More exponential smoothing models

EC 361–001

Prof. Santetti
Spring 2024

Materials

Required readings:

- `Hyndman & Athanasopoulos, ch. 8`
    - sections 8.3—8.7.

Motivation

Motivation

Last time, we were introduced to **exponential smoothing**.

Recall that its **main idea** comprises using **weighted averages**, balancing out the *relative importance* of more recent and older observations.

Now, we move on by incorporating **seasonality** into exponential smoothing methods.

Exponential smoothing with seasonality

Exponential smoothing with seasonality

As we are trying to incorporate more **features** into our models, exponential smoothing methods have to include additional **parameters** and **equations**.

There are **two variations** to this method that differ in the **nature** of the seasonal component:

- The *additive*;
- And the *multiplicative* methods.

The **additive** method is preferred when the seasonal variations are **roughly constant** through the series.

The **multiplicative** method is preferred when the seasonal variations are **changing** proportional to the level of the series.

Exponential smoothing with seasonality

Let us start with the **additive** method.

The **component form** looks like the following:

$$
\begin{aligned}
\text{Forecast equation:} \;\; \hat{y}_{t+h|t} &= \ell_t + hb_t + s_{t+h-m(k+1)} \\
\text{Level equation:} \;\; \ell_t &= \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\
\text{Trend equation:} \;\; b_t &= \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1} \\
\text{Seasonal equation:} \;\; s_t &= \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m}
\end{aligned}
$$

where m denotes the period of seasonality.

Exponential smoothing with seasonality

Then, for **multiplicative** seasonality, we have:

$$\text{Forecast equation: } \hat{y}_{t+h|t} = (\ell_t + hb_t)s_{t+h-m(k+1)}$$

$$\text{Level equation: } \ell_t = \alpha \frac{y_t}{s_{t-m}} + (1-\alpha)(\ell_{t-1} + b_{t-1})$$

$$\text{Trend equation: } b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$$

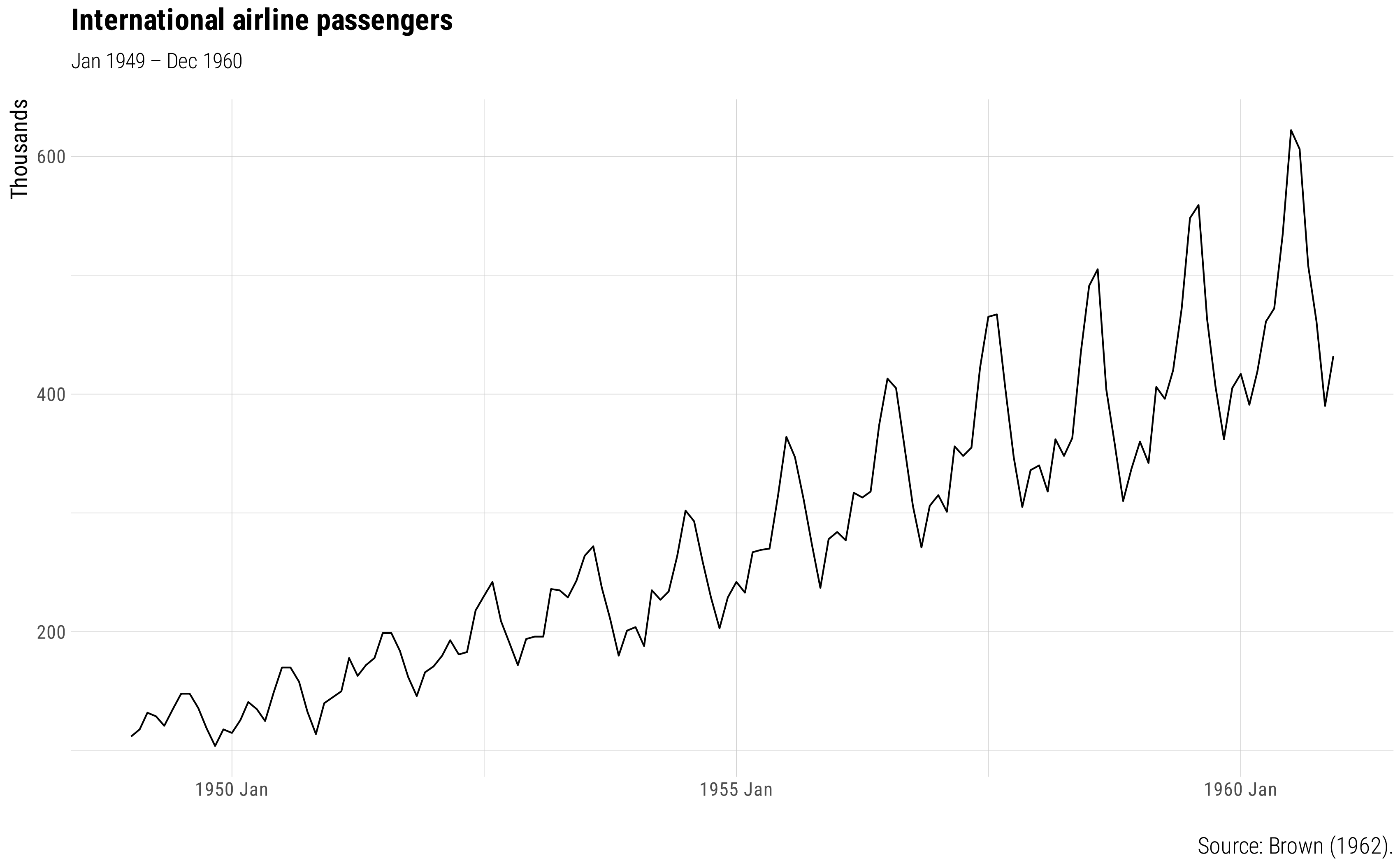
$$\text{Seasonal equation: } s_t = \gamma \frac{y_t}{(\ell_{t-1} + b_{t-1})} + (1-\gamma)s_{t-m}$$

Exponential smoothing with seasonality

Regardless of the method, we may also allow for a **damped trend**, as we have seen in the previous lecture.

