Ekonomisk statistik
Economic statistics

Master course

Daniel Thorburn

Autumn semester 2012

Stockholm University

Assignments

Assignment 3 a (and b)

$$\textit{Laspeyres} \quad I_{2007}^{2008} = \frac{\sum p_{i2} q_{i2}}{\sum p_{i1} q_{12}} = \frac{102.1 * 7.1 + 13.9 * 25 + \ldots}{798.1 * 7.1 + 14.1 * 25 + \ldots} = 1.0199$$

$$\textit{Paasche} \quad I_{2007}^{2008} = \frac{\sum p_{i2} q_{i1}}{\sum p_{i1} q_{11}} = \frac{102.1 * 7.7 + 13.9 * 24 + \ldots}{98.1 * 7.7 + 14.1 * 24 + \ldots} = 1.0201$$

For the other two years the indices are

Laspeyres 1.0125 and 1.0099 Paasches 1.0021 and 1.0090

These figures are too exact, but four decimals are given here so that you may check your own computations

b) Chain index 1.0199 * 1.0125 * 1.0099 = 1.0429

Direct computation 1.0439

Assignment 3c

- Total cost for food first year 98.1*7.1 + 25*14.1 + … + 2220 = 3139.76 + 2220 = 5359.76
- Total costs second, third and fourth year are 3247.37 + 2140 = 5387.37, 3353.7 + 2240 = 5593.7 and 3315.9 + 2320 = 5635.9
- Laspeyres index formula: $$I_1^2 = \frac{\sum \omega_i I_{1,i}^2}{\sum \omega_i}$$ where the weights are consumption first year
- The total index is (5359.76*1.0199 + 5400*1.12) / (5359.76+5400) = 11514.42/10759.76=1.0701
- For the other two years the total indices are (in the same way) 1.1262 and 0.9764

Assignment 4

- The chain index is 1.125*0.857 = 0.964
- and the direct is exactly 1.00
- Note that in spite of the fact that the prices and the quantities are exactly the same the first an last year the chain index is not exactly 1. (This holds for superlative indices)

Time series – Seasonal adjustment

Daniel Thorburn

Ekonomic statistics

Autumn semester 2012

Partition into Components

- The series Y = T + C + S + K + E + I
  - T, ”Trend” is the development in the long run (What is meant by long run differs, ususally one year or more)
  - C ”Cycles” Transient deviations from the trend which are not due to the season nor limited to one measurement, the economic climate, booms and recessions
  - S ”Season” E.g. summer, christmas, school year, vacations, salary day within month … +
  - K ”Calendar effects length of month, Easter time, number of working days …
  - E ”Random Error” effects vvalid only one measurement like measurement errors, weather, lockouts, but sometimes also accrual (periodisering) effects
  - I ”Irregular components” e.g. outliers, sometimes included

