L19

Basic Statistics for Economists

Spring 2020

Department of Statistics

Stockholm University

Repetition χ^2-tests

Concerns **categorical** variables (or categorized numerical)

We have **frequencies** per category (or combination of categories)

- **1** categorical r.v. ⇒ one-way-table ⇒ **Goodness-of-fit test**

Category j	1	2	…	K	Σ
Frequency O_j	O_1	O_2	…	O_K	n

- H_0: $P_1 = p_1,\ P_2 = p_2,\ \dots,\ P_K = p_K$ (can be the same for all j or different)

  H_1: at least one (two) of the above is not as stated in H_0

- $\chi^2 = \sum_{j=1}^{K} \frac{(O_j - E_j)^2}{E_j} \sim \chi^2_{K-1}$ i.e. chi-square distr. with $K-1$ d.f.

$\boxed{E_j = nP_j}$

- Reject H_0 if $\chi^2_{\text{obs}} > \chi^2_{\text{krit}} = \chi^2_{K-1;\alpha}$

Stockholm University

Repetition χ^2-tests, cont.

(or homogeneity test)

- **2** categorical r.v.'s ⇒ two-way-table ⇒ **Independence test**

Frequency O_{ij}	Category j 1	2	3	4	Σ
Category i = 1	O_{11}	O_{12}	O_{13}	O_{14}	R_1
2	O_{21}	O_{22}	O_{23}	O_{24}	R_2
3	O_{31}	O_{32}	O_{33}	O_{34}	R_3
Σ	C_1	C_2	C_3	C_4	n

- H_0: variables 1 and 2 are independent; H_1: they are dependent
- $\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij}-E_{ij})^2}{E_{ij}} \sim \chi^2_{(r-1)(c-1)}$ i.e. $(r-1)(c-1)$ d.f.

$$\boxed{E_{ij} = \frac{R_i C_j}{n}}$$

- Reject H_0 if $\chi^2_{\text{obs}} > \chi^2_{\text{krit}} = \chi^2_{(r-1)(c-1);\alpha}$

Stockholm University

Today

NCT **16.1 – 16.2**

- **Time series**
  - a sequence of observations $\boldsymbol{x_t}$, where the index $\boldsymbol{t}$ denotes a point in time
  - notation: $(\boldsymbol{x_1}, \boldsymbol{x_2}, \dots, \boldsymbol{x_n})$ or $\{\boldsymbol{x_t}\}$, $t = 1, \dots, n$
- **Components**
  - Trend, Seasonality, Cyclical (sv. *konjunktur*), Irregular
  - Additive and Multiplicative models
- **Separation of components**
  - **Moving average** or Regression
  - **Seasonal adjustment** (sv. *säsongsrensning*)

Why time series analysis?

- Study how a r.v. X_t varies over time

- Purposes (from L1):
  - Description *("this is how it looks")*
  - Explanatory, causality *("it's like this because …")*
  - **Forecasting, prediction** (*"what about tomorrow?"*)
  - **Normative, prescriptive** *("if we do this, this will happen")*

> By using available information on how a variable X has varied and changed over time we may, by conditioning on past behavior, be able to get a better idea on future values of X.

- **Better idea = increased precision = decrease in variance**

Stockholm University

Properties of a times series

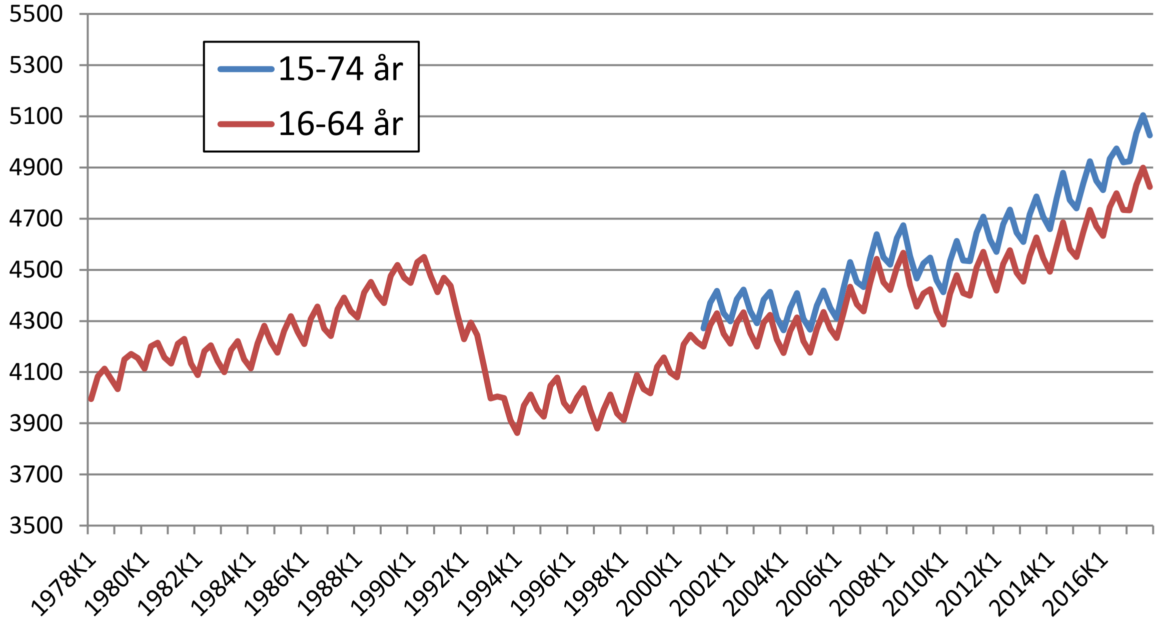
Analysis of the historic behavior with respect to

- **Trends**
  - is it going up or down?
- **Cyclicity** (sv. *konjunktursvängningar*)
  - slow long term variation; boom or recession?
- **Seasonal variation**
  - more ice cream, less *glögg* in the summer
  - **regular periodicity**; typically one-year-cycles; quarter, month
- **Irregular fluctuations**
  - unpredictable shifts that affects the phenomena under study; **random term**

Stockholm University

Ex. Time series Labor Force Survey (LFS)

- No. of people in work, age groups 16-64 and 15-74, national total in thousands by quarter, 1978:Q1 – 2017:Q4



Is the proportion of 65-74 year olds increasing? Or is it an illusion?