An example

An example

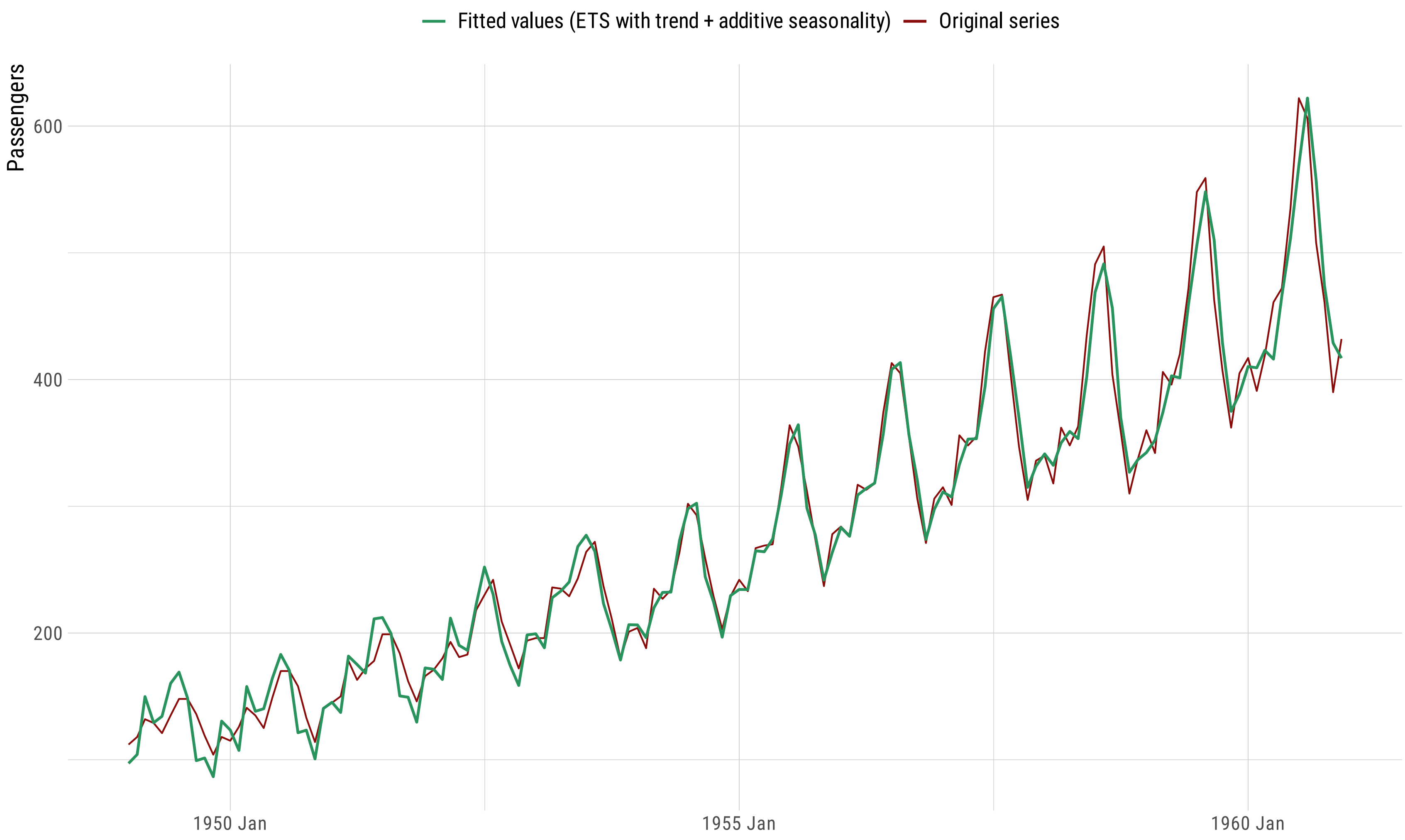

International airline passengers
Jan 1949 – Dec 1960
Thousands
600
400
200
1950 Jan
1955 Jan
1960 Jan
Source: Brown (1962).

An example

```
air_season_A_fit ← air_ts ▷
  model(additive_method = ETS(passengers ~ error("A") + trend("A") + season("A")))

air_season_A_fit ▷
  report()
```

```
#> Series: passengers
#> Model: ETS(A,A,A)
#>   Smoothing parameters:
#>     alpha = 0.9934804
#>     beta  = 0.0001911792
#>     gamma = 0.0005800325
#>
#>   Initial states:
#>      l[0]   b[0]      s[0]     s[-1]     s[-2]    s[-3]   s[-4]    s[-5]    s[-6]    s[-7]     s[-8]    s[-9]    s[-
#>  120.9608 1.3934 -29.18157 -54.38417 -20.71687 15.07266 65.1554 66.18464 33.58223 -4.23203 -8.094607 -3.82047 -34.33
#>
#>   sigma^2:  325.697
#>
#>      AIC     AICc      BIC
#> 1565.872 1570.729 1616.359
```

