L20

Basic Statistics for Economists

Spring 2020

Department of Statistics

Stockholm University

Previous lecture

NCT **16.1 – 16.2**

- **Time-series**
  - sequence of observations $\boldsymbol{x_t}$; index $\boldsymbol{t}$ denotes a point in time
  - notation: $(\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_n})$ or $\{\boldsymbol{x_t}\}$, $t = 1, \ldots, n$
- **Components**
  - Trend, Seasonality, Cyclicity, Irregularities i.e. randomness
  - Additive and Multiplicative models
- **Separation of components**
  - Moving Averages (MA) or Regression or … for de-trending
  - **Seasonal adjustment**

Why seasonal estimates and adjustments?

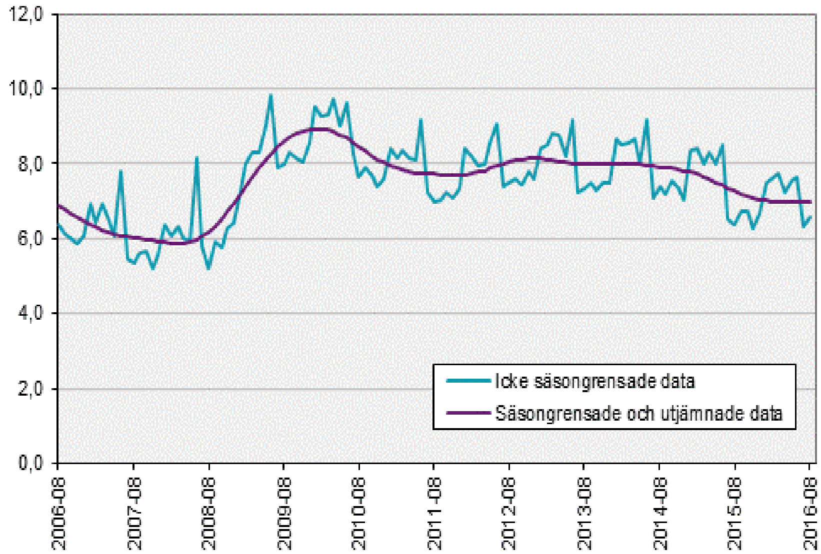
- High-lighting general behavior:
  - Level, direction “up or down?” (descriptive)
  - Easier to compare time-series, general patterns
- Seasonal indices are needed for forecasting (*prognoser*):
  - Extrapolating a trend (*trendframskrivning*)
  - Allocating an annual forecast over seasons
- Causes:
  - Natural phenomena (cold in winter, hot in summer)
  - Human factors, traditions (Christmas, vacations, new car models launched in the autumn)

Stockholm University

Component methods

Summary

The method, despite its simplicity, works well on time series with stable seasonal variations and if the seasonal effects in absolute or relative numbers is fairly constant over time.



Stockholm University

Today

Slightly more advanced method and some extensions that allow for **forecasting** (sv. *prognoser*)

- **Exponential smoothing NCT 16.3**
  - **Simple exponential smoothing, no trend**, no seasonality
    - sv. *exponentiell utjämning*
  - **Holt-Winter’s method with trend**, no seasonality
  - **Holt-Winter’s method with trend and seasonality**

Stockholm University

Forecasts

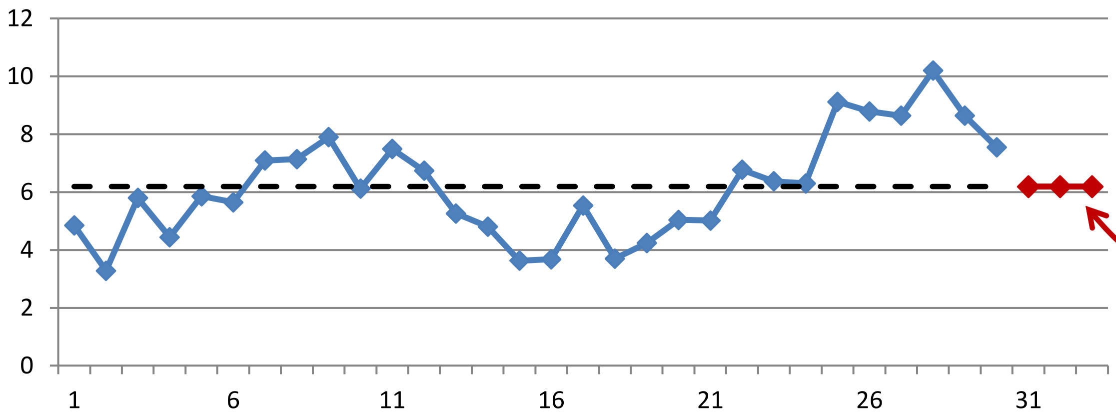
- Present (latest) value $\boldsymbol{x_n}$ and previous $\boldsymbol{x_{n-1}, \ldots, x_1}$ are used to calculate forecasts $\boldsymbol{\hat{x}_{n+1}, \hat{x}_{n+2}, \hat{x}_{n+3}, \ldots}$
- **Future** forecasts: ***out-of-sample forecasts***

Once we have defined a technique for making forecasts, we can of course also forecast for previous time points, when we know what the outcome was; so-called ***in-sample-forecasts***

- "Forecasting" the past: $\boldsymbol{\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n}$
- Used for comparing methods and settings

Naïve forecast 1; mean of the series

- Model: $X_t = \mu + \varepsilon_t$ **level** $= \mu$ relatively **constant** over time
  no trend, cyclic pattern or seasonality



Mean $= \boxed{\overline{\boldsymbol{x}} = \mathbf{6,19}}$

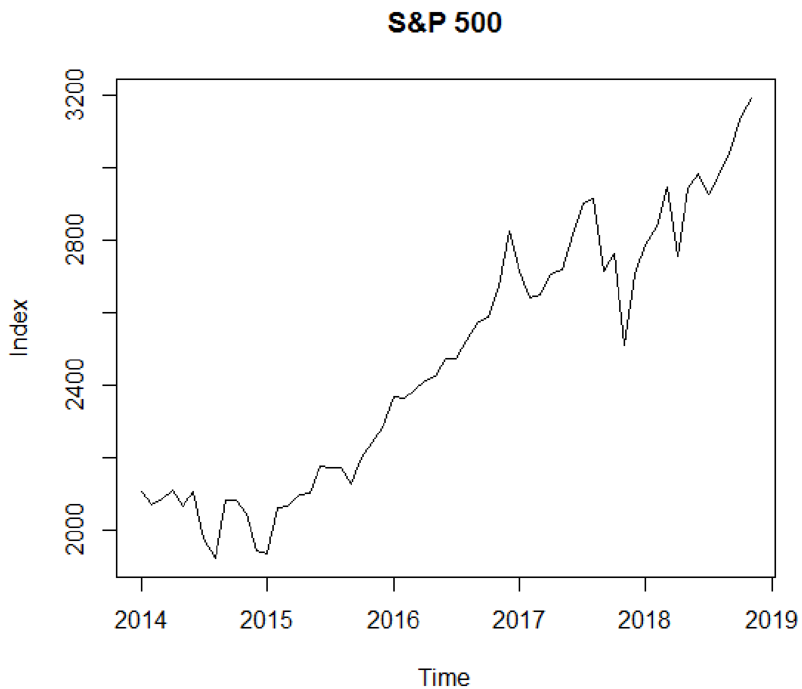
Forecasts at $\boldsymbol{t} = \mathbf{30}$:

