时间序列分析与预测

第七讲

黄嘉平

深圳大学｜中国经济特区研究中心
粤海校区汇文楼办公楼 1510
课程网站 https://huangjp.com/TSAF/

1. 简单指数平滑法

1.1. 历史数据的权重

均值法：用历史数据的均值作为预测值的方法。其定义如下：

$$\hat{y}_{T+h|T} = \overline{y} = \frac{y_1 + \ldots + y_T}{T}$$

朴素法：用最终观测值作为预测值的方法。其定义如下：

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

在预测中，我们应该如何利用历史久远的数据？1920 年的观测值对明年的预测能起到多大的作用？2000 年的观测值呢？在这个问题上，均值法和朴素法各自代表了一种极端情况，均值法给每个历史数据平均分配权重，而朴素法把所有的权重都放在了最新的观测值上。

1950 年代，统计学家们提出了一系列基于对历史数据进行指数加权的预测方法，称为**指数平滑法（exponential smoothing）**。其中最简答的称为**简单指数平滑法（simple exponential smoothing/SES）**，其定义为

$$\hat{y}_{T+h|T} = \alpha y_T + \alpha(1-\alpha)y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + \ldots$$

其中 $0 < \alpha < 1$。我们称 α 为平滑参数（smoothing parameter）。

从定义可知，SES 是介于均值法和朴素法之间的方法，它给近期的观测值赋予更大的权重，给古老的观测值赋予更小的权重。如果我们将所有权重相加，可得

$$\alpha + \alpha(1-\alpha) + \alpha(1-\alpha)^2 + \ldots = \alpha \lim_{t\to\infty} \frac{1-(1-\alpha)^t}{1-(1-\alpha)} = \frac{\alpha}{\alpha} = 1$$

1.2. SES 的递归表达

虽然 SES 的定义中包含无穷多个观测值，但实际上我们只能利用 $y_1, \ldots y_T$，因此实际上需要在有限加权和的基础上附加一个修正项，即

$$\hat{y}_{T+h|T} = \sum_{j=0}^{T-1} \alpha(1-\alpha)^j y_{T-j} + L$$

下面我们推导 L 的表达式。

从定义可知，

$$\begin{aligned}
\hat{y}_{T+h|T} &= \alpha y_T + \alpha(1-\alpha) y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + \ldots \\
&= \alpha y_T + (1-\alpha)\left[\alpha y_{T-1} + \alpha(1-\alpha) y_{T-2} + \alpha(1-\alpha)^2 y_{T-3} \ldots\right] \\
&= \alpha y_T + (1-\alpha)\hat{y}_{T|T-1}
\end{aligned}$$

1.2. SES 的递归表达

令 $h = 1$，并将 T 替换为 $t = 1, \ldots, T$, 可得

$$
\begin{aligned}
\hat{y}_{2|1} &= \alpha y_1 + (1-\alpha)\ell_0 \quad \text{（ℓ_0 是对 y_1 的预测值，后面会对其进行估计）} \\
\hat{y}_{3|2} &= \alpha y_2 + (1-\alpha)\hat{y}_{2|1} \\
&\vdots \\
\hat{y}_{T+h|T} &= \alpha y_T + (1-\alpha)\hat{y}_{T|T-1}
\end{aligned}
$$

将每个等式依次代入其下方的等式，可得

$$
\begin{aligned}
\hat{y}_{3|2} &= \alpha y_2 + (1-\alpha)[\alpha y_1 + (1-\alpha)\ell_0] = \alpha y_2 + \alpha(1-\alpha)y_1 + (1-\alpha)^2\ell_0 \\
&\vdots \\
\hat{y}_{T+h|T} &= \sum_{j=0}^{T-1} \alpha(1-\alpha)^j y_{T-j} + (1-\alpha)^T \ell_0
\end{aligned}
$$

从递归表达 $\hat{y}_{t+1|t} = \alpha y_t + (1-\alpha)\hat{y}_{t|t-1}$ 出发，最终我们获得了基于历史数据 $y_1, \ldots y_T$ 的 SES 定义

$$\hat{y}_{T+h|T} = \sum_{j=0}^{T-1} \alpha(1-\alpha)^j y_{T-j} + (1-\alpha)^T \ell_0$$