An example



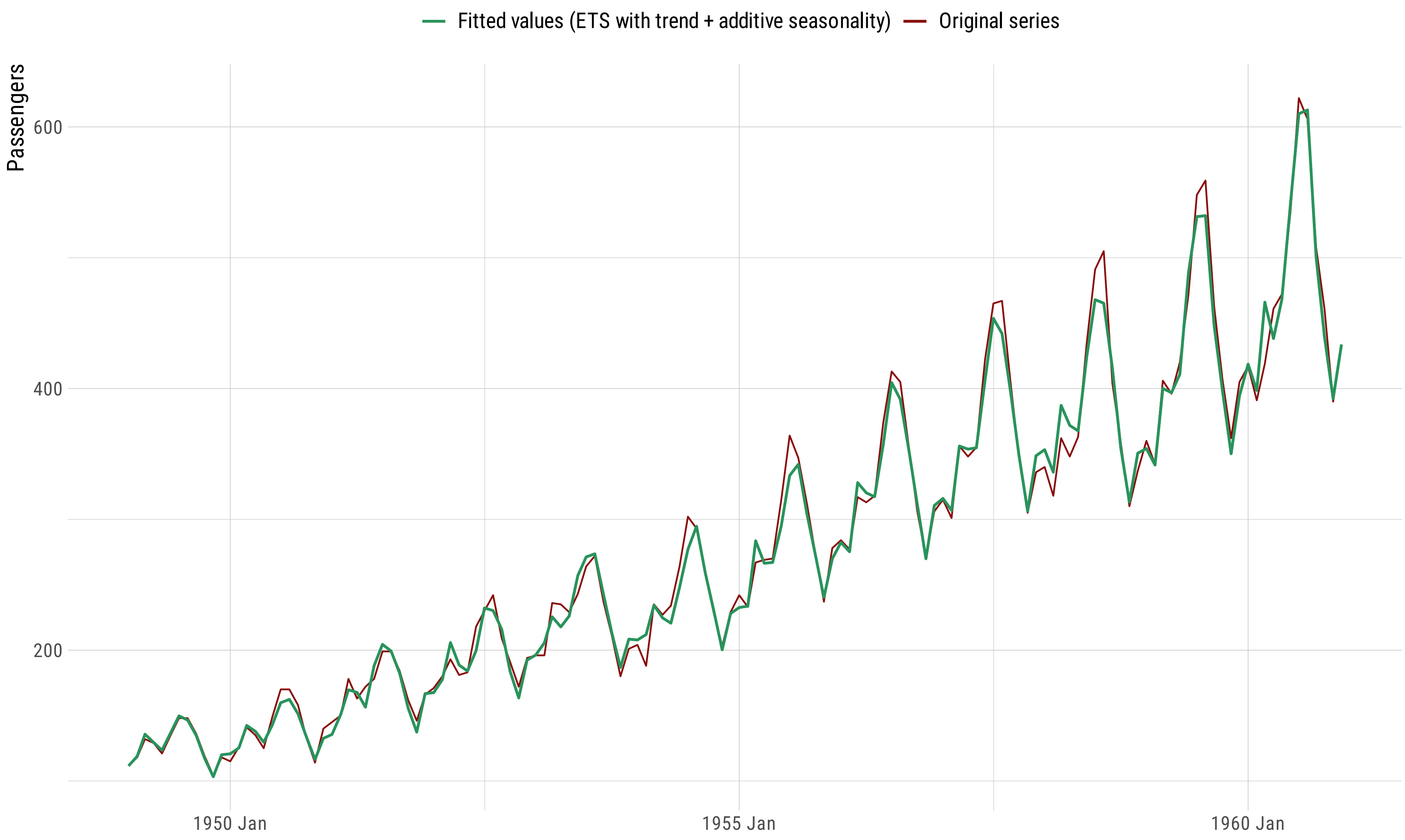
Fitted values (ETS with trend + additive seasonality)
Original series
Passengers
600
400
200
1950 Jan
1955 Jan
1960 Jan

An example

```r
air_season_M_fit <- air_ts |>
  model(additive_method = ETS(passengers ~ error("M") + trend("A") + season("M")))

air_season_M_fit |>
  report()
```

```
#> Series: passengers
#> Model: ETS(M,A,M)
#>   Smoothing parameters:
#>     alpha = 0.3949969
#>     beta  = 0.01070044
#>     gamma = 0.3995392
#>
#>   Initial states:
#>      l[0]     b[0]      s[0]     s[-1]    s[-2]    s[-3]    s[-4]    s[-5]    s[-6]     s[-7]    s[-8]    s[-9]    s
#>  122.3754 1.107367 0.9000411 0.7826691 0.901368 1.047618 1.153707 1.183031 1.083995 0.9786589 1.033162 1.080757 0.95
#>
#>   sigma^2:  0.0016
#>
#>      AIC     AICc      BIC
#> 1398.807 1403.664 1449.294
```