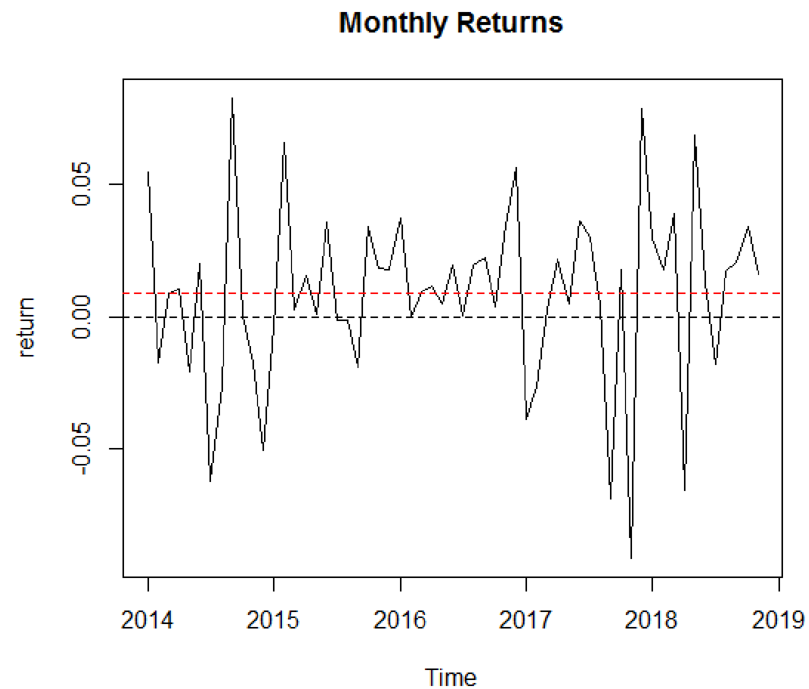
$n + 1 = 31;\ \hat{x}_{31} = 6{,}19$
$n + 2 = 32;\ \hat{x}_{32} = 6{,}19$
$n + 3 = 33;\ \hat{x}_{33} = 6{,}19$

- **Forecast method 1**: $\hat{\boldsymbol{x}}_{\boldsymbol{n+h}} = \overline{\boldsymbol{x}} =$ **mean** of all observations
  - All the time points are equally important! Reasonable?

Stockholm University

Example, Return of S&P 500

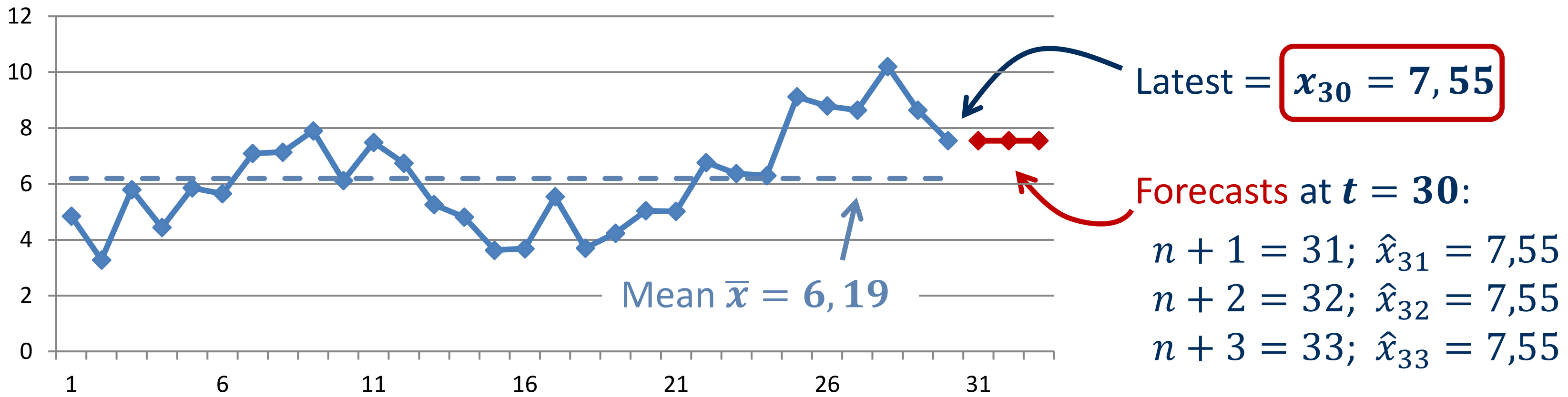




Naïve forecast 2; last observed value

- Model: $X_t = \beta_t + \varepsilon_t$ **level** $= \beta_t$ **slowly changing** over time t

