Time Series Forecasting

Henrika Langen

Central European University
Department of Economics and Business

Winter 2024

Structure of the Lecture

1. Introduction to time series
2. Modeling time series and evaluating forecasting performance
3. **Exponential smoothing models**
4. ARIMA models
5. Regression models and advanced forecasting models
6. Machine Learning and Deep Learning methods

Exponential Smoothing Models

Topics

- Moving average smoothing
- Simple exponential smoothing
- Methods with trend
- Methods with seasonality
- ETS models

Introduction

- This chapter deals with time series models based in decomposing time series in its main components:
  - Trend-cycle component: T_t
  - Seasonal component: S_t
  - Remainder component: R_t
- The components can either be assumed to enter the time series y_t additively:

$$y_t = T_t + S_t + R_t$$

or multiplicatively:

$$y_t = T_t \times S_t \times R_t$$

In general, all time series decomposition approaches follow a similar procedure:

1. Smooth the time series to obtain $\hat{T}_t$
2. Calculate the de-trended series as $y_t - \hat{T}_t$ in case of additive decomposition or $y_t/\hat{T}_t$ in case of multiplicative decomposition
3. Average the observations of the de-trended series for each season separately to obtain $\hat{S}_t$
4. Calculate the remainder component as $\hat{R}_t = y_t - \hat{T}_t - \hat{S}_t$ in case of additive decomposition (or $\hat{R}_t = y_t/(\hat{T}_t\hat{S}_t)$ in case of multiplicative decomposition)

Moving average smoothing:

- Moving average smoothing involves averaging observations within a specific window of k observations before and k observations after each value of the series.
  - $\Rightarrow$ The method smooths out noise and highlights trends.
- A moving average of order a (denoted as a-MA), i.e., with window width a, is defined as:

$$T_t = \frac{1}{a} \sum_{j=-k}^{k} y_{t+j},$$

  where $a = 2k + 1$.
- The larger a, the smoother is the curve.
- There are no values in the smoothed curve for the first and last k points in time.

Moving average smoothing for seasonal time series:

- To smooth out seasonality, set a equal to the seasonal frequency m.
- **Note:** For a symmetric moving average window, a should be odd (k observations before and after y_t and y_t itself).

⇒ For seasonal data with an even frequency, apply a-MA followed by 2-MA to ensure a symmetric window (2 x a -MA).

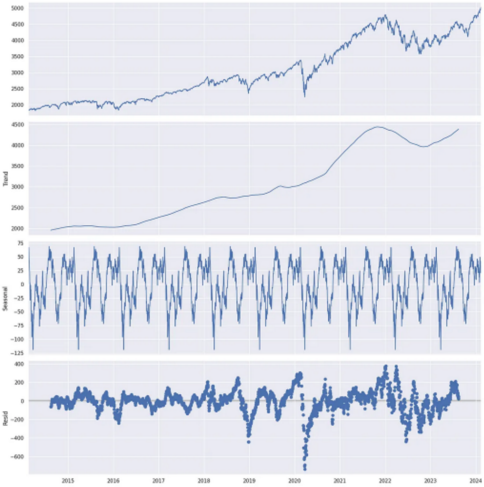
E.g.: Smoothing out quarterly data with yearly seasonality (2 x 4-MA):

$$\begin{aligned}
\hat{T}_t &= \frac{1}{2}\left[\frac{1}{4}\sum_{j=-2}^{1} y_{t+j} + \frac{1}{4}\sum_{j=-1}^{2} y_{t+j}\right] \\
&= \frac{1}{2}\left[\frac{1}{4}(y_{t-2}+y_{t-1}+y_t+y_{t+1}) + \frac{1}{4}(y_{t-1}+y_t+y_{t+1}+y_{t+2})\right] \\
&= \frac{1}{8}y_{t-2} + \frac{1}{4}y_{t-1} + \frac{1}{4}y_t + \frac{1}{4}y_{t+1} + \frac{1}{8}y_{t+2} \qquad (\widehat{=}\ \text{weighted moving average})
\end{aligned}$$

Classical time series decomposition

1. Choose between additive and multiplicative decomposition:
   - Additive decomposition is suitable if seasonal variation remains relatively constant over time.
   - Multiplicative decomposition is preferable if seasonal variation increases over time (should only be used for purely positive data).
2. Smooth the time series via moving average smoothing to obtain $\hat{T}_t$.
3. Calculate the de-trended series as $y_t - \hat{T}_t$ (additive decomposition) or $y_t/\hat{T}_t$ (multiplicative decomposition).
4. Average the observations of the de-trended series for each season $1, \ldots, m$ separately to obtain $\hat{S}_t$.
5. Calculate the remainder component as $\hat{R}_t = y_t - \hat{T}_t - \hat{S}_t$ (or $\hat{R}_t = y_t/(\hat{T}_t\hat{S}_t)$, respectively)

Time Series Decomposition



Classical additive decomposition of the S&P500 index price.