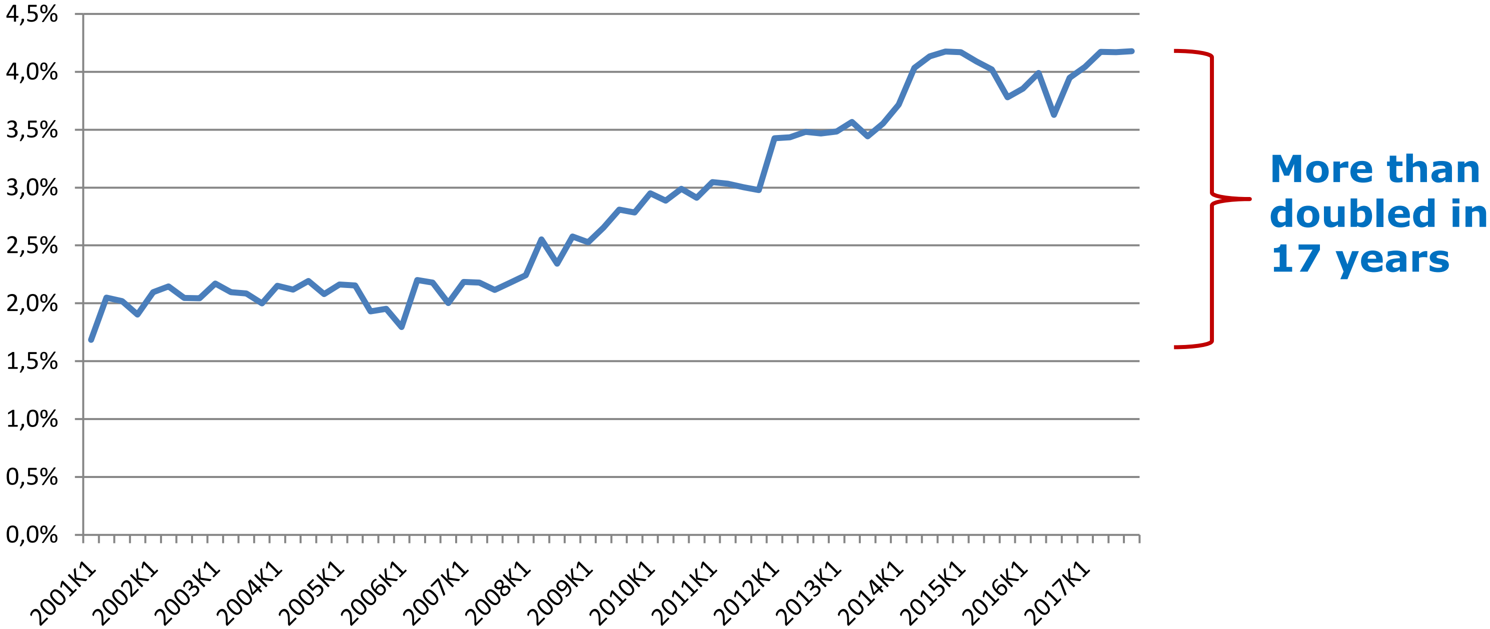
$$\text{Proportion}(\%) = 100 \cdot \frac{\text{Blue} - \text{Red}}{\text{Blue}}$$

Source: SCB (2018) Statistical database, www.scb.se, 2018-02-23

Stockholm University

Ex. Labor Force Survey , cont.

- Proportion of age group 65-74 (+ 15) in percent (%), nationwide by quarter, 2001:Q1 – 2017:Q4



Source: SCB (2018) Statistical database, www.scb.se, 2018-02-23

Stockholm University

Comparability over time

A general **quality aspect** of statistics

Ex. Substantial changes in the definitions for the LFS (EU)

- From 2005:Q2 the target population was expanded from 16-64 to include 15-74 year olds.
- Up until 2007:Q4 full-time students who had applied for and were able to take a job were attributed to the category "Not in the labor force‟; since October 2007, they are classified as unemployed.

- How do these changes in the target population affect **comparability** over time?

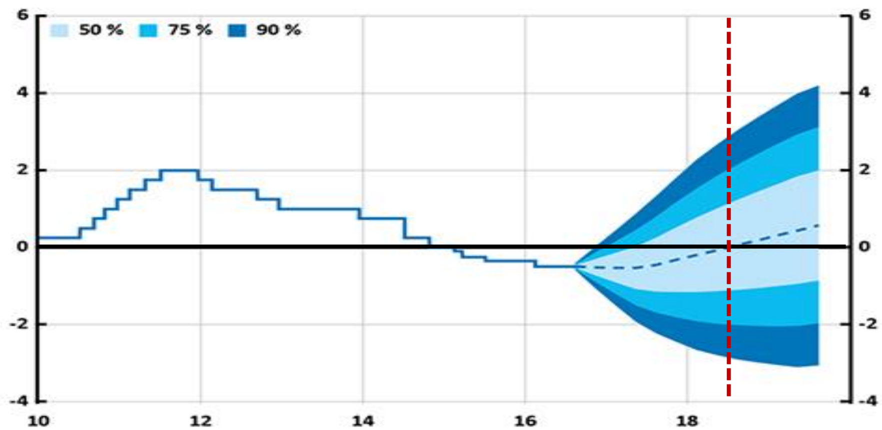
  Rhetorical question! We will in the following assume unchanged definitions and 100% comparability!