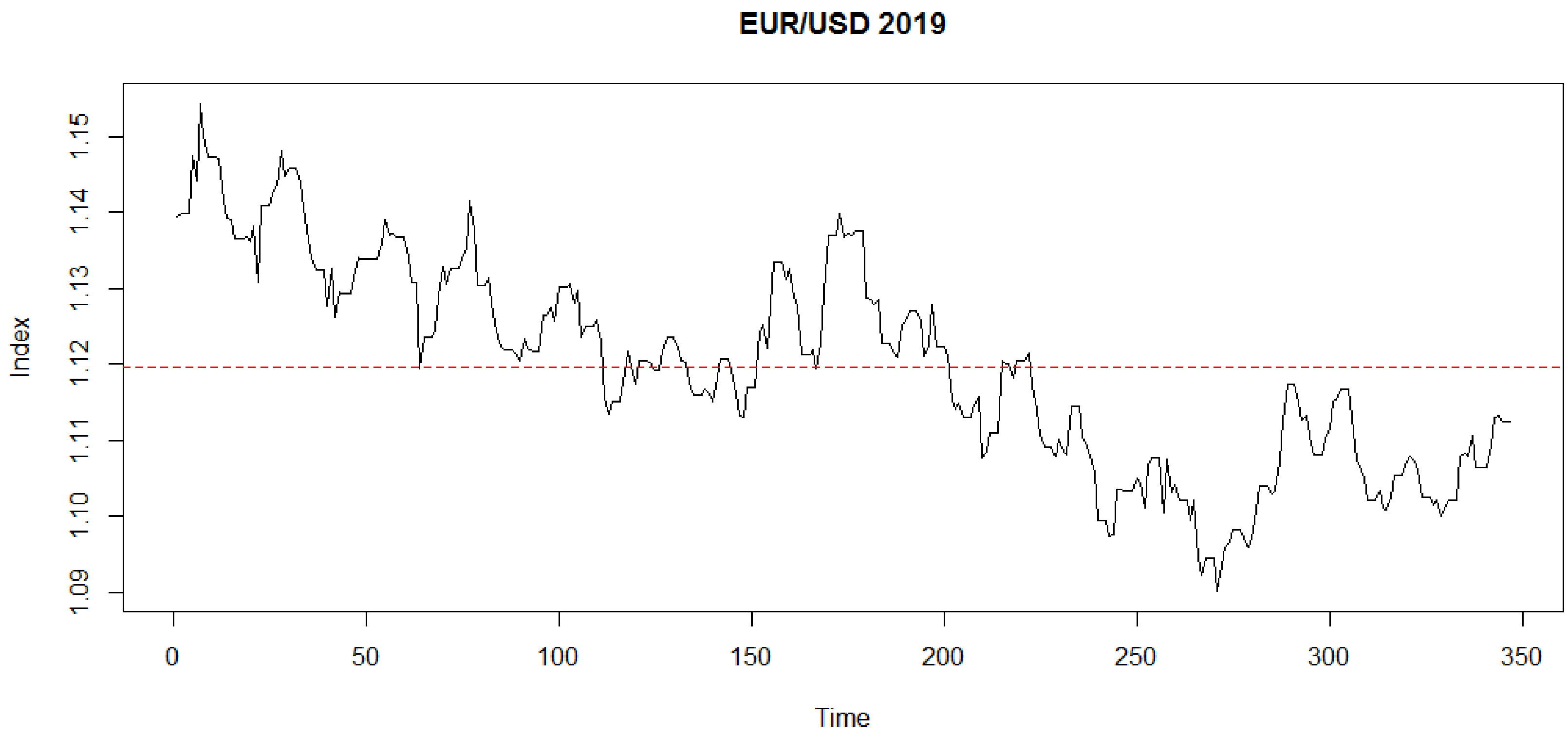
  Cyclic patterns (no trend or season)



- **Forecast method 2**: $\hat{x}_{n+h} = x_n =$ **latest observation**
  - Only the latest observation matters! No history! Reasonable?

Stockholm University

Example, EUR/USD exchange rate



Stockholm University

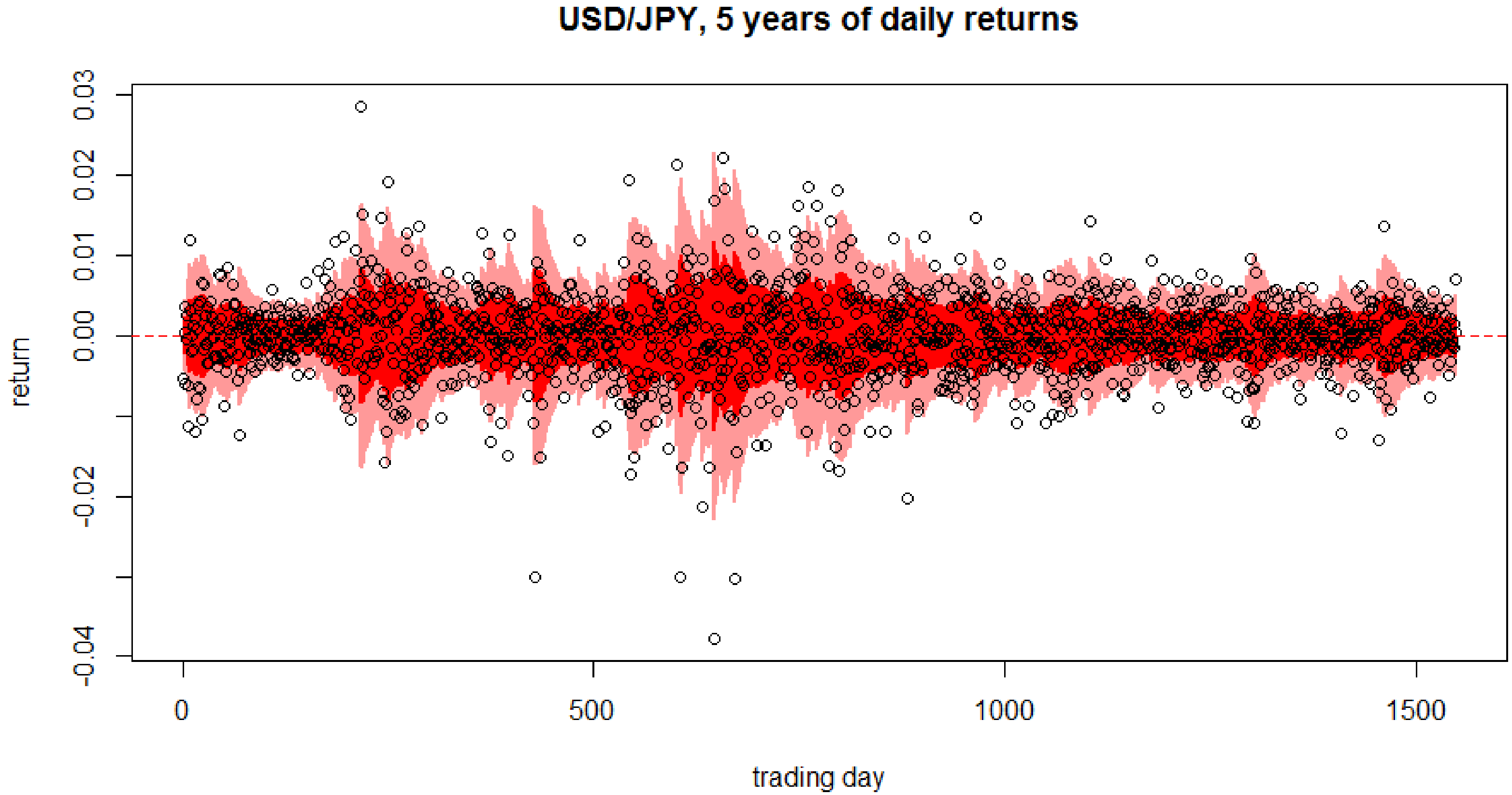
Simple exponential smoothing

- Something in between methods 1 and 2
- An estimate for the current value/level $\hat{x}_t$ at time t is given by

$$\hat{x}_t = \underbrace{\alpha x_t}_{\substack{\alpha \times \text{current observation} \\ \text{at time } t}} + \underbrace{(1-\alpha)\hat{x}_{t-1}}_{\substack{(1-\alpha) \times \text{level estimate from} \\ \text{the previous time point, } t-1}} \qquad \boxed{0 < \alpha < 1}$$

- At every time point t a new level estimate $\hat{x}_t$ is calculated by weighting together the current observation (x_t) with the most recent level estimate ($\hat{x}_{t-1}$)
- Weighted with the **smoothing constant** α