Source:
https://medium.com/@ipeksahbazoglu/time-series-decomposition-and-moving-average-smoothing-fd61b558ea36

Forecasting with classical time series decomposition

- The decomposed time series can be written as:

$$y_t = \hat{S}_t + \hat{A}_t \quad \text{for additive decomposition or}$$
$$y_t = \hat{S}_t \cdot \hat{A}_t \quad \text{for multiplicative decomposition,}$$

  where $\hat{A}_t = \hat{T}_t + \hat{R}_t$ (or, $\hat{A}_t = \hat{T}_t \cdot \hat{R}_t$, respectively) is the seasonally adjusted component.
- To obtain a forecast $y_{T+h|T}$, determine $\hat{A}_{T+h|T}$ and $\hat{S}_{T+h|T}$ and then combine them:
  - Estimate $\hat{A}_{T+h|T}$ using any non-seasonal forecasting method like a random walk with drift model (or more advanced methods, see later sections).
  - Estimate $\hat{S}_{T+h|T}$ as the last observed seasonal component of the same season or the average of the seasonal components of the respective season.
  - Combine the two by multiplication or addition, respectively.

Advantages and disadvantages of classical time series decomposition:

+ The approach is very intuitive and easy to follow.
- There are no estimates available for the first and last k observations.
- The seasonal component is assumed to be the same over all periods (or to increase proportionally to the trend-cycle component).
- Depending on m, the trend-cycle estimate may over-smooth the series.

$\Rightarrow$ Several alternative approaches address these issues while still being based on the basic steps used in classical decomposition (e.g. X11, SEATS and STL decomposition)

Simple exponential smoothing:

- For simple exponential smoothing, each value is replaced by a weighted average of itself and previous observations, with the weights decreasing as the observations go further back in time.
- The exponentially smoothed statistic s_t is given by:

$$\begin{aligned} s_0 &= y_0 \\ s_t &= \alpha y_t + (1-\alpha)s_{t-1} \quad \text{for } t > 0, \end{aligned}$$

  where $0 \leq \alpha \leq 1$ is the smoothing parameter, with smaller values of α resulting in a smoother curve.
- Substitution shows:

$$\begin{aligned} s_t &= \alpha y_t + (1-\alpha)s_{t-1} \\ &= \alpha y_t + (1-\alpha)\left[\alpha y_{t-1} + (1-\alpha)s_{t-2}\right] = \alpha y_t + \alpha(1-\alpha)y_{t-1} + (1-\alpha)^2 s_{t-2} \\ &= \alpha y_t + \alpha(1-\alpha)y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} + (1-\alpha)^3 s_{t-3} \\ &= \alpha[y_t + (1-\alpha)y_{t-1} + \ldots + (1-\alpha)^{t-1}y_1] + (1-\alpha)^t y_0 \end{aligned}$$

Smoothing Methods

Simple exponential smoothing vs. moving average smoothing

- A series that is smoothed with simple exponential smoothing contains all values, while a-MA smoothed series do not contain the first and last k observations.
- Simple exponential smoothing only works well when there is no trend and no seasonality in the data; moving average smoothing also works well for series with trend and seasonality.
- Simple exponential smoothing captures irregular or random fluctuations well, while moving average smoothing may smooth them out.

$\Rightarrow$ Use simple exponential smoothing for stable time series without trend or seasonality in order to capture short-term fluctuations.

$\Rightarrow$ Use moving average smoothing for stabilizing trends, reducing noise and capturing long-term patterns.

Forecasting with simple exponential smoothing (only advisable for series without trend):

The simple exponential smoothing approach can be easily adapted for forecasting by simply replacing the s_t with $\hat{y}_{t|t-1}$.