此定义中的 α 和 ℓ_0 为未知参数，需要利用数据进行估计。常用的方法是最小二乘法，也就是找出使残差 $e_t = y_t - \hat{y}_{t|t-1}$ 的平方和最小的参数值，表达为下面的最小化问题

$$\min_{\alpha, \ell_0} \sum_{t=1}^{T} e_t^2$$

1.3. SES 的成份表达式

指数平滑模型的另一种表达方法是将其分解成不同的成份，然后针对每个成份进行递归表达。在 SES 中仅包含一个成份，我们称之为水平（level）。后面我们还将加入趋势和季节成份。

SES 的成份表达式为

$$
\begin{aligned}
\text{预测方程} \quad \hat{y}_{t+h|t} &= \ell_t \\
\text{平滑方程} \quad \ell_t &= \alpha y_t + (1-\alpha)\ell_{t-1}
\end{aligned}
$$

成份表达式的优点是可以明确地体现出预测的逻辑，同时可以针对不同的成份设定不同的方程。后面介绍的其他种类的指数平滑模型都使用成份表达式。

预测方程强调了 SES 中不包含趋势和季节成份，因此适用范围也比较窄。

前面介绍的模型中都不包含随机成份，导致这类模型虽然可以进行点预测，但是无法进行区间预测。换句话说，由于缺少随机误差项，使得前面介绍的模型无法获得预测值的分布信息。

那么如何加入误差项呢？主要分为加法误差和乘法误差两种。

首先我们将 SES 的平滑方程改写为以下形式

$$\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1} = \ell_{t-1} + \alpha(y_t - \ell_{t-1}) = \ell_{t-1} + \alpha e_t$$

其中 e_t 是 t 时间点上的残差。如果我们将 e_t 替换为误差项 ε_t 并假设它服从独立正态分布 $\mathrm{NID}(0, \sigma^2)$，就得到了**加法误差模型**

$$\begin{aligned} y_t &= \ell_{t-1} + \varepsilon_t \\ \ell_t &= \ell_{t-1} + \alpha\varepsilon_t \end{aligned}$$

如果我们考虑相对误差

$$\varepsilon_t = \frac{y_t - \hat{y}_{t|t-1}}{\hat{y}_{t|t-1}} = \frac{y_t - \ell_{t-1}}{\ell_{t-1}}$$

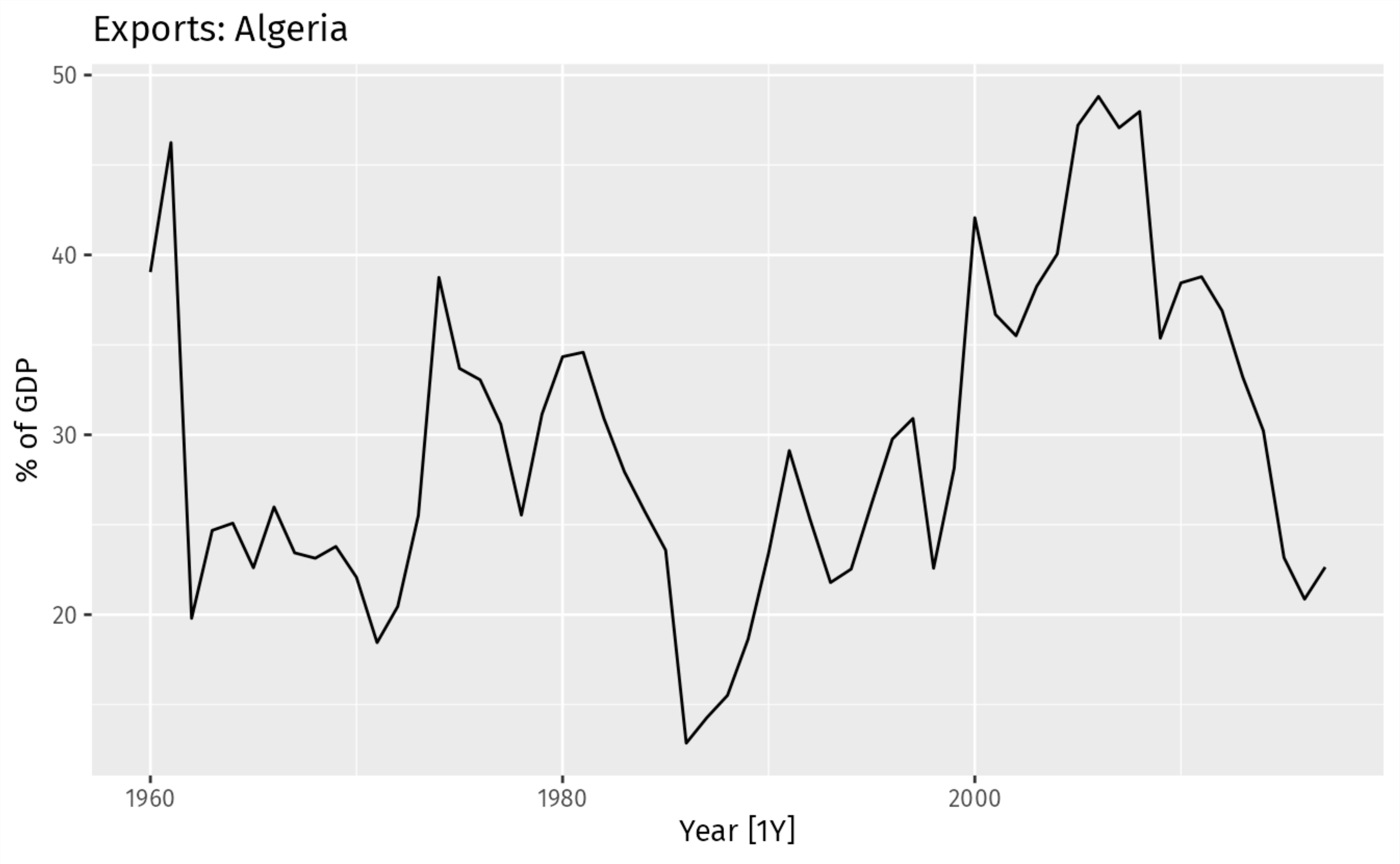
则残差和误差项的关系变为 $e_t = \ell_{t-1}\varepsilon_t$，并得到**乘法误差模型**