Example, USD/JPY estimated volatility



Understanding the formula

- Expand the formula (see NCT p 698):

$$\hat{x}_t = \alpha x_t + (1-\alpha)\cdot \boxed{\hat{x}_{t-1}} = \alpha x_t + (1-\alpha)\cdot \boxed{[\alpha x_{t-1} + (1-\alpha)\hat{x}_{t-2}]}$$

$$= \alpha x_t + \alpha(1-\alpha)x_{t-1} + (1-\alpha)^2 \cdot \boxed{\hat{x}_{t-2}}$$

$$= \alpha x_t + \alpha(1-\alpha)x_{t-1} + (1-\alpha)^2 \cdot \boxed{[\alpha x_{t-2} + (1-\alpha)\hat{x}_{t-3}]}$$

$$= \alpha x_t + \alpha(1-\alpha)x_{t-1} + \alpha(1-\alpha)^2 x_{t-2} + (1-\alpha)^3 \cdot \boxed{\hat{x}_{t-3}}$$

… repeat …

$$= \alpha x_t + \alpha(1-\alpha)x_{t-1} + \alpha(1-\alpha)^2 x_{t-2} + \alpha(1-\alpha)^3 x_{t-3} + \alpha(1-\alpha)^4 x_{t-4} + \cdots$$

- $\hat{x}_t$ = **weighted mean** of all observations; the largest weight given to the current observation and then exponentially decreasing weights.

Stockholm University

Forecasting with exponential smoothing

- At time $\boldsymbol{t} = \boldsymbol{n}$ forecasts for the **future** $n+1, n+2, \ldots$ are

$$\boxed{\widehat{\boldsymbol{x}}_{\boldsymbol{n}+\boldsymbol{h}} = \widehat{\boldsymbol{x}}_{\boldsymbol{n}}} = \alpha x_n + (1-\alpha)\hat{x}_{n-1} \quad \text{where } h = 1, 2, 3, \ldots$$

i.e. the latest smoothed value is the forecast for all future

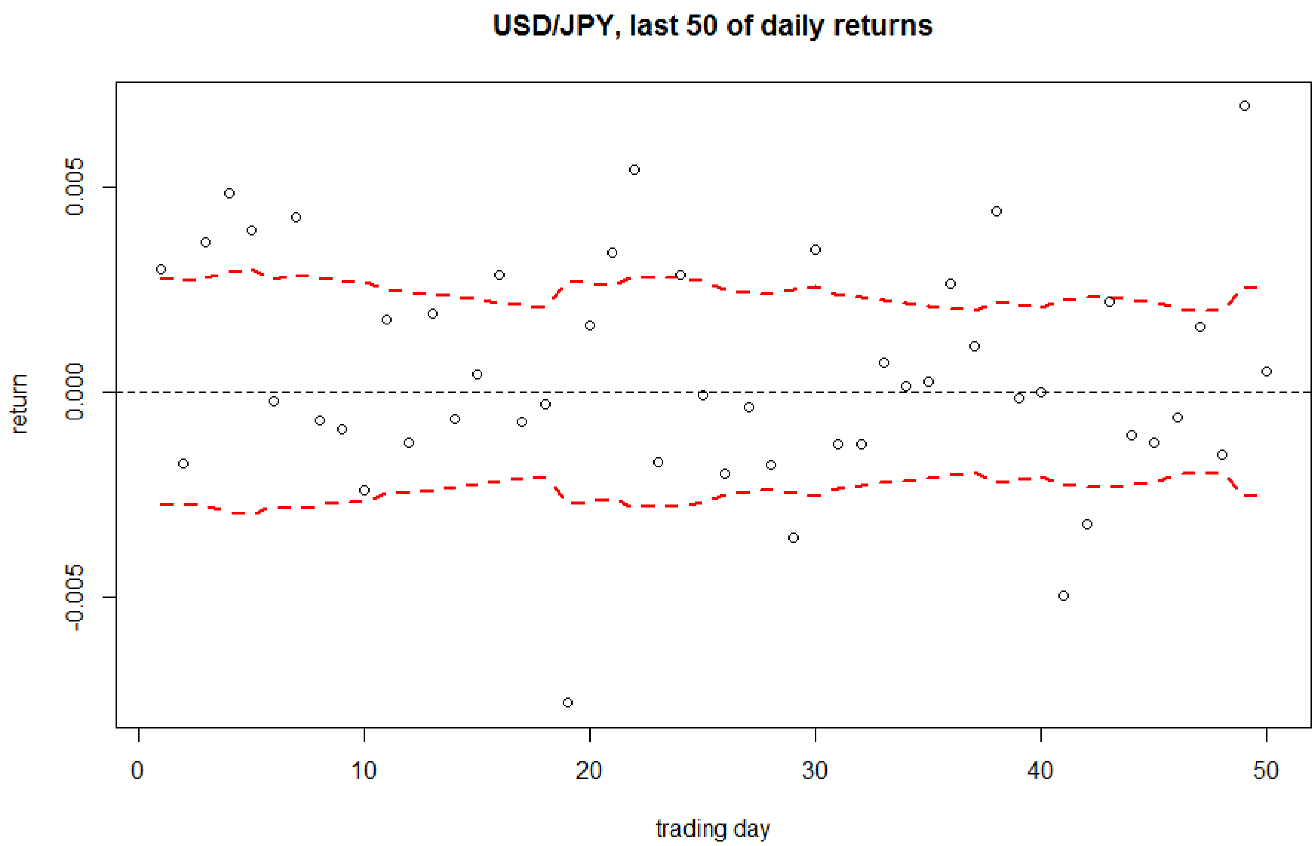
- Considers only the **current level**, <u>not</u> trend or seasonality

- **In-sample-forecast** for time $\boldsymbol{t}$ is equal to the most recently smoothed value, i.e. from the **time prior to the current** i.e.

$$\boxed{\widetilde{\boldsymbol{x}}_{\boldsymbol{t}} = \widehat{\boldsymbol{x}}_{\boldsymbol{t}-\boldsymbol{1}}} = \alpha x_{t-1} + (1-\alpha)\hat{x}_{t-2} \quad \text{where } t = 2, \ldots, n$$