- The forecast for time $T+1$ is given by a weighted average of y_T and the forecast for time T based on all observations before T, i.e. $\hat{y}_{T|T-1}$:

$$
\begin{aligned}
\hat{y}_{T+1|T} &= \alpha y_T + (1-\alpha)\hat{y}_{T|T-1} \\
&= \alpha y_T + (1-\alpha)\left[\alpha y_{T-1} + (1-\alpha)\hat{y}_{T-1|T-2}\right] = \alpha y_T + \alpha(1-\alpha)y_{T-1} + (1-\alpha)^2\hat{y}_{T-1|T-2} \\
&= \ldots \\
&= \alpha[y_T + (1-\alpha)y_{T-1} + \ldots + (1-\alpha)^{T-1}y_1] + (1-\alpha)^T y_0
\end{aligned}
$$

- The resulting forecast is flat, i.e., $\hat{y}_{T+h|T} = \hat{y}_{T+1|T} \quad \forall h > 2$.
- The optimal smoothing parameter α is identified by minimizing the sum of squared residuals. Additionally, one may also use optimization to find a different starting value than y_0.

Forecasting with Trend

- **Holt's linear trend method** (also called double exponential smoothing) extends simple exponential smoothing to allow for forecasting series with a trend.
- Approach:

$$\text{Forecast equation:} \quad \hat{y}_{t+h|t} = l_t + hb_t$$

$$\text{Level equation:} \quad l_t = \alpha y_t + (1-\alpha)\cdot \underbrace{(l_{t-1} + b_{t-1})}_{\text{1-step-ahead training forecast}}$$

$$\text{Trend equation:} \quad b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1},$$

  where l_t denotes an estimate of the level and b_t an estimate of the trend at time t. $0 \leq \alpha \leq 1$ and $0 \leq \beta \leq 1$ are smoothing parameters for level and trend, respectively
- The optimal values for the smoothing parameters as well as l_0 and b_0 can be identified by minimizing the sum of squared residuals.
- The resulting forecasts for $T+h$ lie on a straight line with slope b_T.

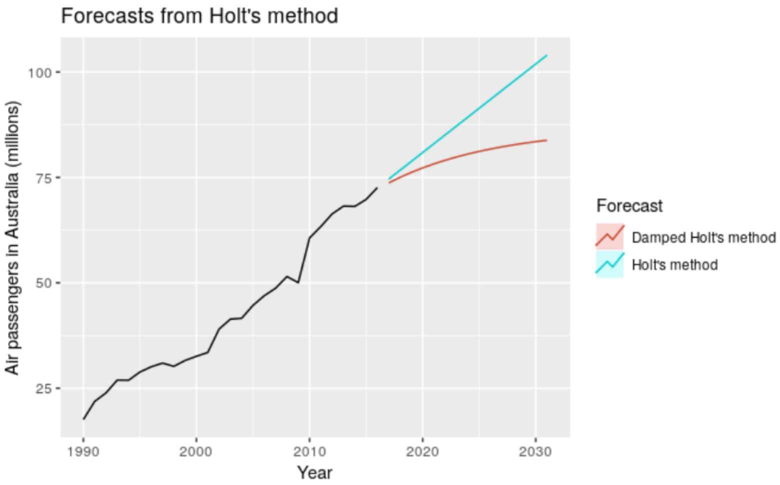
Forecasting with Trend

- **Damped trend methods** are advancements of Holt's linear trend method that gradually flatten the forecasted trend over time, as empirical evidence suggests that forecasts with linear trends tend to overshoot for longer forecast horizons.
- To flatten the forecasted trend, a damping parameter $0 \leq \phi \leq 1$ is introduced:

$$\begin{aligned} \text{Forecast equation:} \quad & \hat{y}_{t+h|t} = l_t + (\phi + \phi^2 + \ldots + \phi^h) b_t \\ \text{Level equation:} \quad & l_t = \alpha y_t + (1-\alpha)(l_{t-1} + b_{t-1}) \\ \text{Trend equation:} \quad & b_t = \beta(l_t - l_{t-1}) + (1-\beta)\phi b_{t-1}, \end{aligned}$$

- The optimal value of ϕ, the smoothing parameters as well as l_0 and b_0 can be determined by minimizing the sum of squared residuals.

Forecasting with Trend



Forecasting the total annual number of passengers for air carriers registered in Australia (in millions of passengers). $\phi = 0.9$ for the damed approach.

Source: FPP

Forecasting with Seasonality

- **Holt-Winters' seasonal method:**

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

$$\text{Level equation:} \quad l_t = \alpha \underbrace{(y_t - s_{t-m})}_{\text{seasonally adjusted obs.}} + (1-\alpha) \underbrace{(l_{t-1} + b_{t-1})}_{\text{non-seasonal forecast}}$$

$$\text{Trend equation:} \quad b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1}$$

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

  where k is the integer part of $(h-1)/m$, i.e., the number of complete seasonal cycles between T and $T+h$ and α, β and γ are smoothing parameters.

