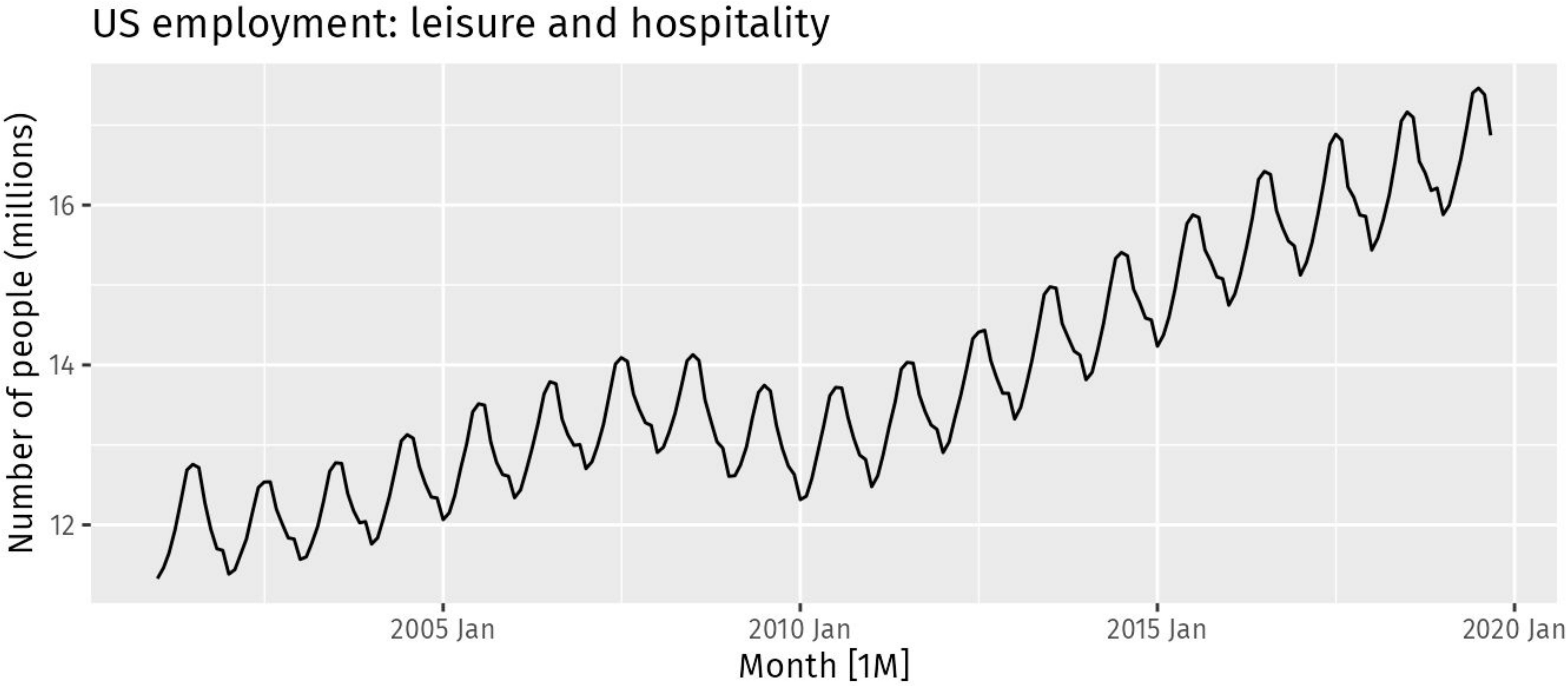
Modelo ARIMA não sazonal

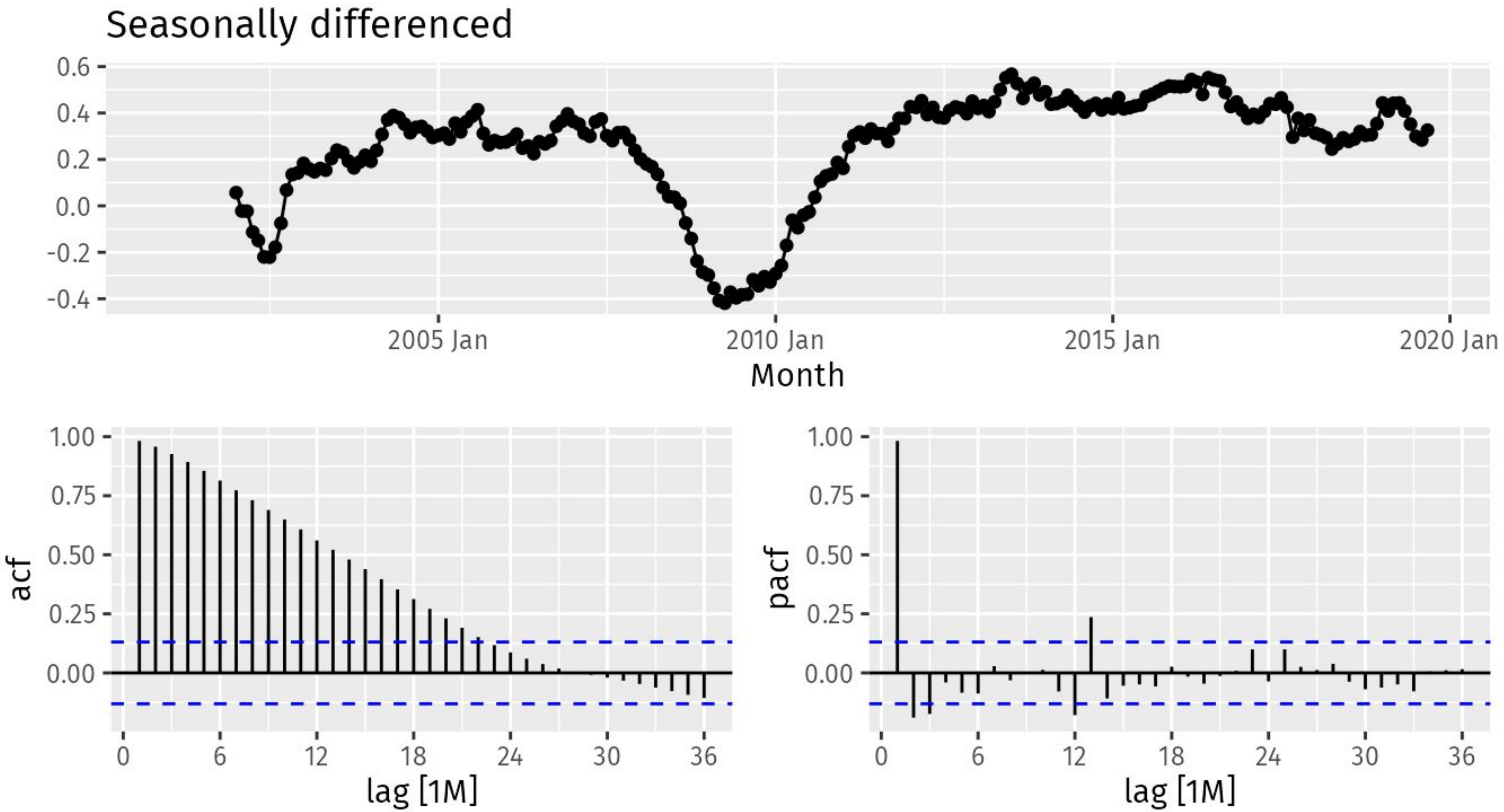
Para estimar o modelo ARIMA usa-se MLE (maximum likelihood estimation), que encontra valores para os parâmetros que maximizam a probabilidade de encontrar os dados observados.