Stockholm University

Example, USD/JPY estimated volatility

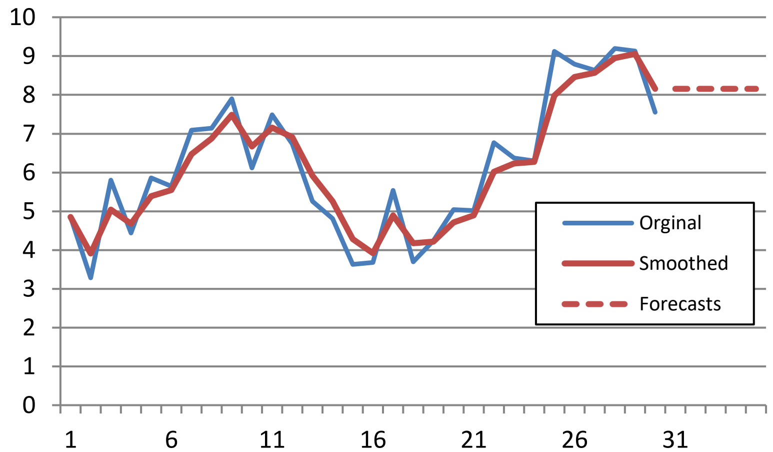


Stockholm University

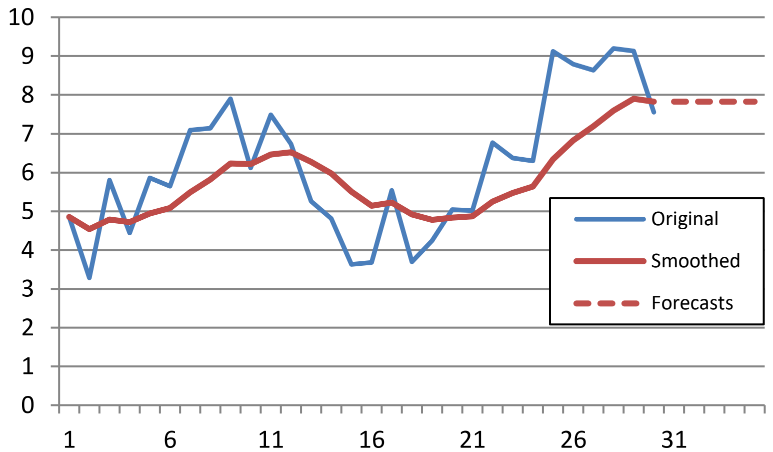
The smoothing constant

- Larger α gives more weight to the latest observation
  – less smoothing
- Smaller α gives more weight to “older” data
  – more smoothing

$\boxed{0 < \alpha < 1}$



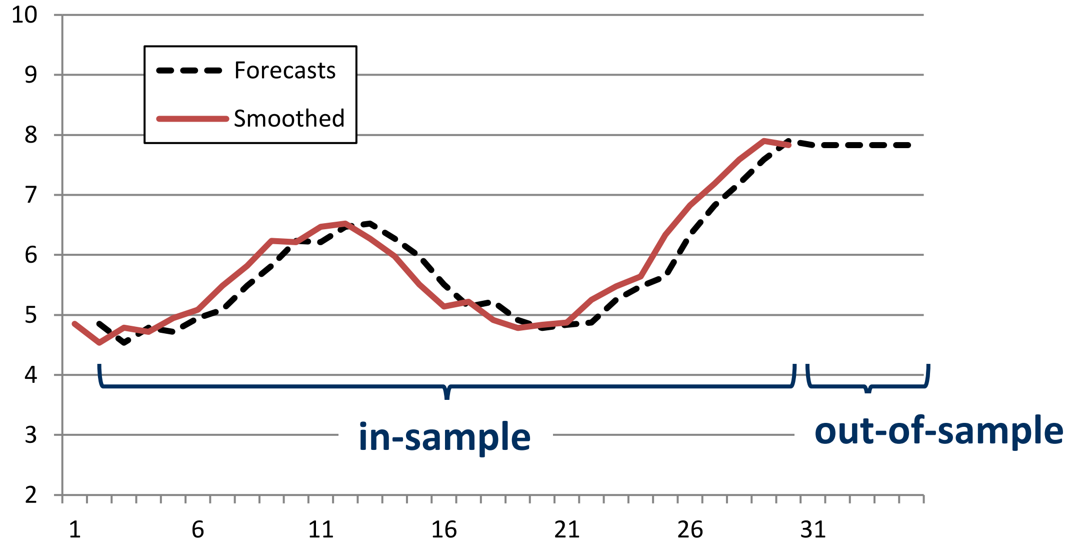
$\alpha = 0,6$



$\alpha = 0,2$

Smoothed values and forecasts

- **In-sample-forecast** for time t is equal to the smoothed value of the **time-point prior to the current**: $\tilde{x}_t = \hat{x}_{t-1}$
- Causes the in-sample-forecasts to be **offset** by one time-point forward:



Stockholm University

Choosing a start value $\hat{x}_1$

We don't have a value x_0 for the time just before the series starts!
How do we start off our calculations? Two (three) ways:

1. NCT: set $\hat{x}_1 = x_1$ and proceed with $\hat{x}_2$, $\hat{x}_3$, … accordingly
2. Alt: set x_0 = mean value of the k first ($k = 1,2,3,\ldots$), then proceed with $\hat{x}_1$, $\hat{x}_2$, … accordingly
3. Excel: set $\hat{x}_2 = x_1$ and proceed with $\hat{x}_3$, $\hat{x}_4$, … but they use an entirely different formula! ($\hat{x}_t = (1-\alpha)x_{t-1} + \alpha\hat{x}_{t-1}$)

- 1 and 2 yield a bit different results but if the series is long the difference is negligible when we get to the end of the series.

Example 1

t	x_t	$\hat{x}_t$	Forecast
1	4,85	4,850	*
2	3,28	3,908	4,850
3	5,80	5,043	3,908
4	4,44	4,681	5,043
⋮	⋮	⋮	⋮
28	9,20	8,948	8,571
29	9,13	9,057	8,948
30	7,55	8,153	9,057
31	*	*	8,153
32	*	*	8,153
33	*	*	8,153

Smoothing constant: $\boldsymbol{\alpha = 0,6}$

Start value: $\hat{x}_1 = x_1 = \boxed{4,85}$

$$\hat{x}_2 = 0{,}6 \cdot 3{,}28 + 0{,}4 \cdot \boxed{4,850} = \boxed{3,908}$$

$$\hat{x}_3 = 0{,}6 \cdot 5{,}80 + 0{,}4 \cdot \boxed{3,908} = \boxed{5,043}$$

⋮

Forecasts:

- in-sample: $\tilde{x}_t = \hat{x}_{t-1}$
- out-of-sample: $\hat{x}_{30+h} = \hat{x}_{30} = \mathbf{8.153}$

Stockholm University

Choosing the smoothing constant α

- Constant $\boldsymbol{\alpha}$ determines how "smoothed" it'll be. Best choice?
- Criteria for choosing optimal $\boldsymbol{\alpha}$ based on in-sample-forecasts
- E.g. choose $\boldsymbol{\alpha}$ so that the sum-of-squared-differences between observed value and in-sample-forecast is minimized:

Observed $= \boldsymbol{x_t}$ Forecast $= \tilde{\boldsymbol{x}}_t = \hat{\boldsymbol{x}}_{t-1}$ $\quad \boldsymbol{t = 2, 3, \dots, n}$