Stockholm University

Forecasting with uncertainty intervals

Repo rate
in %, quarterly averages

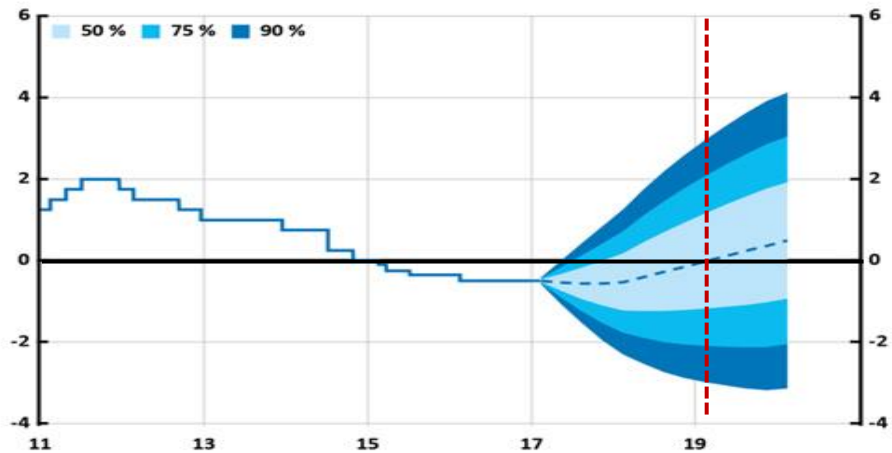
Red line marks the point when the rate is expected to turn to positive



Forecast in Sep 2016

Again five months later:

New forecast and a new **red line**

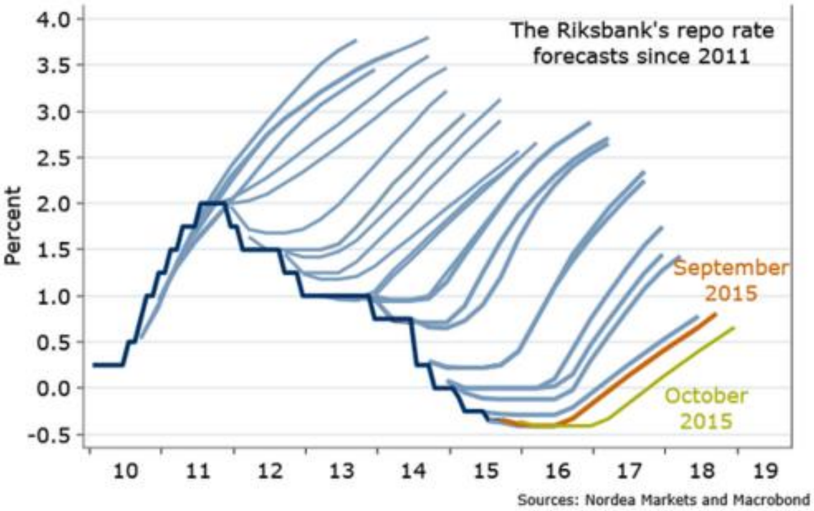


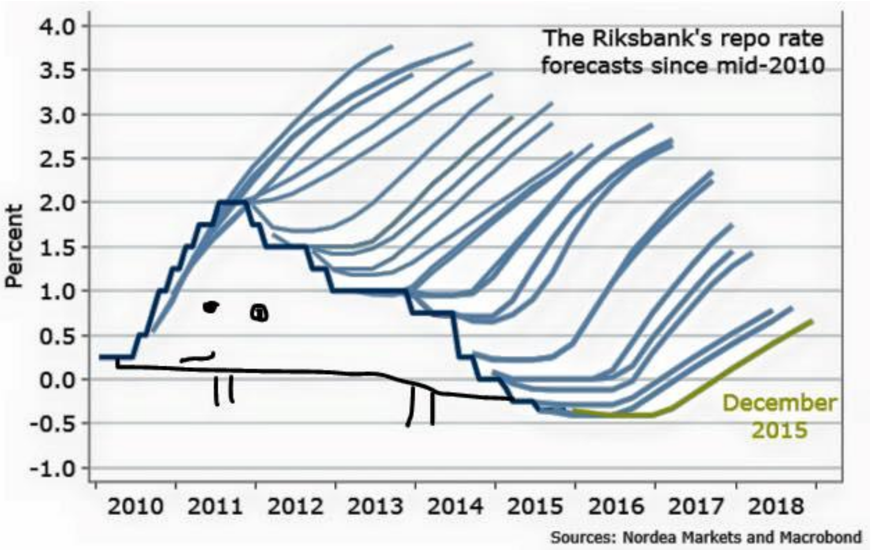
Forecast in Feb 2017

Source: Central bank of Sweden (2016, 2017) Penningpolitisk Rapport, www.riksbank.se

Stockholm University

Repo rate and the hedgehog





Stockholm University

Component models

Explorative, descriptive method! No inference!

"Simple" model with the four components:

- T_t = **Trend** – direction (up or down), an average slope
- C_t = **Cyclic** – wave-like movements, long term

  The trend and the cyclic components are often merged into one component, especially if the time span is not very long; C_t is dropped from the model and T_t may vary, "local" slope, see p 29.

- S_t = **Season** – repeated regular variation with fixed period
- I_t = **Irregularities** – unpredictable (random) variation

Stockholm University

Additive and multiplicative models

The parts are **added**: $\boxed{X_t = T_t + C_t + S_t + I_t}$

- Seasonal component is constant in <u>absolute</u> terms

The parts are **multiplied**: $\boxed{X_t = T_t \cdot C_t \cdot S_t \cdot I_t}$

- Seasonal component is constant in <u>relative</u> terms
- By taking **logarithms** a multiplicative model becomes

$$\ln X_t = \ln T_t + \ln C_t + \ln S_t + \ln I_t$$

i.e. the model is transformed to an **additive** model

Stockholm University

Trend component

Linear trend: $T_t = \beta \cdot t$

- β = the average **absolute** increase in x from t to $t+1$

Exponential trend: $T_t = \beta^t$

- β = the average **relative** increase in x from t to $t+1$
- Ex. $\beta = 1{,}04$ corresponds to a 4 % average increase
- Logarithm: $\ln T_t = t(\ln \beta) = \beta' \cdot t$ yields an additive model

Other trend models: quadratic trend, polynomial, ….