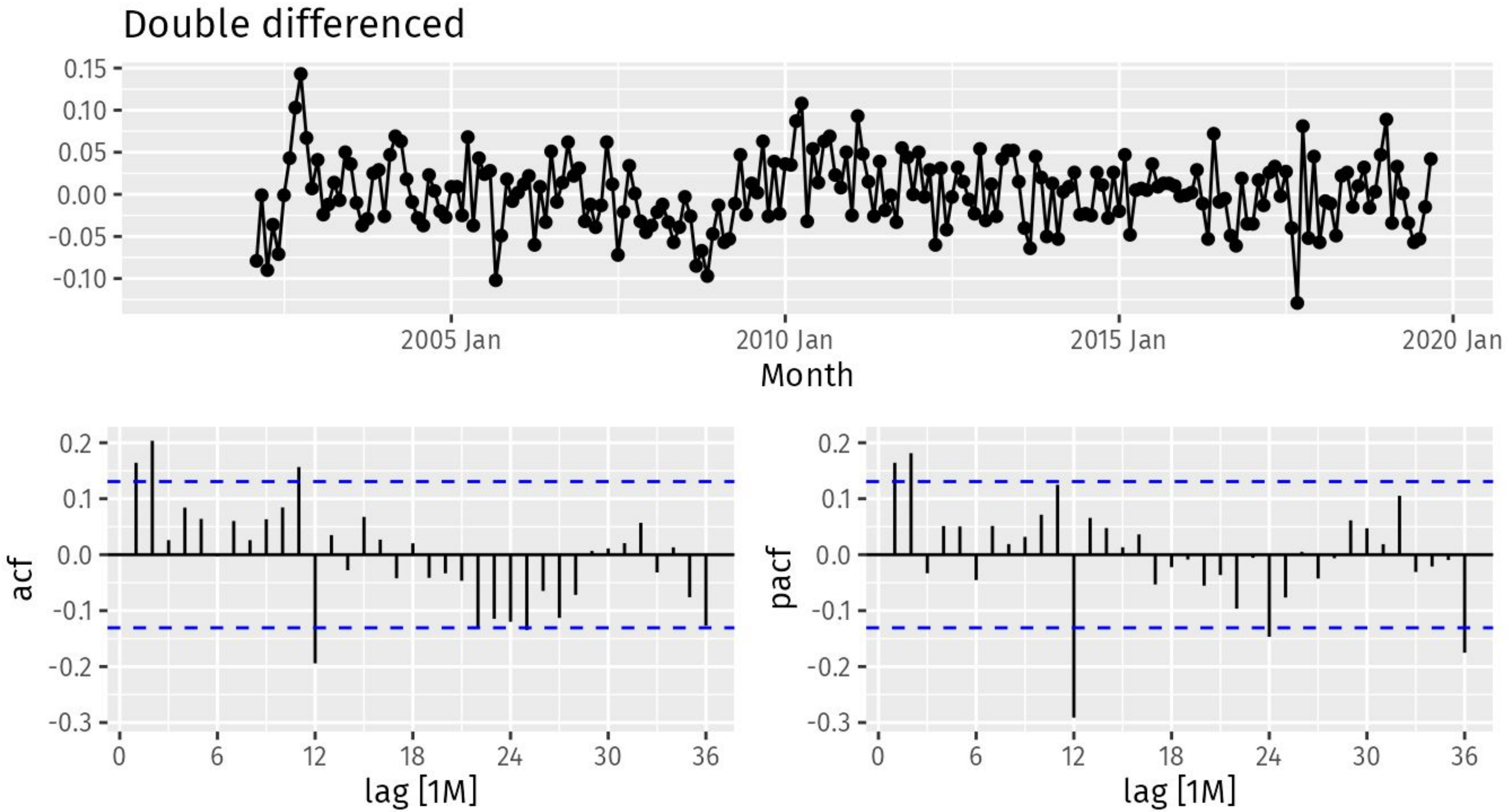
Modelo ARIMA sazonal

Para modelar observações com sazonalidade, incluí-se termos adicionais sazonais.

$$\text{ARIMA} \quad \underbrace{(p, d, q)}_{\substack{\uparrow \\ \text{Non-seasonal part} \\ \text{of the model}}} \quad \underbrace{(P, D, Q)_m}_{\substack{\uparrow \\ \text{Seasonal part} \\ \text{of the model}}}$$


US employment: leisure and hospitality
Number of people (millions)
16
14
12
2005 Jan
2010 Jan
2015 Jan
2020 Jan
Month [1M]


Seasonally differenced
0.6
0.4
0.2
0.0
-0.2
-0.4
2005 Jan
2010 Jan
2015 Jan
2020 Jan
Month
acf
1.00
0.75
0.50
0.25
0.00
0
6
12
18
24
30
36
lag [1M]
pacf
1.00
0.75
0.50
0.25
0.00
0
6
12
18
24
30
36
lag [1M]


Double differenced
0.15
0.10
0.05
0.00
-0.05
-0.10
2005 Jan
2010 Jan
2015 Jan
2020 Jan
Month
acf
0.2
0.1
0.0
-0.1
-0.2
-0.3
0
6
12
18
24
30
36
lag [1M]
pacf
0.2
0.1
0.0
-0.1
-0.2
-0.3
0
6
12
18
24
30
36
lag [1M]

Obrigado!