$$\begin{aligned} y_t &= \ell_{t-1}(1 + \varepsilon_t) \\ \ell_t &= \ell_{t-1}(1 + \alpha\varepsilon_t) \end{aligned}$$

加法误差模型和乘法误差模型给出的点预测值都一样，但是预测区间不同。加法误差模型更适合误差大小不随时间变化的序列，乘法误差模型则正相反。

加入误差的模型也称作**状态空间模型（state space model）**。除了能进行区间预测，这类统计学模型的另一个好处是允许使用 AICc 等测度进行模型选择。

1.5. `global_economy` 数据集中的阿尔及利亚出口额



1.5. `global_economy` 数据集中的阿尔及利亚出口额

拟合指数平滑模型的函数是 `ETS()`。下面的代码利用加法误差 SES 模型对阿尔及利亚的出口额序列进行了拟合。

```
algeria_economy <- global_economy |>
  filter(Country == "Algeria")
fit <- algeria_economy |>
  model(ETS(Exports ~ error("A") + trend("N") + season("N")))
```

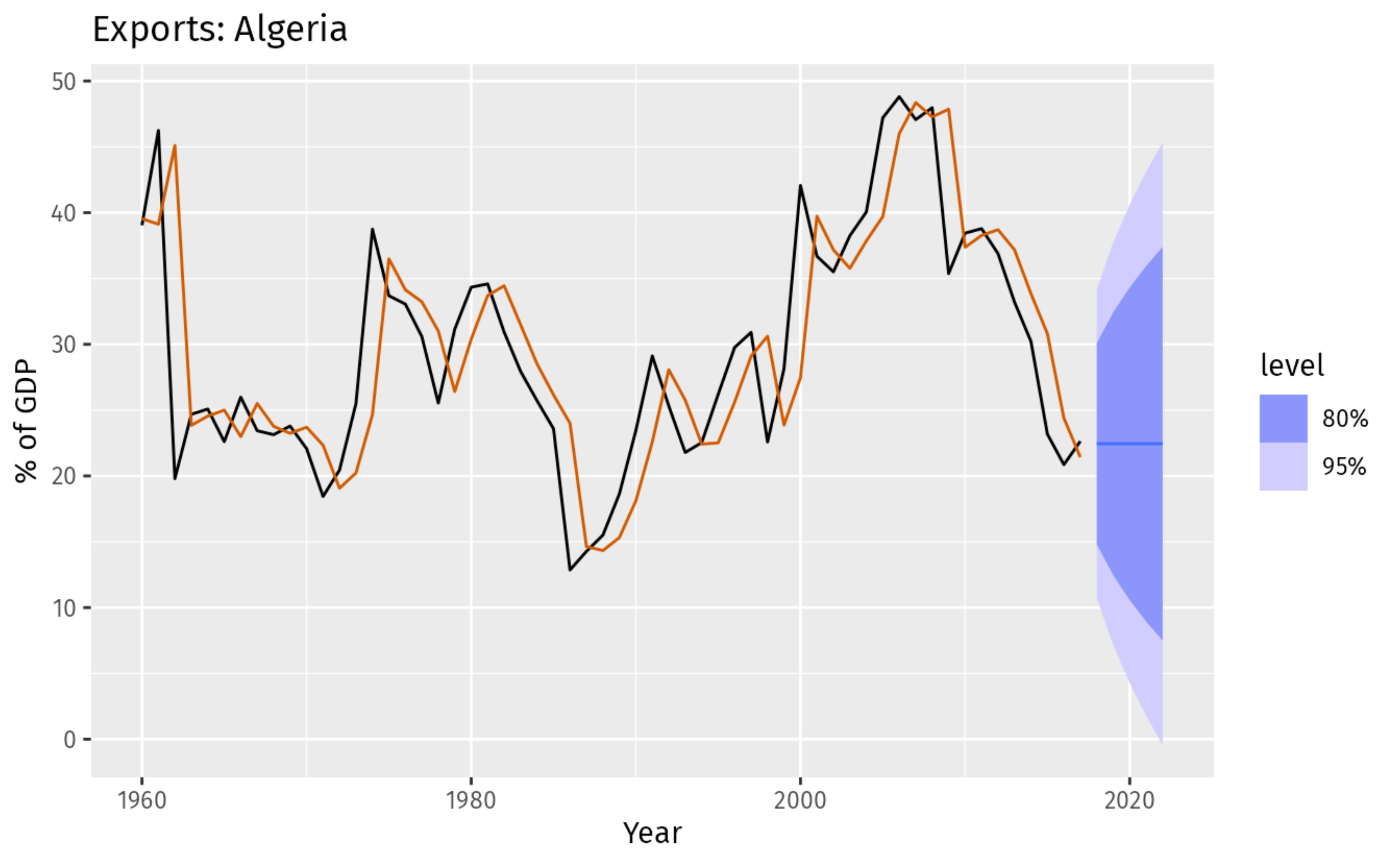
`ETS()` 中的模型部份由误差（error）、趋势（trend）和季节（season）三项组成，其中误差项可以指定为加法（`"A"`）或乘法（`"M"`），而趋势和季节项被指定为 `"N"` 则代表不包含该项，因此这是一个加法误差 SES 模型。

```r
report(fit)
```

```
Series: Exports
Model: ETS(A,N,N)
  Smoothing parameters:
    alpha = 0.8399875

  Initial states:
   l[0]
 39.539

  sigma^2:  35.6301

     AIC     AICc      BIC
446.7154 447.1599 452.8968
```

参数的估计值分别为 $\hat{\alpha} \approx 0.84, \hat{\ell}_0 \approx 35.5$。同时也给出了误差项方差的估计值和 AICc 等测度的计算结果。

1.5. `global_economy` 数据集中的阿尔及利亚出口额



2. 指数平滑法的分类和模型选择

2.1. Holt 线性趋势模型

Holt (1957) 在 SES 中加入了趋势项，其加法误差模型为

$$
\begin{aligned}
y_t &= \ell_{t-1} + b_{t-1} + \varepsilon_t \\
\ell_t &= \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t \\
b_t &= b_{t-1} + \beta\varepsilon_t
\end{aligned}
$$

对应的成份表达式为

$$
\begin{aligned}
\text{预测方程} \quad \hat{y}_{t+h|t} &= \ell_t + h b_t \\
\text{水平方程} \quad \ell_t &= \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\
\text{趋势方程} \quad b_t &= \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*) b_{t-1}, \quad \beta^* = \beta/\alpha
\end{aligned}
$$

Holt, C. C. (1957). *Forecasting seasonals and trends by exponentially weighted averages* (ONR Memorandum No. 52). Carnegie Institute of Technology, Pittsburgh USA. Reprinted in the *International Journal of Forecasting*, 2004.