An example

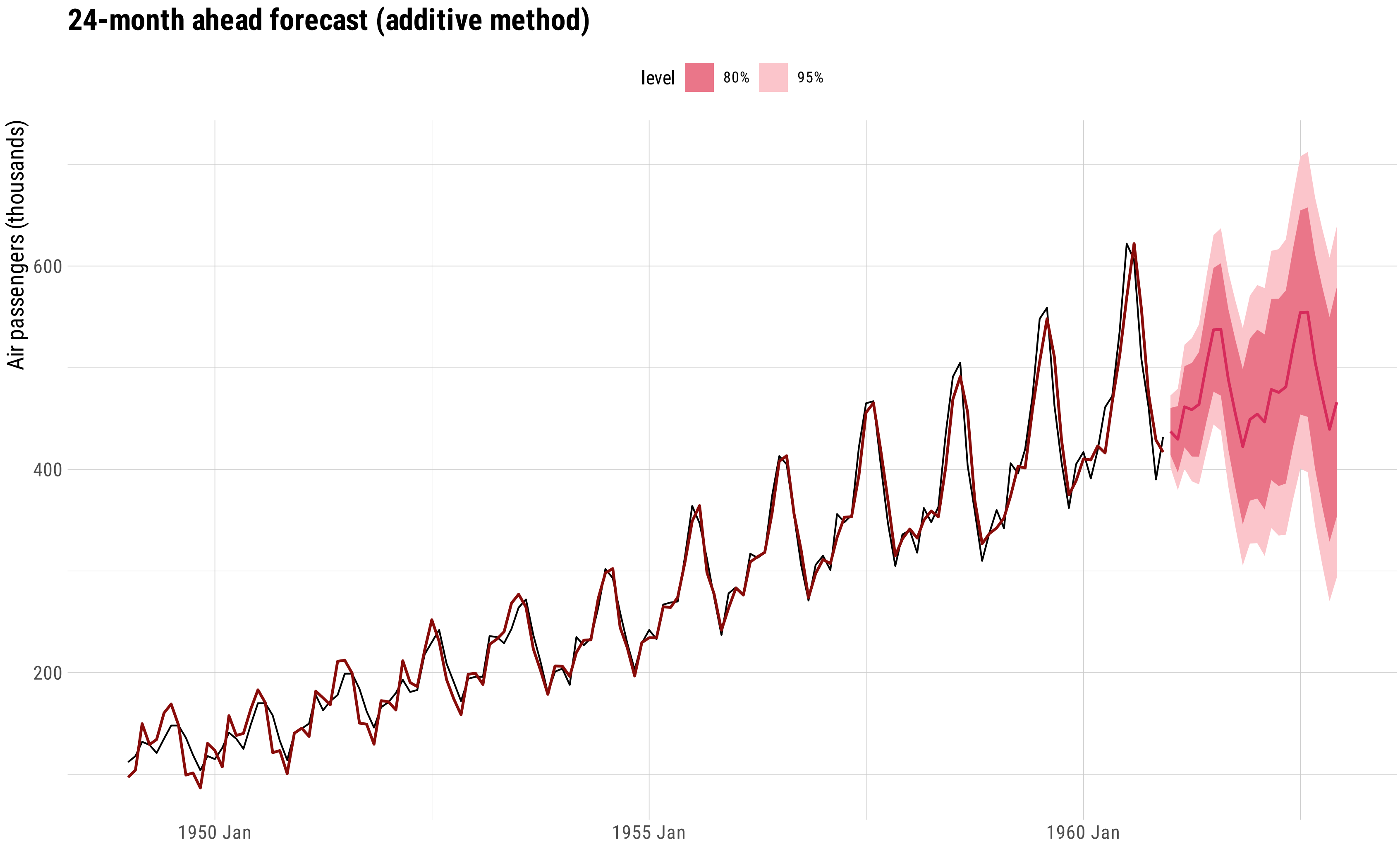


Fitted values (ETS with trend + additive seasonality)
Original series
Passengers
600
400
200
1950 Jan
1955 Jan
1960 Jan

An example

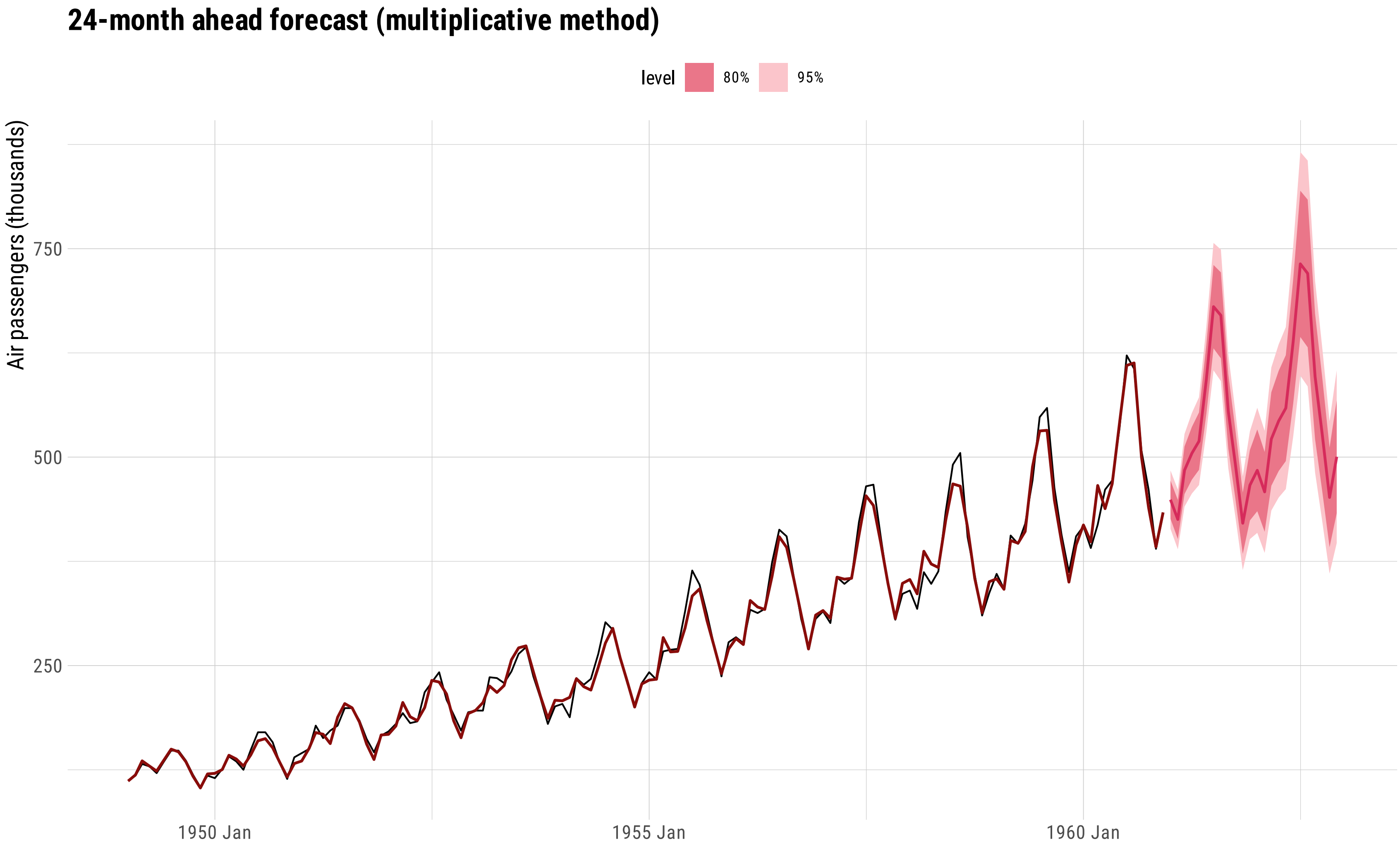
```
air_season_A_fc ← air_season_A_fit ▷
  forecast(h = 24)

air_season_M_fc ← air_season_M_fit ▷
  forecast(h = 24)
```

An example



24-month ahead forecast (additive method)
level
80%
95%
Air passengers (thousands)
600
400
200
1950 Jan
1955 Jan
1960 Jan

An example



24-month ahead forecast (multiplicative method)
level
80%
95%
Air passengers (thousands)
750
500
250
1950 Jan
1955 Jan
1960 Jan

ETS taxonomy

ETS taxonomy

As you may have noticed, different **variations** of exponential smoothing are possible.