- By "estimating" the trend we can separate the seasonal component from the time series

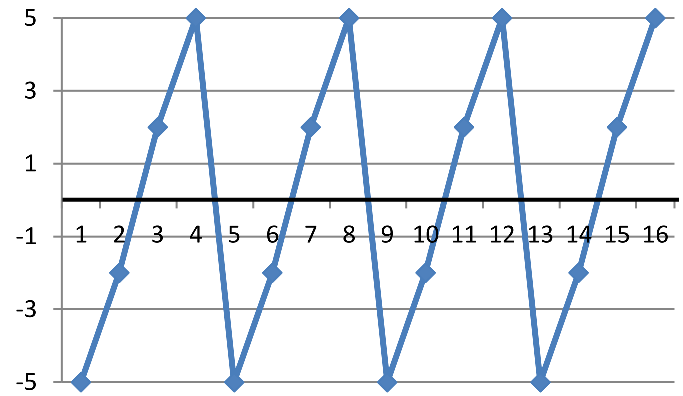
Seasonal component

- **Constant over time**
  - Quarterly data: 4 separate $\boldsymbol{S}$-values: $S = (S_{\mathrm{Q1}}, S_{\mathrm{Q2}}, S_{\mathrm{Q3}}, S_{\mathrm{Q4}})$
  - Monthly data: 12 separate $\boldsymbol{S}$-values: $S = (S_{\mathrm{Jan}}, S_{\mathrm{Feb}}, S_{\mathrm{Mar}}, \dots, S_{\mathrm{Dec}})$

Ex. an **additive** model for quarters: $\boldsymbol{S} = (\boldsymbol{-5}, \boldsymbol{-2}, \boldsymbol{2}, \boldsymbol{5})$

- Each values represent the **absolute** seasonal effect that is added $(\boldsymbol{X_t} = \boldsymbol{T_t} + \boldsymbol{C_t} + \boldsymbol{S_t} + \boldsymbol{I_t})$

- Average is $\overline{\boldsymbol{S}} = \boldsymbol{0}$



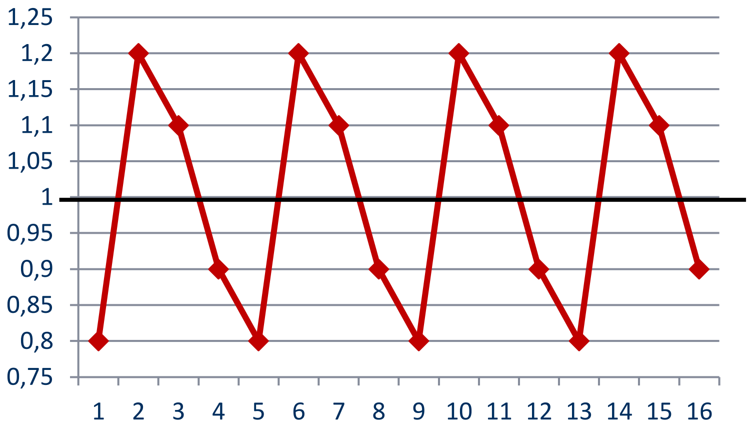
Stockholm University

Seasonal component, cont.

Ex. a **multiplicative** model, quarters: $\boldsymbol{S} = (\mathbf{0,80; 1,20;\ 1,10; 0,90})$

- Each value represents the **relative** seasonal effect that is multiplied $(\boldsymbol{X_t} = \boldsymbol{T_t} \cdot \boldsymbol{C_t} \cdot \boldsymbol{S_t} \cdot \boldsymbol{I_t})$

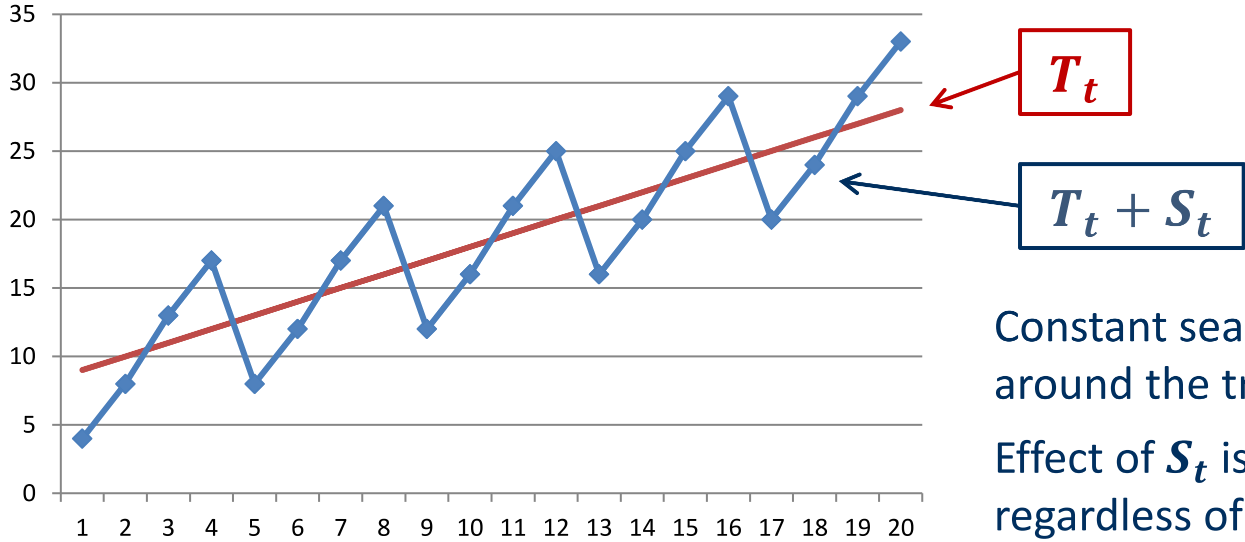
- Average is $\overline{\boldsymbol{S}} = \mathbf{1}$



Stockholm University

Ex. Additive model

- **Positive linear** trend $\boldsymbol{T_t}$, linear over time
- Absolutely constant seasonal effect $\boldsymbol{S} = (\boldsymbol{-5}, \boldsymbol{-2}, \boldsymbol{2}, \boldsymbol{5}, \dots)$