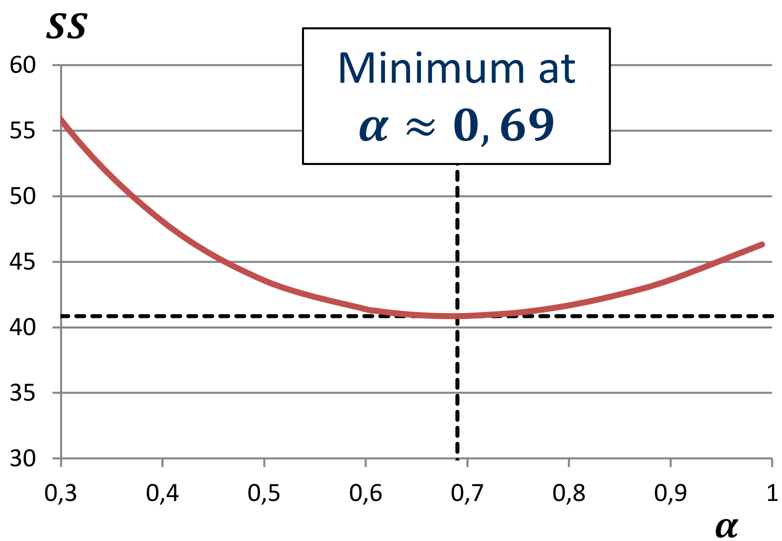
Minimize $\quad \boldsymbol{SS} = \sum_{t=2}^{n} e_t^2 = \sum_{t=2}^{n} (x_t - \hat{x}_{t-1})^2$

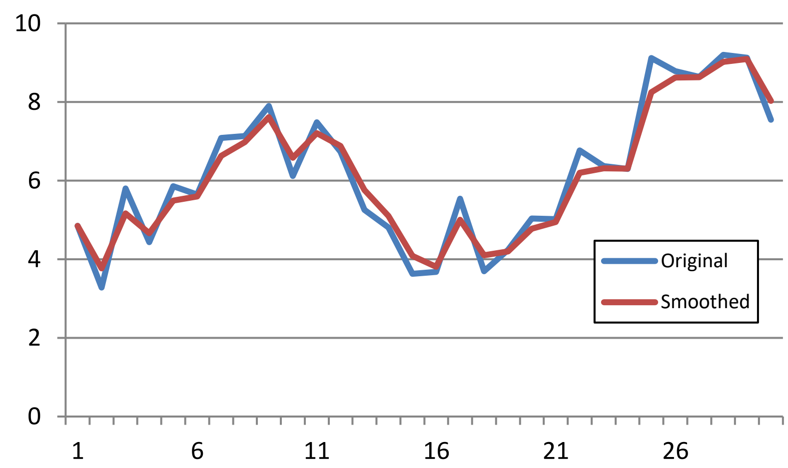
- Statistical software often have built-in procedures for finding optimal value for $\boldsymbol{\alpha}$

Stockholm University

Example 1, cont.

- Same time series as before:



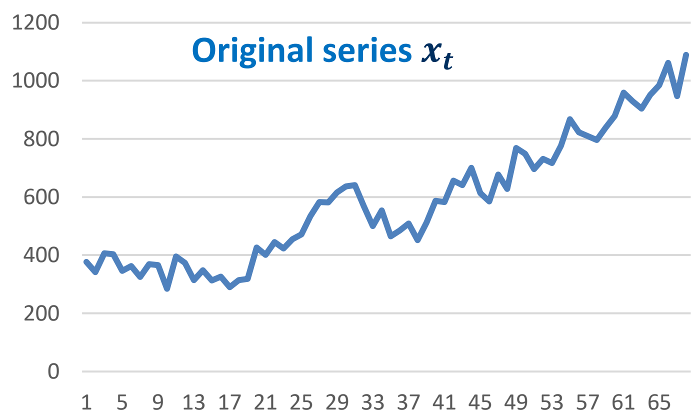


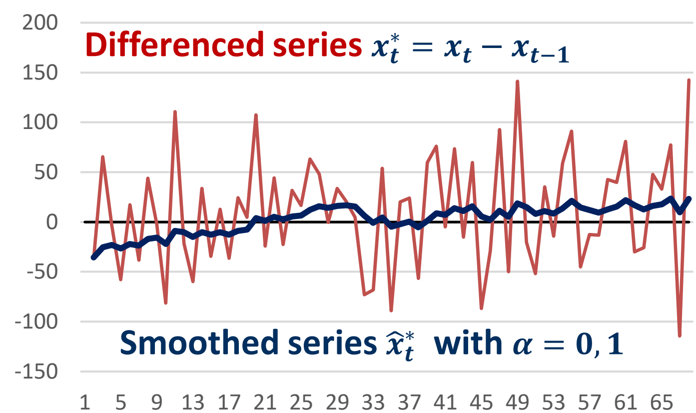
Actually minimum at $\alpha = 0,68548$

Stockholm University

Differencing a series with trend

- Compute differences $x_t^* = x_t - x_{t-1}$, the **change in level**
- Exponential smoothing on the new series $\{x_t^*\}$
- Eliminates trend (partially), often yields a fairly constant level





- More difficult to interpret results and to forecast

Stockholm University

Holt’s method, smoothing with trend

alt. **Holt-Winter’s metod** alt. **Double exponential smoothing**

- Considers **trend** (slope, direction), not seasonality
- Two formulas:

Provides an extra term, a direction, up or down

Level: $$\hat{x}_t = \underbrace{\alpha x_t}_{\substack{\alpha \times \text{current observation} \\ \text{at time } t}} + \underbrace{(1-\alpha)(\hat{x}_{t-1} + T_{t-1})}_{\substack{(1-\alpha) \times (\text{level} + \text{trend estimates}) \\ \text{at time } (t-1)}}$$

Trend: $$T_t = \underbrace{\beta(\hat{x}_t - \hat{x}_{t-1})}_{\substack{\beta \times \text{recent change in estimated} \\ \text{level from time } t-1 \text{ to } t}} + \underbrace{(1-\beta)T_{t-1}}_{\substack{(1-\beta) \times \text{most recent trend estimate} \\ \text{at time } t-1}}$$

Holt's method, cont.

- At every time-point $\boldsymbol{t}$ a new estimate of the level $\widehat{\boldsymbol{x}}_{\boldsymbol{t}}$ is given by weighing together the current $\boldsymbol{x}_{\boldsymbol{t}}$ and the previous level estimate $\widehat{\boldsymbol{x}}_{\boldsymbol{t-1}}$ which in turn is based on "older" information
- The weighting is determined by $\boldsymbol{\alpha}$ for the level and $\boldsymbol{\beta}$ for the trend
- The trend is updated by taking the **level change** between the two latest smoothed level estimates: $\widehat{\boldsymbol{x}}_{\boldsymbol{t}} - \widehat{\boldsymbol{x}}_{\boldsymbol{t-1}}$

Same questions as before but more:

- Determine <u>two</u> start values: $\boldsymbol{T_0}, \boldsymbol{T_1}$, alt. $\boldsymbol{T_2}$ and $\widehat{\boldsymbol{x}}_{\boldsymbol{0}}$, $\widehat{\boldsymbol{x}}_{\boldsymbol{1}}$ alt. $\widehat{\boldsymbol{x}}_{\boldsymbol{2}}$
- Choosing <u>two</u> smoothing constants: $\boldsymbol{\alpha}$ och $\boldsymbol{\beta}$

Holt's method, cont.