It all *depends on* the features we would like to explore from our data, and also whether these features (especially **seasonality**) behaves in a *constant* or *non-constant* way.

Thus, it is possible to develop a **taxonomy** of exponential smoothing methods:

Table 8.5: A two-way classification of exponential smoothing methods.

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)

The textbook also includes all *mathematical expressions* for each model's **component form**.

Further thoughts

Further thoughts

The exponential smoothing methods we have studied so far only produce **point forecasts**.

The main issue concerning this fact is that these methods do not deliver **prediction intervals**.

This way, exponential smoothing **methods** are not the same as **statistical models**.

> *A statistical model is a stochastic (or random) data generating process that can produce an entire forecast distribution.*

Further thoughts

Given that we need a **forecast distribution** in order to produce a **full forecast**, we need to specify the **stochastic** part of our methods.

Recall, from a few lectures ago, our definition of **forecast errors**:

> A **forecast error** is the difference between an *observed* value and its *forecast*.

Further thoughts

Formally,

$$e_{T+h} = y_{T+h} - \hat{y}_{T+h\,|\,T}$$

Coming back to **Simple Exponential Smoothing** (SES) methods, we can turn the component form into a statistical model:

$$y_t = \ell_{t-1} + \varepsilon_t$$

$$\ell_t = \ell_{t-1} + \alpha\varepsilon_t$$

where $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$.

The same "*error correction*" approach can be adopted for the other exponential smoothing methods.

Model selection

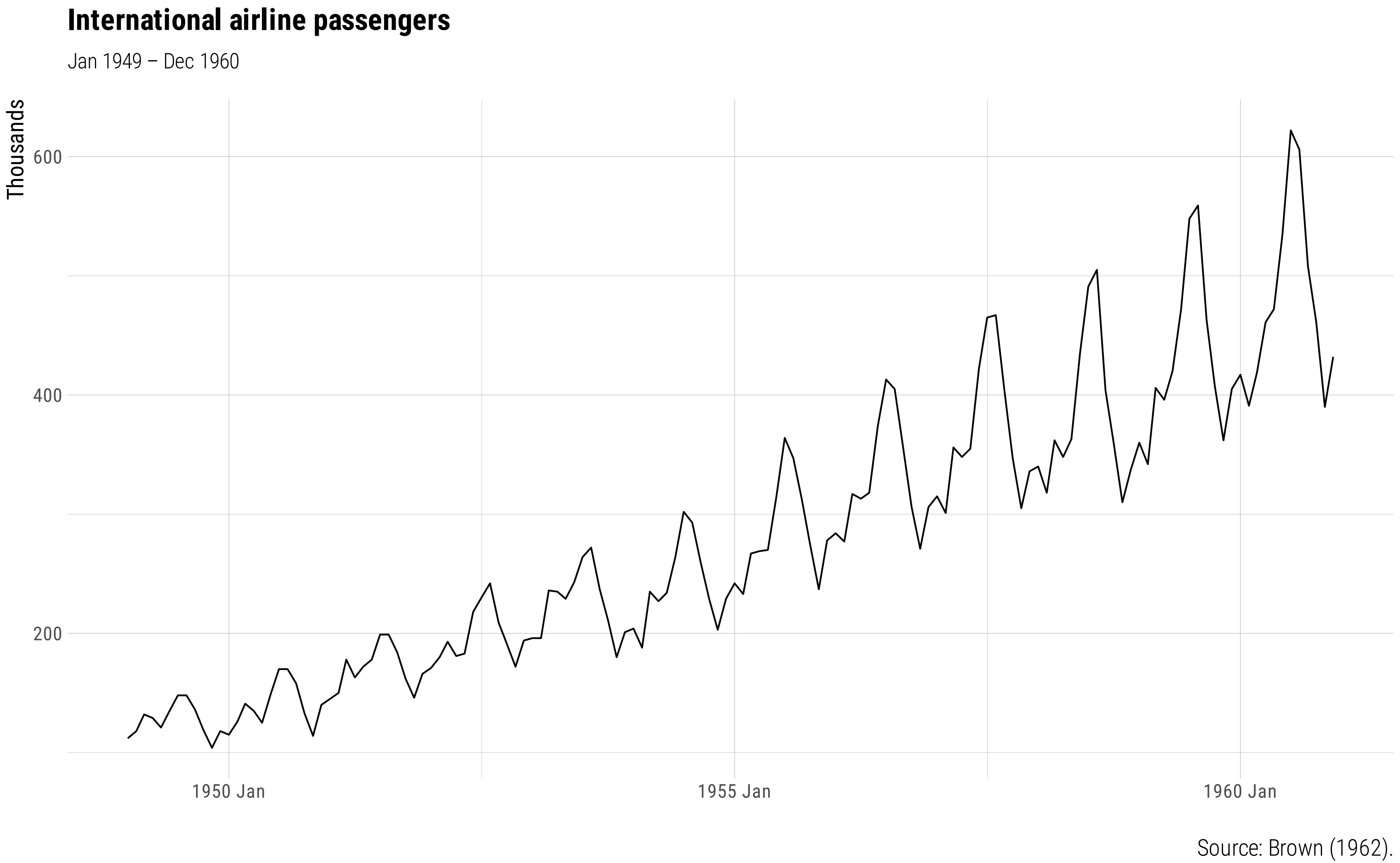
Model selection

Models estimated via exponential smoothing have **36 possible variations**, depending on how we specify its **E**rror, **T**rend, and **S**easonal components.