- Additionally to this additive approach there is a multiplicative one, better suited when seasonal variation changes proportionally to the level (see e.g. FPP).
- Plus, a damped trend may be incorporated into the above model.

The different exponential smoothing approaches can be combined in various ways:

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)

Source: FPP

Forecasting with ETS Models

To obtain prediction intervals for the exponential smoothing forecasts, we must specify how the error enters the model (additively or multiplicatively).
The full model specification for exponential smoothing forecasting is known as ETS ("Error Trend Seasonality"):

- The **Error** can enter the model "Additively" (**A**) or "Multiplicatively" (**M**)
- **Trend** can be "None" (**N**), "Additive" (**A**), "Additive damped" (**Ad**)
- **Seasonality** can be "None" (**N**), "Additive" (**A**) or "Multiplicative" (**M**).

ETS models are denoted as E(..., ..., ...) with type of error, trend and seasonality respectively in the blanks.
E.g. simple exponential smoothing, i.e., a model with additive error, no trend and no seasonality, is denoted as E(A,N,N)

To estimate the error component in an exponential smoothing model, we can rearrange the model equations such that the residual $e_t = y_t - \hat{y}_{t|t-1}$ is included:

E.g. in the **simple exponential smoothing** model with

$$\text{Forecast equation:} \quad \hat{y}_{t+1|t} = l_t$$

$$\text{Level equation:} \quad l_t = \alpha y_t + (1-\alpha) l_{t-1}$$

we can express y_t as a function of the residual e_t, which - under the assumption that the **error enters additively** - is defined as $e_t = y_t - \hat{y}_{t|t-1} = y_t - l_{t-1}$:

$$\Leftrightarrow \; y_t = l_{t-1} + e_t$$

We can rewrite the level equation as:

$$l_t = \alpha y_t + (1-\alpha) l_{t-1} = l_{t-1} - \alpha \underbrace{(y_t - l_{t-1})}_{=e_t} = l_{t-1} + \alpha e_t$$

So, the equations of the model can be rewritten as:

$$y_t = l_{t-1} + e_t$$

$$l_t = l_{t-1} + \alpha e_t, \quad \text{where } e_t \text{ enters additively.}$$

Particularities of ETS models and estimation

- The probability distribution of the residuals e_t is usually assumed to be e_t $NID(0, \sigma^2)$ (i.e., normally and independently distributed white noise ε_t with mean 0 and variance σ^2)
- The parameters in ETS models, particularly the smoothing parameters, can be estimated by minimizing least squares or with maximum likelihood estimation.
- For model specifications of different ETS models see FPP 7.5.

Estimating ETS Models

Model selection

- To determine the most suitable ETS model for a given time series, i.e., in terms of Error, Trend and Seasonality component, one can use information criteria (AIC, AIC$_c$, BIC).
- Once, a model is selected...
  - ... check the ACF of the residuals.
  - $\Rightarrow$ The residuals should resemble white noise. If there is autocorrelation left in the residuals, try to find a better-fitting model.
  - examine the residuals by plotting a histogram.
  - $\Rightarrow$ The residuals should be approximately normally distributed.

Forecasting with ETS Models

Forecasting: Once, a model is selected

- ETS models produce the same point forecasts as the corresponding exponential smoothing model. Additionally, they model the error, enabling the estimation of prediction intervals.
- Forecasts can be obtained by setting $e_t = 0$, estimating $\hat{y}_{T+1|T}$ based on the corresponding equation and then again using $\hat{y}_{T+1|T}$ to estimate $\hat{y}_{T+2|T}$, etc.
- The estimation of prediction intervals is comparably complex, but prediction intervals are provided when forecasting with ETS models in Python.

ETS Models

Advantages and disadvantages of ETS models and which data to use them with

+ ETS models effectively capture trend and seasonality in various types of time series, making them widely used in time series modeling and forecasting.
+ They are flexible and adapt to changing patterns and trends in the data, without requiring strong assumptions about the form of the time series.
+ They handle outliers and noise well by smoothing them out.
+ ETS models account for autocorrelation by expressing the different components of the series in t as a function of their past.
- ETS models may not capture certain complex patterns, such as abrupt changes in seasonality, as seasonality is assumed to remain constant or change proportionally to the trend.
- The accuracy of forecasts can be affected by the choice of smoothing parameters, and finding optimal values for these parameters may be computationally intensive.