2.2. 趋势衰减模型（damped trend model）

Holt 模型的趋势为线性，且会一直保持不变。这在长期预测中效果并不好，因为大多数序列的趋势都会随时间变化。Gardner & McKenzie (1985) 在该模型的基础上允许趋势项的影响随时间的流逝而越来越弱。

令 ϕ 为衰减参数（damping parameter，教科书中文版中译为阻尼参数），则趋势衰减模型的成份表达式为

$$
\begin{aligned}
\text{预测方程} \quad \hat{y}_{t+h|t} &= \ell_t + (\phi + \phi^2 + \dots + \phi^h) b_t \\
\text{水平方程} \quad \ell_t &= \alpha y_t + (1-\alpha)(\ell_{t-1} + \phi b_{t-1}) \\
\text{趋势方程} \quad b_t &= \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}
\end{aligned}
$$

虽然衰减参数的理论范围是 $0 < \phi < 1$，但通常将它限制在 $0.8 \le \phi \le 0.98$。

Gardner, E. S., & McKenzie, E. (1985). Forecasting trends in time series. *Management Science*, 31(10), 1237–1246.

2.3. 澳大利亚人口数据

```
aus_economy <- global_economy |>
  filter(Code == "AUS") |>
  mutate(Pop = Population / 1e6)

aus_economy |>
  model(
    `Holt's method` = ETS(Pop ~ error("A") +
                          trend("A") + season("N")),
    `Damped Holt's method` = ETS(Pop ~ error("A") +
                                 trend("Ad", phi = 0.9) +
                                 season("N"))
        # trend 中的 Ad 代表加法衰减。这里指定了衰减参数的值，而不是进行估计
  ) |>
  tidy() # 以 tibble 的形式显示参数的估计结果，和 coef() 类似
```

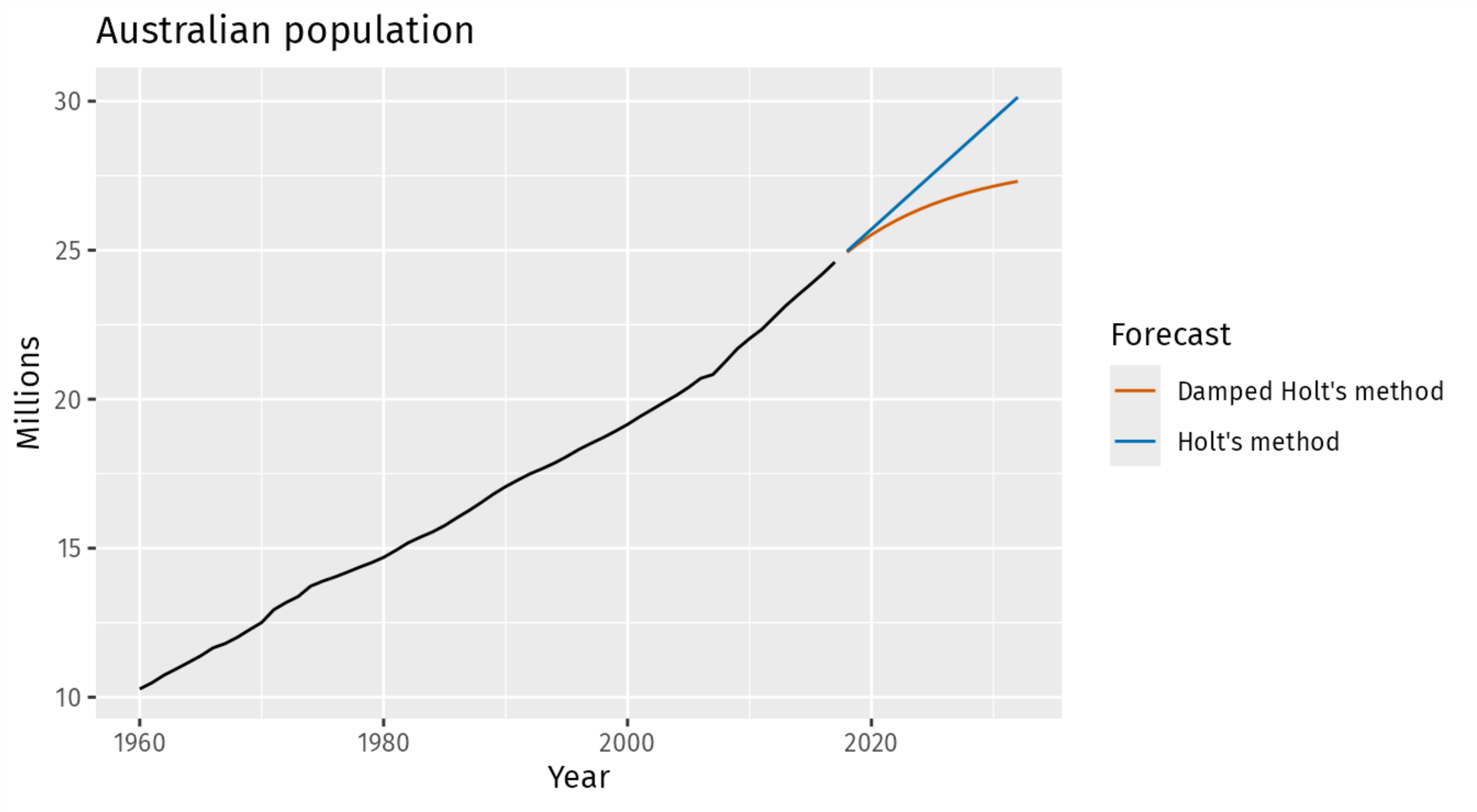
2.3. 澳大利亚人口数据

```
# A tibble: 9 × 4
  Country   .model                term  estimate
  <fct>     <chr>                 <chr>    <dbl>
1 Australia Holt's method         alpha    1.00
2 Australia Holt's method         beta     0.327
3 Australia Holt's method         l[0]    10.1
4 Australia Holt's method         b[0]     0.222
5 Australia Damped Holt's method alpha    0.868
6 Australia Damped Holt's method beta     0.817
7 Australia Damped Holt's method phi      0.9
8 Australia Damped Holt's method l[0]    10.0
9 Australia Damped Holt's method b[0]     0.273
```