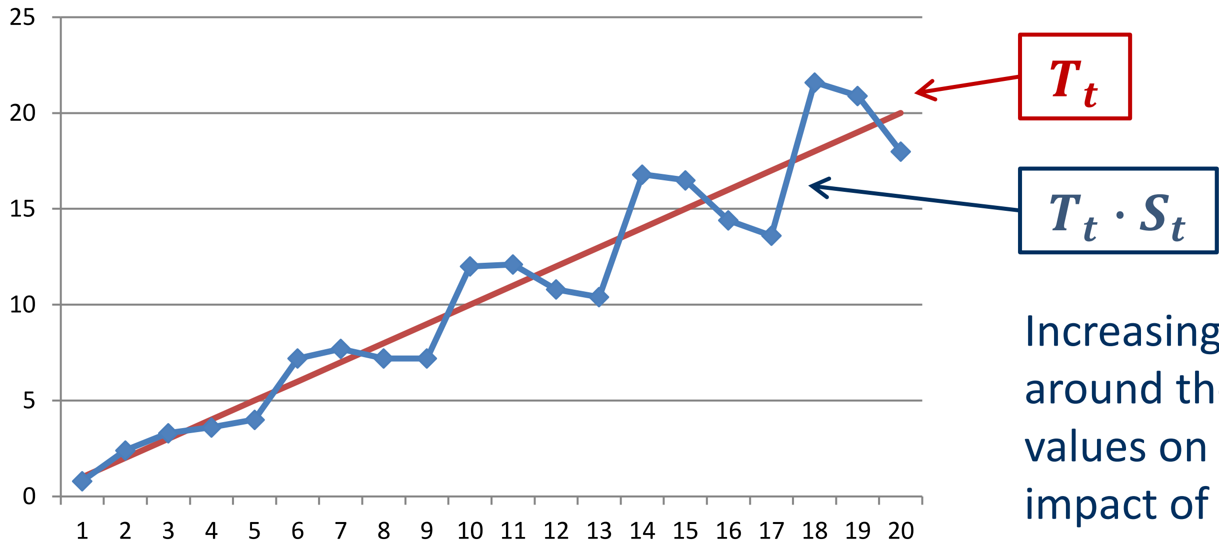


Constant seasonal variation around the trend.

Effect of $\boldsymbol{S_t}$ is the same regardless of the value of $\boldsymbol{T_t}$

Ex. Multiplicative model

- **Positive linear** trend T_t, linear over time
- Relative constant seasonal effect $S = (0,80; 1,20;\ 1,10; 0,90)$

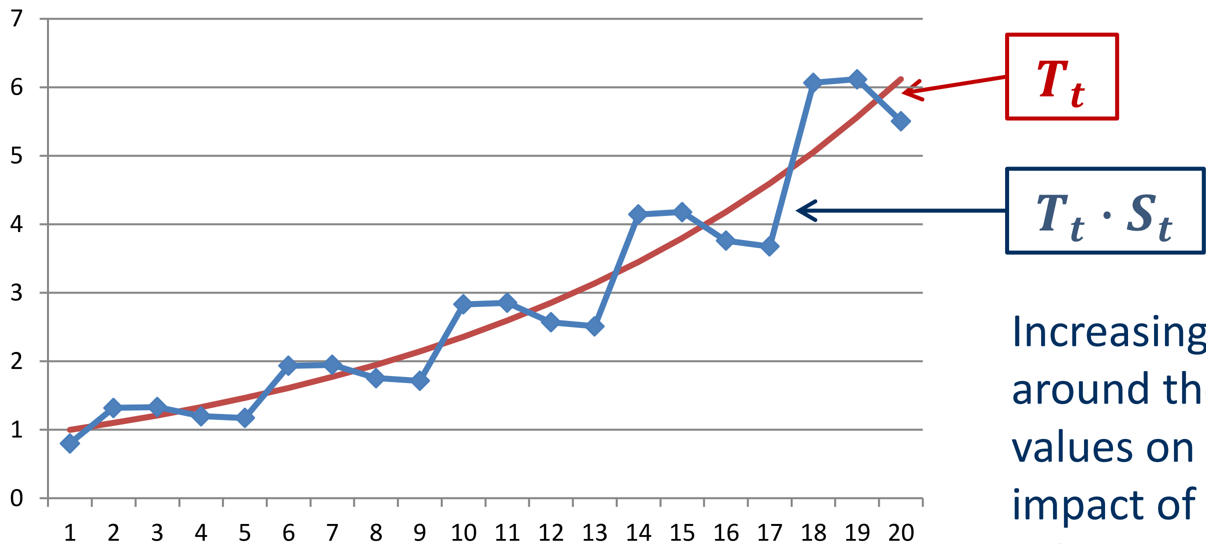


Increasing seasonal variation around the trend; for large values on T_t the absolute impact of S_t is larger but its *relative* impact is constant

Stockholm University

Ex. Multiplicative model

- **Positive exponential** trend $\boldsymbol{T_t}$ (e.g. 10% increase/year, $\beta^t = 1{,}10^t$)
- Relative constant seasonal effect $\boldsymbol{S} = (\mathbf{0{,}80; 1{,}20; 1{,}10; 0{,}90})$



Increasing seasonal variation around the trend; for large values on $\boldsymbol{T_t}$ the absolute impact of $\boldsymbol{S_t}$ is larger but its *relative* impact is constant

Stockholm University

Procedure for seasonal adjustment

- Described in **NCT section 16.2 p. 692-696**
- Assumes a multiplicative seasonal component

General case:

1. Estimate the trend (or trend + cycle)
   - regression (NCT 11.3)
   - non-parametrically using moving averages (NCT 16.2)
   - should one take logarithms first? exponential trend?
2. De-trend the time-series, subtract the trend from the series
3. Estimate the seasonal indices
4. Seasonally adjust the original time-series

Estimating the trend using regression

- Estimate the model $\boldsymbol{X_t = \beta_0 + \beta_1 t + \varepsilon_t}$
- $\boldsymbol{X_t}$ is the dependent variable and $\boldsymbol{t = 1, 2, \dots, n}$ is the explanatory variable (independent variable)