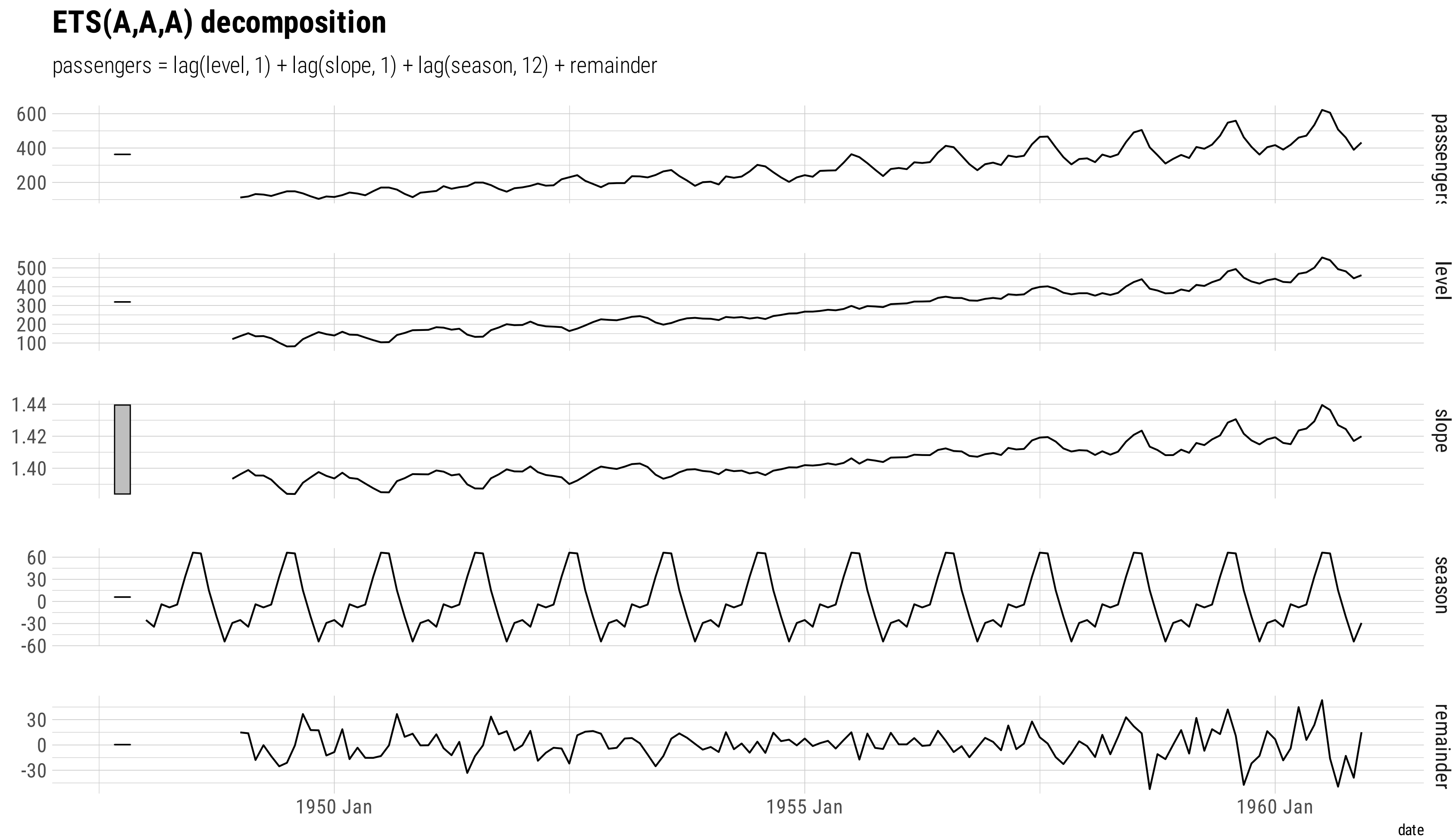
- *But which one should we choose*?

A good **starting point** is estimating the simplest possible model, accounting for the presence of **each feature** (or lack thereof).

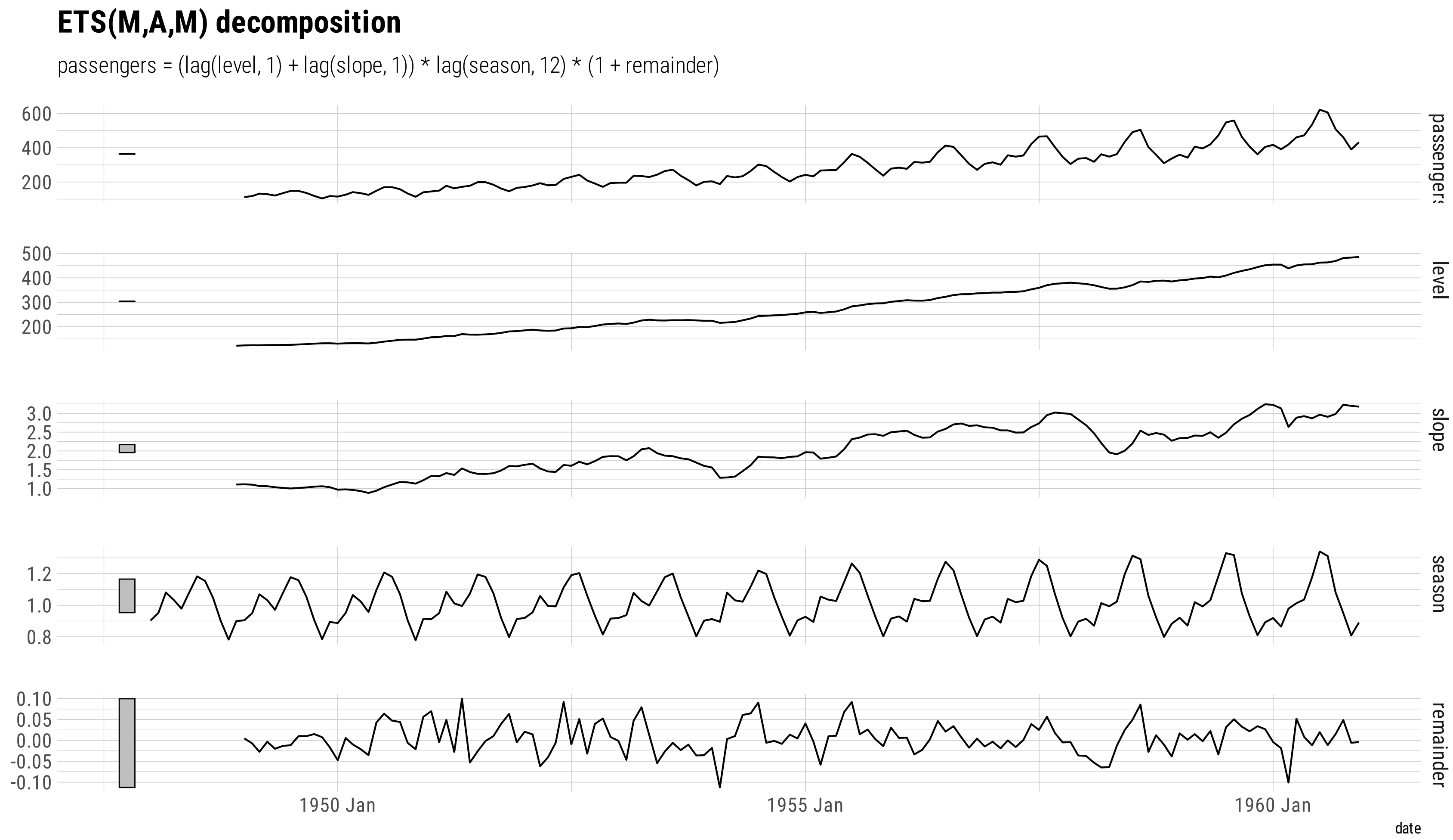
Model selection



```r
air_season_A_fit |>
  components() |>
  autoplot()
```



```
air_season_M_fit ▷
  components() ▷
  autoplot()
```