注意这里给出的是 β 的估计值（对应误差模型），不是成份表达式中的 β^*。

（如果对 ϕ 进行估计，则估计值为 0.98）

2.3. 澳大利亚人口数据



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

2.4. Holt-Winters 趋势季节模型

Winters (1960) 在 Holt 线性趋势模型的基础上加入了季节项。季节项也可以分为加法和乘法两种，其中加法模型的成份表达式如下

$$
\begin{aligned}
\text{预测方程} \quad & \hat{y}_{t+h|t} = \ell_t + hb_t + s_{t+h-m(k+1)} \\
\text{水平方程} \quad & \ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\
\text{趋势方程} \quad & b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1} \\
\text{季节性方程} \quad & s_t = \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m} \\
& \quad = \gamma^*(y_t - \ell_t) + (1-\gamma^*)s_{s-m}, \quad \gamma^* = \gamma/(1-\alpha)
\end{aligned}
$$

这里的 m 是季节性周期，k 是 $(h-1)/m$ 的整数部份，这确保预测方程中的季节项来自数据中的最后一年。

Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science*, 6(3), 324–342.

2.5. 指数平滑模型的分类

依照误差项、趋势项、季节项的不同建模方式，指数平滑模型可以包含 18 种不同的设定：

加法误差 (A)				乘法误差 (M)			
趋势	季节			趋势	季节		
	N	A	M		N	A	M
N	(A,N,N)	(A,N,A)	(A,N,M)	N	(M,N,N)	(M,N,A)	(M,N,M)
A	(A,A,N)	(A,A,A)	(A,A,M)	A	(M,A,N)	(M,A,A)	(M,A,M)
Ad	(A,Ad,N)	(A,Ad,A)	(A,Ad,M)	Ad	(M,Ad,N)	(M,Ad,A)	(M,Ad,M)

`ETS()` 函数可以指定其中任意一种设定。例如 ETS(M,Ad,A) 对应的命令是

```
ETS(Y ~ error("M") + trend("Ad") + season("A"))
```

也可以用 `ETS(Y)` 让函数自动选择最优模型，这种情况下采用的判断标准是 AICc。

2.6. 澳大利亚国内旅行数据

`tourism` 数据集包含了澳大利亚国内多日旅行的季度累计次数数据

```
aus_holidays <- tourism |>
  filter(Purpose == "Holiday") |> # 限定旅行目的为 Holiday
  summarise(Trips = sum(Trips)/1e3) # 对不同地区的数据进行加总

fit <- aus_holidays |>
  model(ETS(Trips)) # 自动选择模型

report(fit) # 显示拟合结果
```

2.6. 澳大利亚国内旅行数据

```
Series: Trips
Model: ETS(M,N,A)
  Smoothing parameters:
    alpha = 0.3484054
    gamma = 0.0001000018

  Initial states:
      l[0]       s[0]      s[-1]      s[-2]    s[-3]
  9.727072 -0.5376106 -0.6884343 -0.2933663 1.519411

  sigma^2:  0.0022

      AIC     AICc      BIC
226.2289 227.7845 242.9031
```

`ETS()` 函数选择了 ETS(M,N,A) 模型，即乘法误差、无趋势、加法季节性。参数包括 α, γ, ℓ_0 以及季节项的四个初始值 $s_0, s_{-1}, s_{-2}, s_{-3}$。