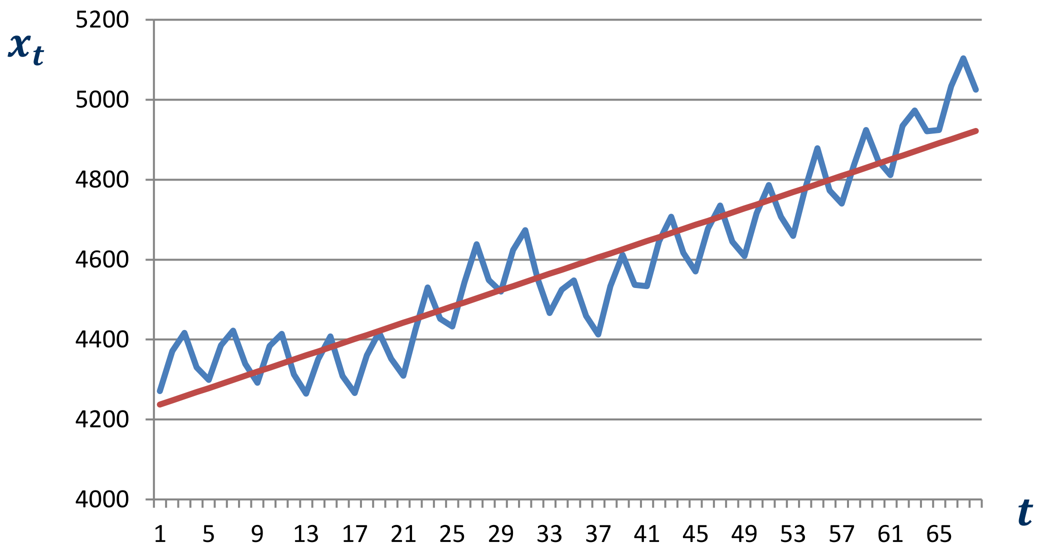
- **Descriptive**
  - e.g. ”on average $\boldsymbol{x_t}$ increases by $\boldsymbol{b_1}$ per time period”
- **Using the trend to produce simple forecasts**:

$$\boldsymbol{\widehat{x}_{n+k} = b_0 + b_1 \cdot (n+k)} + \text{seasonal component}$$

Stockholm University

Ex. LFS data, regression

- Number of occupied, thousands (blue)
- Estimated **trend** using **regression line** (red):
  - note that we aren't missing any values at the beginning and the end!



$$\hat{x}_t = 4227,6 + 10\,210 \cdot t$$

An *average increase* of about 10 210 per quarter within the observed time frame.

$$R^2 = 85,5\ \%$$

Stockholm University

Moving averages (MA), non-seasonal

- Procedure for highlighting the underlying components, making it “smoother” and less jagged due to random effects
- Calculated successively by **moving** through the series, taking the **average** of the closest observations :
- Ex. **centered 3-point MA**: $$x_t^* = \frac{x_{t-1} + x_t + x_{t+1}}{3}$$
  - short weight system $\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$
  - little smoothing effect
- Ex. **centered 5-point MA**: $$x_t^* = \frac{x_{t-2} + x_{t-1} + x_t + x_{t+1} + x_{t+2}}{5}$$
  - longer weight system $\left(\frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}\right)$
  - larger smoothing effect

Centered moving average

- Centered 3-point MA (*3-termers centrerat glidande medelvärde*)

t	x_t	3-point MA x_t^*
1	3751	*
2	3776	3791,33
3	3847	3817,67
4	3830	3852,67
5	3881	3870,00
6	3899	3899,00
⋮	⋮	⋮
42	3936	3921,67
44	3911	3912,67
45	3891	3885,67
46	3855	*

$$\boxed{x_2^* = \frac{x_1 + x_2 + x_3}{3}}$$

$$\boxed{x_3^* = \frac{x_2 + x_3 + x_4}{3}}$$

NOTE!
We lose data points in the beginning and the end of the series (*)

Stockholm University

Centered moving average, cont.

- Centered 5-point MA

t	x_t
1	3751
2	3776
3	3847
4	3830
5	3881
6	3899
⋮	⋮
42	3936
44	3911
45	3891
46	3855