Model selection

When errors are of a **multiplicative** nature, simply minimizing the squared residuals is not sufficient for for model esimation.

Thus, a better approach is **Maximum Likelihood** (ML).

Maximum Likelihood implies that, given the model we have chosen, what is the probability of obtaining the data we observe (y_t)?

- Thus, a *large likelihood* is associated with a *good model*.

Model selection

Based on Maximum Likelihood, ETS models can make use of so-called **information criteria** for model comparison and selection.

One of them is the **Akaike Information Criterion (AIC):**

$$\mathrm{AIC} = -2\log(L) + 2k$$

where L is the likelihood of the model and k is the total number of *parameters* and *initial states* that have been estimated (including the residual variance).

For smaller samples, an alternative measure is the **corrected AIC ($\mathbf{AIC_c}$)**:

$$\mathrm{AIC_c} = \mathrm{AIC} + \frac{2k(k+1)}{T-k-1}$$

Model selection

Out of the 36 possible variations of ETS models, three of them usually produce **numerically unstable** estimations:

1. ETS(A, N, M);
2. ETS(A, A, M);
3. ETS(A, A_d, M).

Therefore, the above models are hardly used in practice.

Models with **multiplicative errors** are useful when the data are **strictly positive**, but are **not** numerically stable when the data contain *zeros* or *negative values*.

Model selection

```
air_season_A_fit ▷
  report()
```

```
#> Series: passengers
#> Model: ETS(A,A,A)
#>   Smoothing parameters:
#>     alpha = 0.9934804
#>     beta  = 0.0001911792
#>     gamma = 0.0005800325
#>
#>   Initial states:
#>      l[0]   b[0]      s[0]     s[-1]     s[-2]    s[-3]  [illegible]
#>  120.9608 1.3934 -29.18157 -54.38417 -20.71687 15.07266 [illegible]
#>
#>   sigma^2:  325.697
#>
#>      AIC     AICc      BIC
#> 1565.872 1570.729 1616.359
```

```
air_season_M_fit ▷
  report()
```

```
#> Series: passengers
#> Model: ETS(M,A,M)
#>   Smoothing parameters:
#>     alpha = 0.3949969
#>     beta  = 0.01070044
#>     gamma = 0.3995392
#>
#>   Initial states:
#> [illegible]
#> [illegible]
#>
#>   sigma^2:  0.0016
#>
#>      AIC     AICc      BIC
#> 1398.807 1403.664 1449.294
```

Model selection

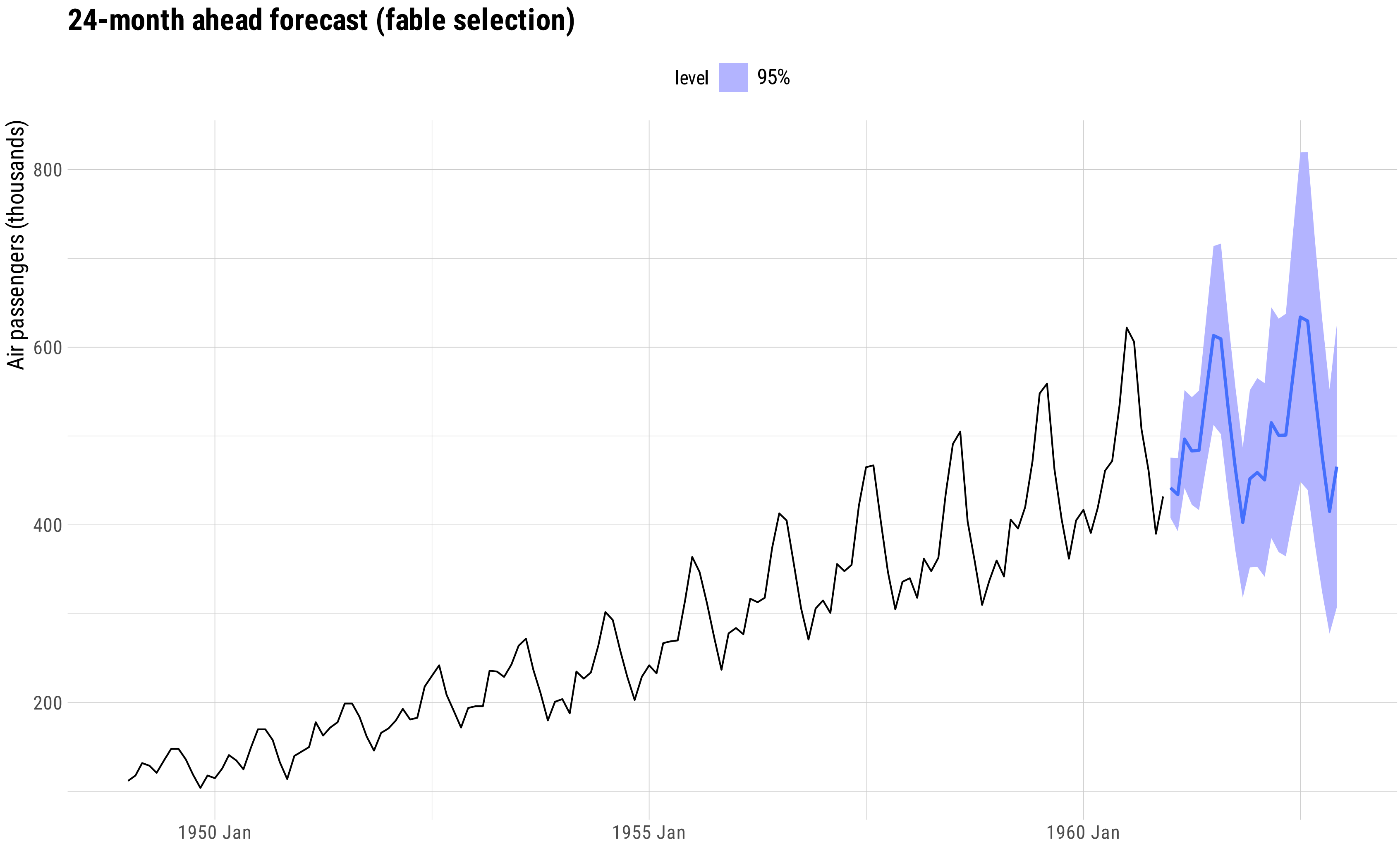
Fortunately for us, the `ETS()` function will deliver the model specification that **minimizes** the **corrected AIC**.