2.6. 澳大利亚国内旅行数据

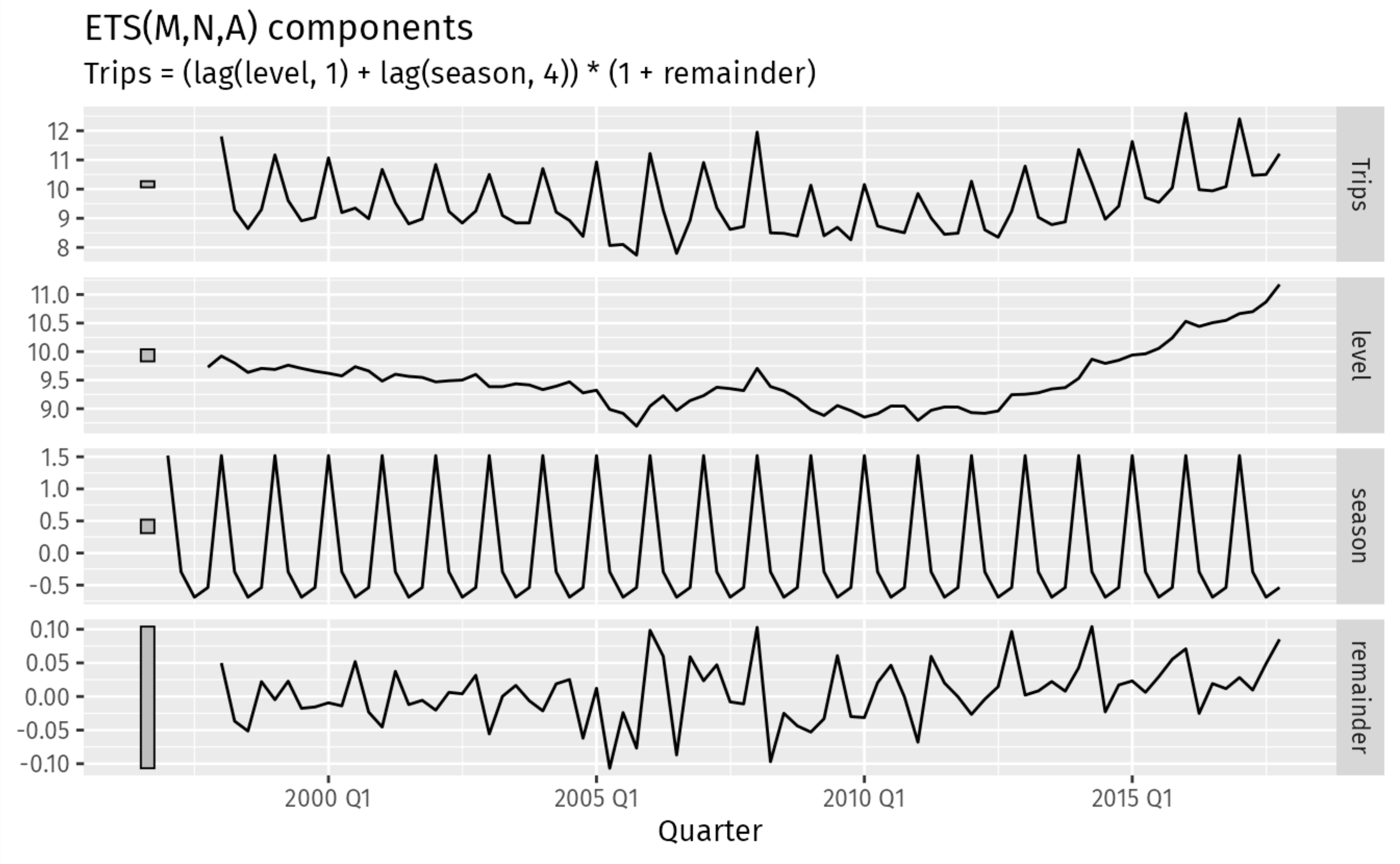
对拟合结果进行成份分解:

```
components(fit) |>
  autoplot() +
  labs(title = "ETS(M,N,A) components")
```