5-point MA x_t^*
*
*
3817,0
3846,6
3874,8
3888,0
⋮
3916,0
3902,2
*
*

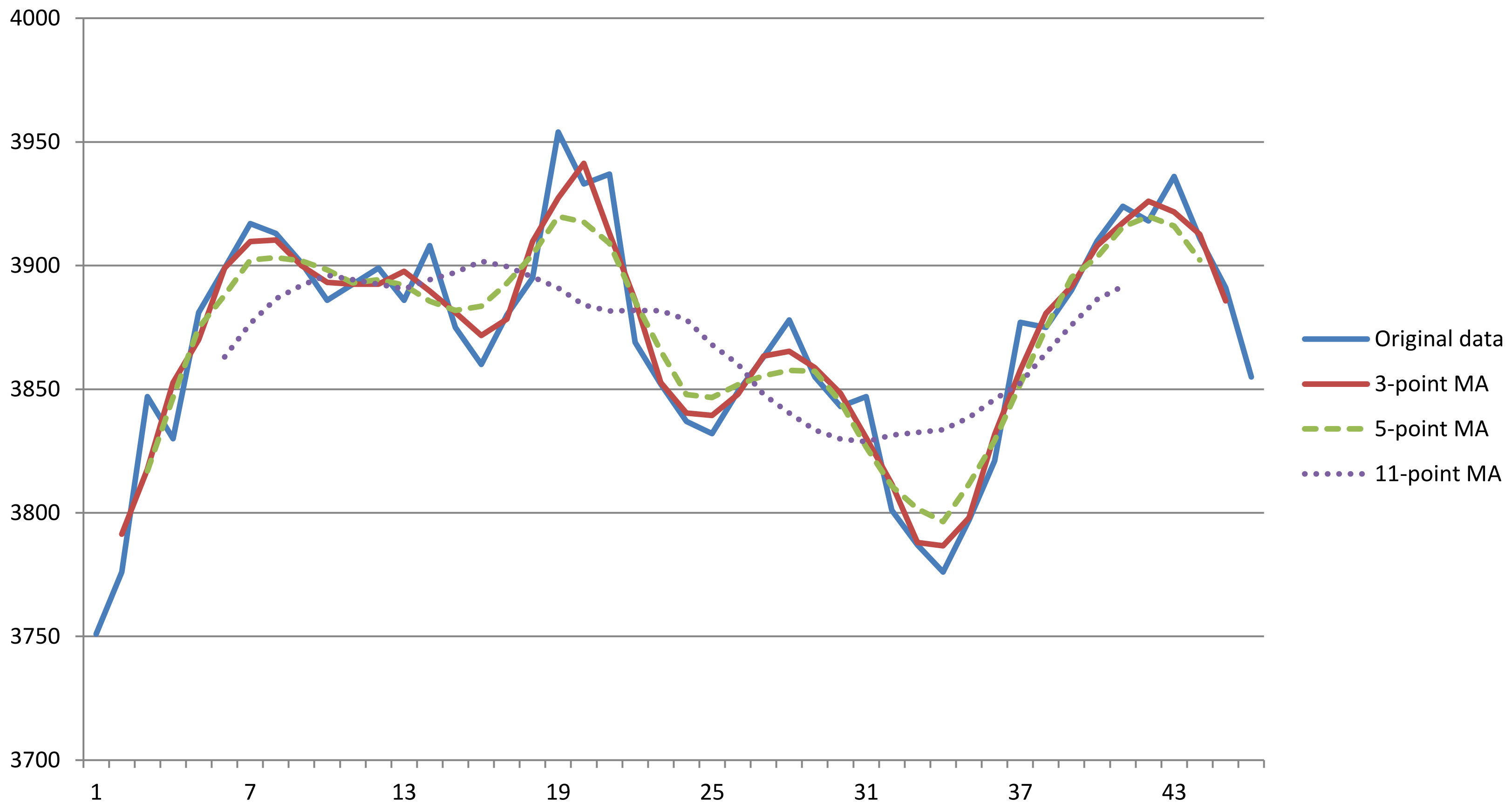
$$\boxed{x_3^* = \frac{x_1 + x_2 + x_3 + x_4 + x_5}{5}}$$

$$\boxed{x_4^* = \frac{x_2 + x_3 + x_4 + x_5 + x_6}{5}}$$

NOTE!
The more points included in the average the more smoother the time-series but we lose even more data points (*)

Stockholm University

Illustrating the smoothing effect



MA with seasonal time-series

- Same as before but with a different weight system:

- Ex. **centered** with **quarterly data**:
  - weight system $\left(\frac{1}{8}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{8}\right)$

$$x_t^* = \frac{x_{t-2}}{8} + \frac{x_{t-1}}{4} + \frac{x_t}{4} + \frac{x_{t+1}}{4} + \frac{x_{t+2}}{8}$$

Less influence from the endpoints

- Ex. **centered** with **monthly data**:
  - weight system $\left(\frac{1}{24}, \frac{1}{12}, \frac{1}{12}, \dots, \frac{1}{12}, \frac{1}{24}\right)$

$$x_t^* = \frac{x_{t-6}}{24} + \frac{x_{t-5}}{12} + \cdots \frac{x_{t-1}}{12} + \frac{x_t}{12} + \frac{x_{t+1}}{12} + \cdots + \frac{x_{t+5}}{12} + \frac{x_{t+6}}{24}$$

Stockholm University

Centered MA, seasonal data

- Can be calculated in two steps:

t	x_t
1	3751
2	3776
3	3847
4	3830
5	3881
6	3899
⋮	⋮
42	3936
44	3911
45	3891
46	3855

4-point MA
*
3801,00
3833,50
3864,25
3881,75
⋮
⋮
3914,00
3898,25
*

Centered 4-point MA x_t^*
*
*
3817,250
3848,875
3873,000
3892,125
⋮
3918,125
3906,125
*
*

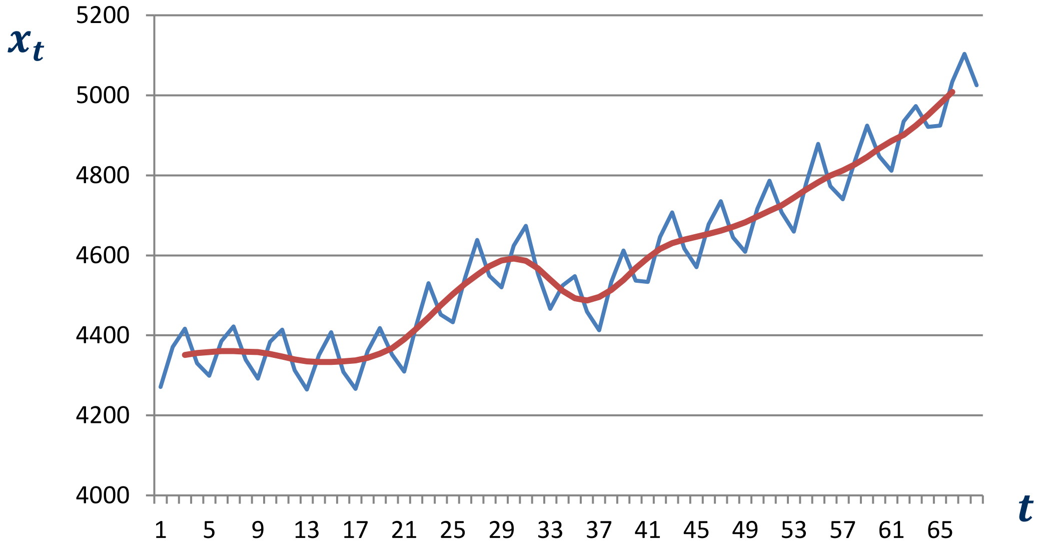
$$\boxed{x_{2,5}^* = \frac{x_1 + x_2 + x_3 + x_4}{4}}$$

$$\boxed{x_{3,5}^* = \frac{x_2 + x_3 + x_4 + x_5}{4}}$$

$$\boxed{x_3^* = \frac{x_{2,5}^* + x_{3,5}^*}{2}}$$

Ex. LFS data, centered MA

- Number of occupied, thousands (blue)
- Centered 4-point **moving average** (red)
  - note that we are missing two values at the start and end!