NCT suggests **start values** as follows:

1. Set $\hat{x}_2 = x_2$ and $T_2 = x_2 - x_1$
2. Then calculate a) $\hat{x}_3 = \alpha x_3 + (1-\alpha)(\hat{x}_2 + T_2)$

   b) $T_3 = \beta(\hat{x}_3 - \hat{x}_2) + (1-\beta)T_2$
3. Continue with $\hat{x}_t$ and T_t alternately for $t = 4, 5, \dots n$

NCT p. 703 describes an alternative procedure using Minitab

Optimal smoothing constants α and β can be determined in the same way as we demonstrated on slides 16-17 (NCT p. 699) or by experimenting. Reasonable forecasts?

Stockholm University

Forecasting *h* time-points forward

- Holt's model: $$\hat{x}_{n+h} = \hat{x}_n + h \cdot T_n$$

Last level estimate at time n

Last trend estimate at time n

- Extrapolation, i.e. out-of-sample forecasting, based on a linear trend

Example 2: Holt's method

t	x_t	$\hat{x}_t$	$\hat{x}_t - \hat{x}_{t-1}$	T_t	Forecast
1	4,85	*	*	*	*
2	3,28	3,280	*	-1,570	*
3	5,80	**4,164**	0,884	**0,148**	1,710
4	4,44	4,389	0,225	0,202	4,312
⋮	⋮	⋮	⋮	⋮	⋮
29	9,13	9,218	0,001	0,063	9,425
30	7,55	8,069	-1,149	-0,786	9,281
31					7,284
32					6,498
33					5,712

Smoothing constants: $\alpha = 0,6$, $\beta = 0,7$

Starting values:

$\hat{x}_2 = x_2 = 3,28$

$T_2 = x_2 - x_1 = 3{,}28 - 4{,}85 = -1,57$

Next:

$$\hat{x}_3 = 0{,}6 \cdot 5{,}80 + 0{,}4 \cdot (3{,}28 - 1{,}57) = \\ = 4,164$$

$$T_3 = 0{,}7 \cdot (4{,}164 - 3{,}280) - 0{,}3 \cdot 1{,}570 = \\ = 0,148$$

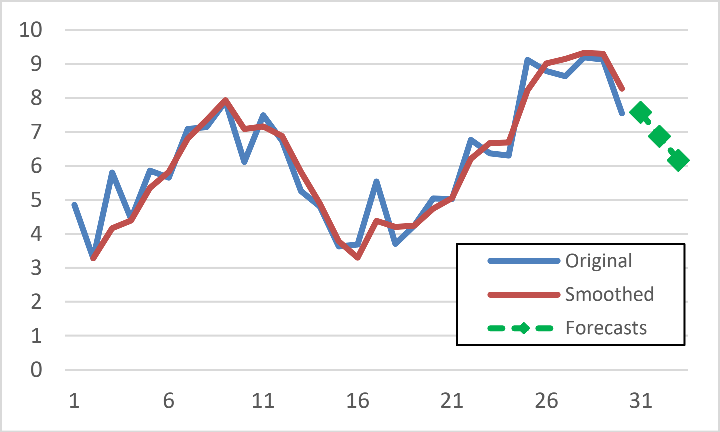
Forecasts:

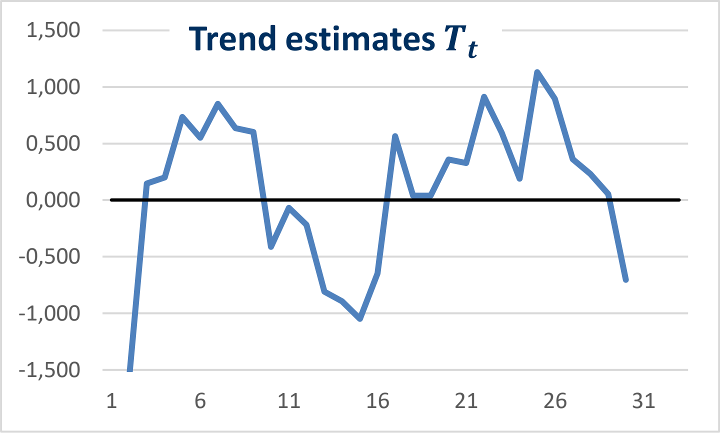
- in-sample: $\tilde{x}_t = \hat{x}_{t-1} + T_{t-1}$

- out-of-sample: $\hat{x}_{30+h} = \hat{x}_{30} + h \cdot T_{30}$

Stockholm University

Example 2, cont.





Note that future out-of-sample forecasts are decreasing. The slope is equal to the most recent trend estimate T_{30}.

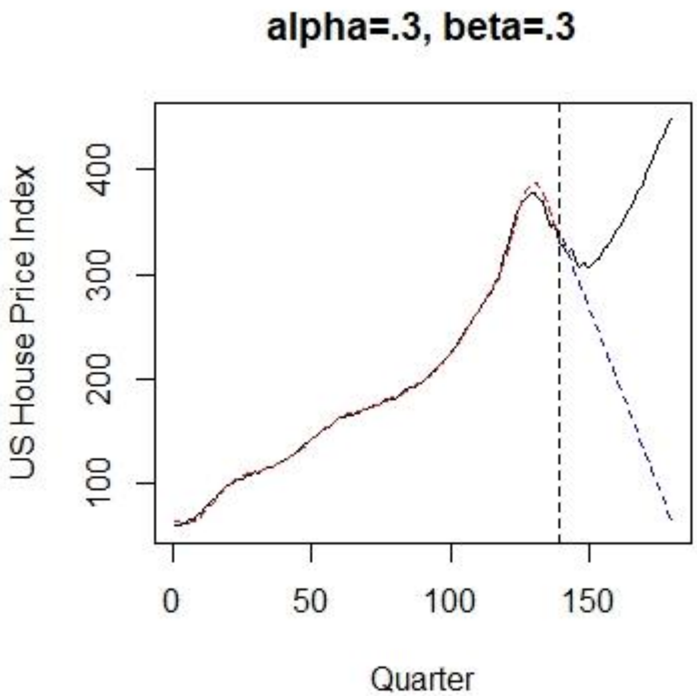
Compare to simple smoothing slide 12.