- Of course, when we do not specify a model ourselves.

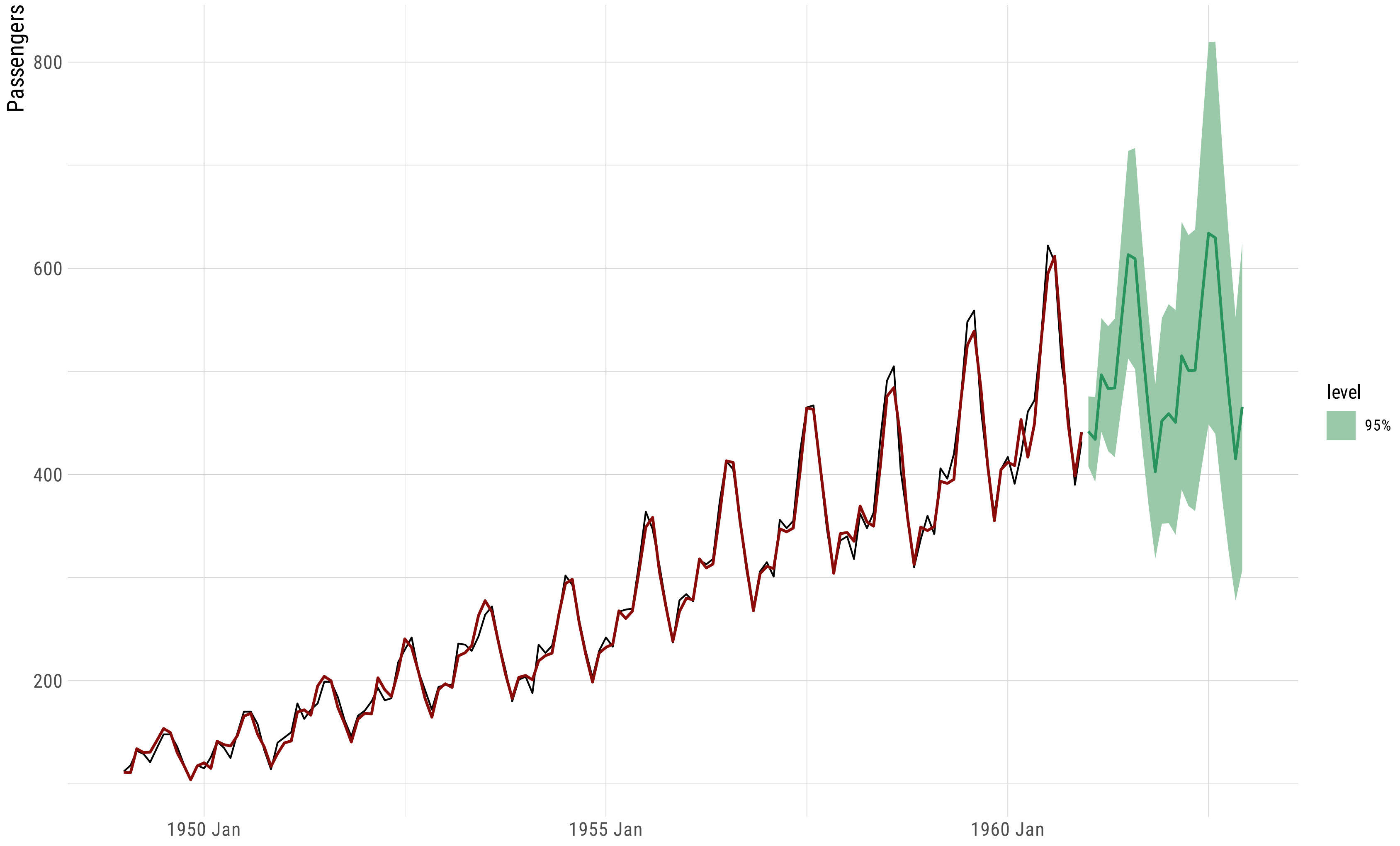
```
air_pass_ets <- air_ts |>
  model(ETS_best_model = ETS(passengers))

air_pass_ets
```

```
#> # A mable: 1 x 1
#>   ETS_best_model
#>          <model>
#> 1  <ETS(M,Ad,M)>
```

Model selection



Model selection



Next time: ARIMA models