Captures cyclicity (*konjunktur*) better than simple linear regression, i.e. $(T_t + C_t)$; the slope can change (slowly) over time.

These features are lost with the regression model since it is an *average trend*; same slope for the entire series.

Stockholm University

Estimating the seasonal indices

1. After estimating the trend the time-series we **de-trend** it:
   - Additive model: $\boldsymbol{x_t = (T_t + C_t) + S_t + I_t} \;\Rightarrow\; \boxed{\boldsymbol{x_t - x_t^* \approx S_t + I_t}}$
   - Multiplicative model: $\boldsymbol{x_t = (T_t \cdot C_t) \cdot S_t \cdot I_t} \;\Rightarrow\; \boxed{\boldsymbol{x_t / x_t^* \approx S_t \cdot I_t}}$

We'll be focusing on the **multiplicative** model (NCT p. 692-696)

2. Group the de-trended data points by quarter (or month)
3. For each of the four quarters (or 12 months), calculate the median of the de-trended data points
4. Adjust the medians so they sum to 4 (or 12)

Estimating the seasonal indices, cont.

- Example, LFS: calculate x_t^* and **de-trend**: x_t/x_t^*

t	x_t	x_t^*	x_t/x_t^*
2001:Q1	4271.1	*	*
2001:Q2	4371.4	*	*
2001:Q3	4417.0	4350.950	1.01518
2001:Q4	4330.2	4356.225	0.99403
2002:Q1	4299.3	4358.638	0.98639
2002:Q2	4385.4	4360.375	1.00574
2002:Q3	4422.3	4360.475	1.01418
⋮	⋮	⋮	⋮
2016:Q4	4920.8	4950.750	0.99395
2017:Q1	4924.4	4979.450	0.98894
2017:Q2	5033.9	5008.775	1.00502
2017:Q3	5103.8	*	*
2017:Q4	5025.2	*	*

Group each quarter:

	Q1	Q2	Q3	Q4	
2001	*	*	1.01518	0.99403	
2002	0.98639	1.00574	1.01418	0.99530	
2003	0.98473	1.00686	1.01544	0.99378	
⋮	⋮	⋮	⋮	⋮	
2016	0.98483	1.00686	1.01003	0.99395	
2017	0.98894	1.00502	*	*	
Median	0.98425	1.00422	1.01606	0.99455	3.99909
***S* index**	**0.98448**	**1.00445**	**1.01629**	**0.99478**	4

Adjustment: seasonal index for quarter q

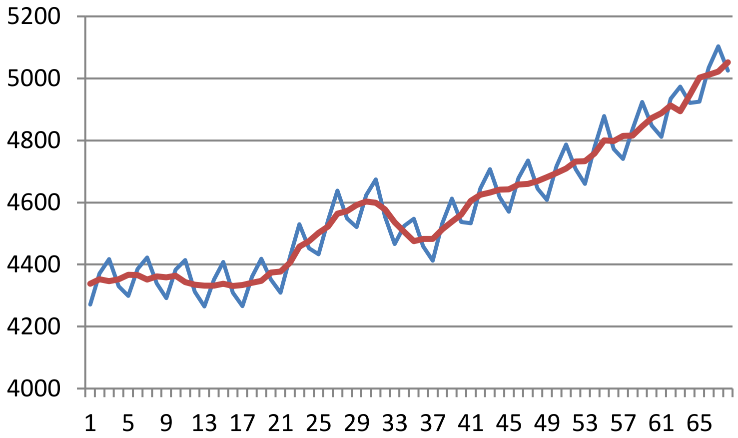
$= S_q = 4 \cdot median_q \,/\, (\sum median_k)$

Stockholm University

Seasonal adjustment

- $S = (S_{Q1}, S_{Q2}, S_{Q3}, S_{Q4}) = (0{,}98448;\ 1{,}00445;\ 1{,}01629;\ 0{,}99478)$
- Adjust: $\boxed{x_t/S_Q = \textbf{seasonally adjusted value}}$

t	x_t	S_Q	x_t/S_Q
2001Q1	4271.1	0.98448	4338.44
2001Q2	4371.4	1.00445	4352.02
2001Q3	4417.0	1.01629	4346.20
2001Q4	4330.2	0.99478	4352.93
2002Q1	4299.3	0.98448	4367.09
2002Q2	4385.4	1.00445	4365.96
⋮	⋮	⋮	⋮
2016Q4	4920.8	0.99478	4946.63
2017Q1	4924.4	0.98448	5002.04
2017Q2	5033.9	1.00445	5011.58
2017Q3	5103.8	1.01629	5021.99
2017Q4	5025.2	0.99478	5051.58



Compared to the moving averages on slide 25 the seasonal adjusted series also captures the irregularities.

Stockholm University

Comments

- A model with parameters but no assumed distributions etc.
  - Seasonal indices can be viewed as parameters but only point estimates, no inference is involved (no confidence intervals, tests)
  - Trend/slope with the regression method is parametric
  - Moving averages non-parametric,
  - Mainly an explorative method to study time-series from various perspectives (i.e. components)

- More modern methods that utilize some of the basic concepts:
  - Ex. US Census Bureau X-13ARIMA-SEATS Seasonal Adjustment Program: https://www.census.gov/srd/www/x13as/
  - Ex. EU & Bank of Spain TRAMO/SEATS: https://ec.europa.eu/eurostat/sa-elearning/tramoseats

- Computers! No one does this manually for long time-series!

Next time

A little more advanced method and extension thereof enabling forecasts

- Exponential smoothing NCT 16.3
  - Simple exponential smoothing
  - Holt-Winter's method with trend
  - Holt-Winter's method with trend and seasonality

- Some spare time for reflection … ?