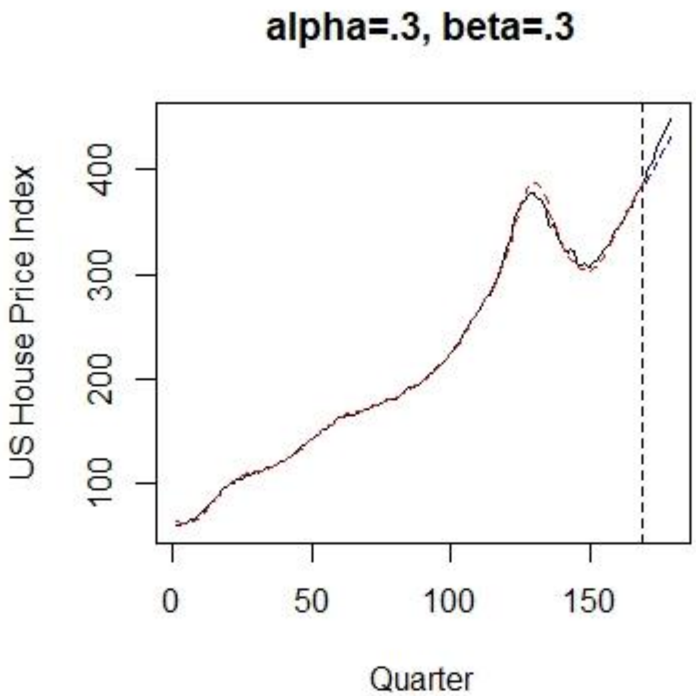
Note that the trend estimates T_t as expected vary over time. The method takes into consideration local changes in trend and cyclic patterns.

Compare to the regression method which provides a constant average trend across the entire time period.

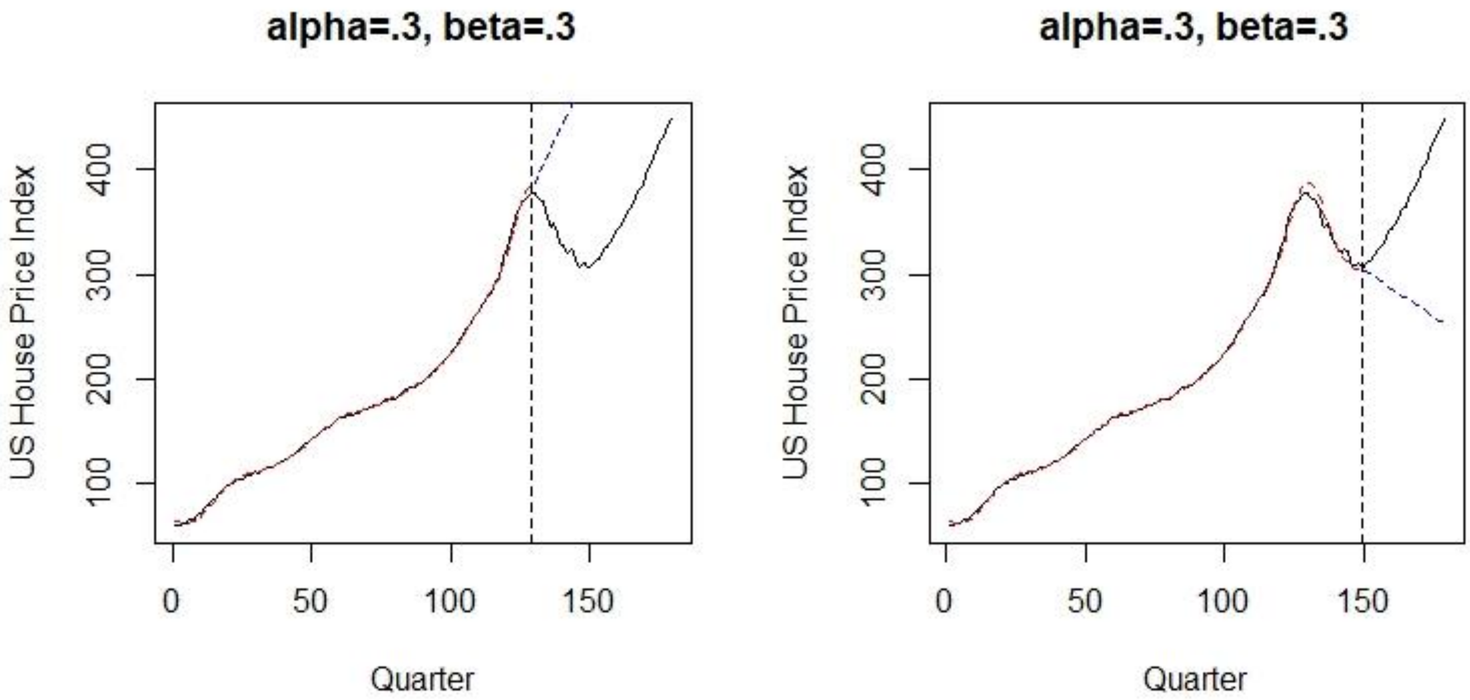
Stockholm University

Example, US House Price Index '74-'19





Example, US House Price Index '74-'19



Holt-Winter's method with seasonality

- Takes into account both **trend** and **seasonal variation**
- Three formulas:

Level: $$\hat{x}_t = \alpha x_t + (1-\alpha)(\hat{x}_{t-1} + T_{t-1})$$

Trend: $$T_t = \beta(\hat{x}_t - \hat{x}_{t-1}) + (1-\beta)T_{t-1}$$

Season: $$F_t = \gamma \cdot \frac{x_t}{\hat{x}_t} + (1-\gamma)F_{t-s}$$

- s = seasonal length; 4 for quarterly, 12 for monthly data
- Three constants and three start values to determine
- Extremely painful to do by hand! **Statistical software**!

Stockholm University

Summary

- Exponential smoothing foremost a forecasting technique
- Larger weights to most recent (i.e. latest) observations
- Retains the history of the time-series (exponentially decreasing weights)
- How much weight is determined by smoothing constants $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$
- Adjusts dynamically to changes in level, trend and seasonality
- Smoothly follows slowly changing patterns in the time-series

Summary, cont.

- Reasonably constant (or slowly changing) level, <u>no</u> trend nor seasonality:
  - **simple exponential**
- Reasonably constant linear trend, no seasonality:
  - **Holt’s method, double exponential**
- Reasonably constant linear trend plus multiplicative seasonality:
  - **Holt-Winter’s method**

Advanced methods

- AR, ARMA, ARIMA, Seasonal ARIMA
  - ARIMA = Autoregressive Integrated Moving Average
  - “Autoregressive” resembles the idea of exponential smoothing
- ARCH/GARCH
- VAR
  - Vector autoregressive, multivariate i.e. several time-series
- TRAMO/SEATS
  - Time series Regression with ARIMA noise, Missing data and Outliers/Seasonal Adjustment of Time-Series
- …
- Some of the above included in the course “Financial statistics”