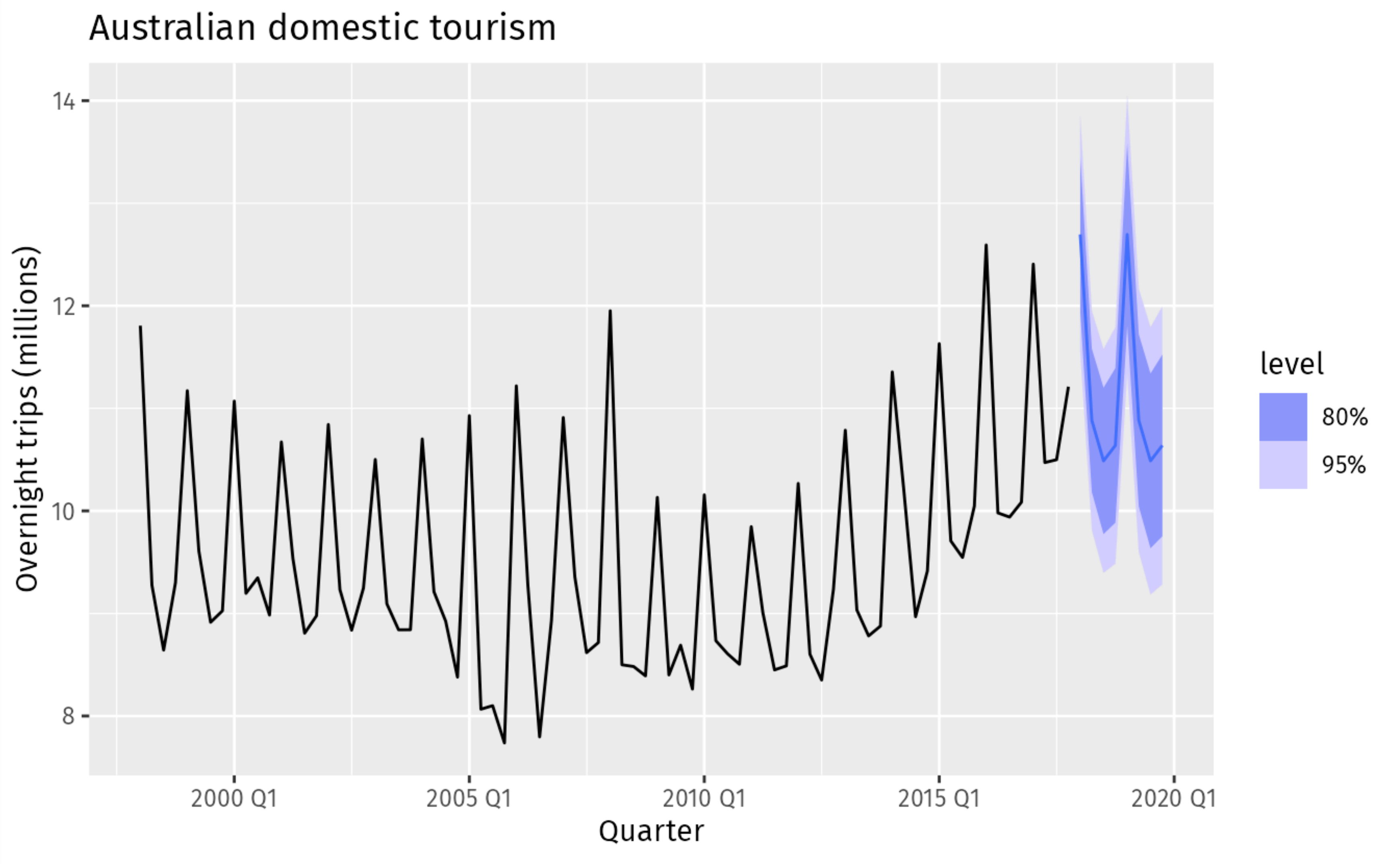
根据拟合后的模型进行未来两年（$h = 8$）的预测:

```
fit |>
  forecast(h = 8) |>
  autoplot(aus_holidays)+
  labs(title="Australian domestic tourism",
        y="Overnight trips (millions)")
```

2.6. 澳大利亚国内旅行数据



2.6. 澳大利亚国内旅行数据



3. 课后练习

3. 课后练习

- 学习教科书第 8 章（Exponential Smoothing）中的内容，并尝试在自己的电脑上复现书中的结果。

- 利用 `aus_holidays` 数据回答下列问题：

  1. 手动选择指数平滑模型的几个设定，分别进行拟合，并比较各自的 AICc。可利用 `glance()` 函数。
  2. 观察每个设定的成份分解图，你能总结出哪些异同?
  3. 挑选最优设定进行预测。一年后的点预测值是什么?