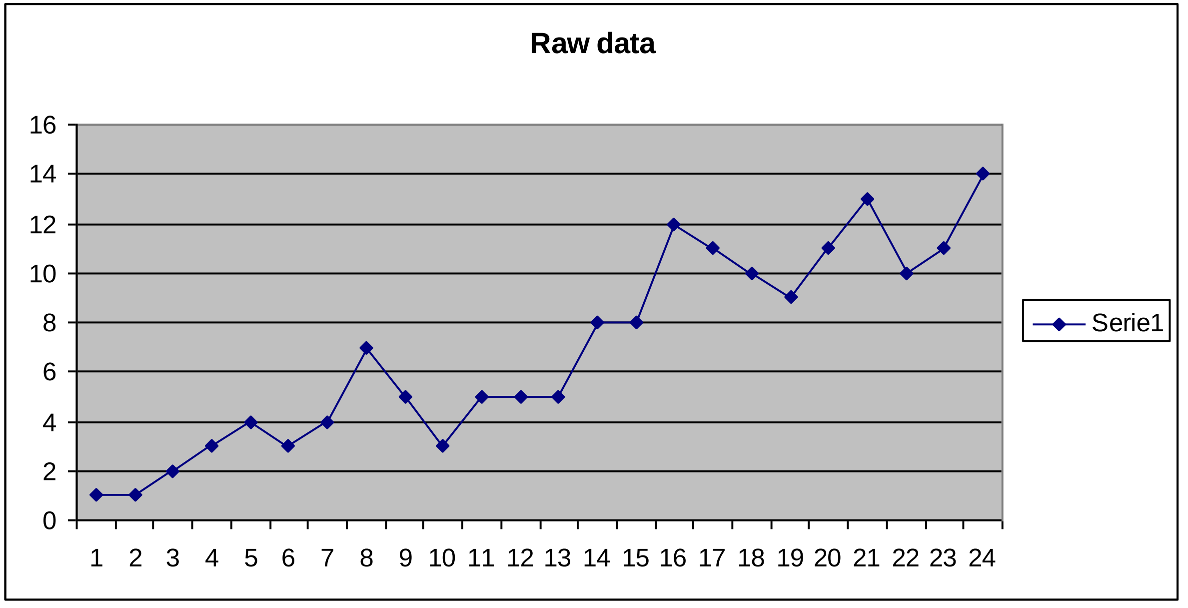
Simple methods

- Comparison with the same period last year/month/period (The Swedish economy (quarterly GDP) has gone up with 4.5 % since last year)
- Traditional
  - Trend* = Moving average (e.g. the sum of the last period (12 months))
  - Season* = (weighted) average for some (e.g.5) of the last observed values for the same season
  - Error* = Observed value - Trend* - Season*
- Seasonal adjusted series is sometimes Trend* and sometimes Y – Error*. That depends on the objective

Simple example: Model Y = T + S + E

- Data (quarterly data)

1 1 2 3 4 3 4 7 5 3 5 5 5 8 8 12 11 10 9 11 13 10 11 14



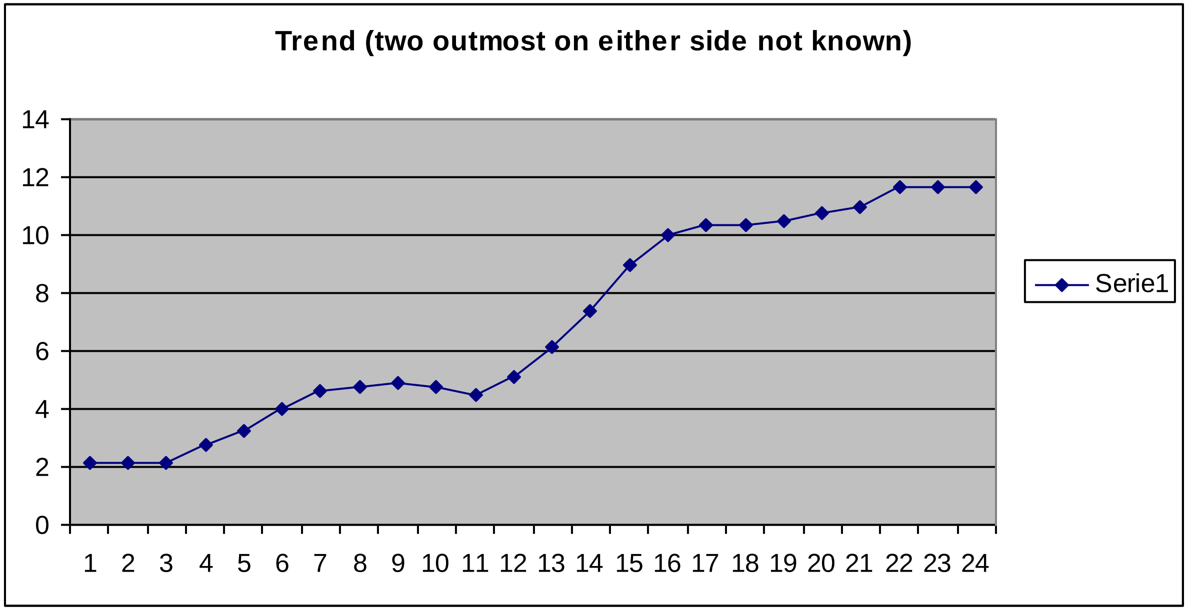
Simple example: Model Y = T + S + E

- Data (quarterly data)

  1 1 2 3   4 3 4 7   5 3 5 5   5 8 8 12   11 10 9 11   13 10 11 14

- Make a 4-quarterly average ((½*1+1+2+3+½*4)/4=2.1) for T*

  - - 2.1 2.8   3.2 4 4.6 4.8   4.9 4.8 4.5 5.1 … 10.4 10.5 10.8 11 11.6 - -



A simple example:
Model Y = T + S + E

- Data (quarterly data)

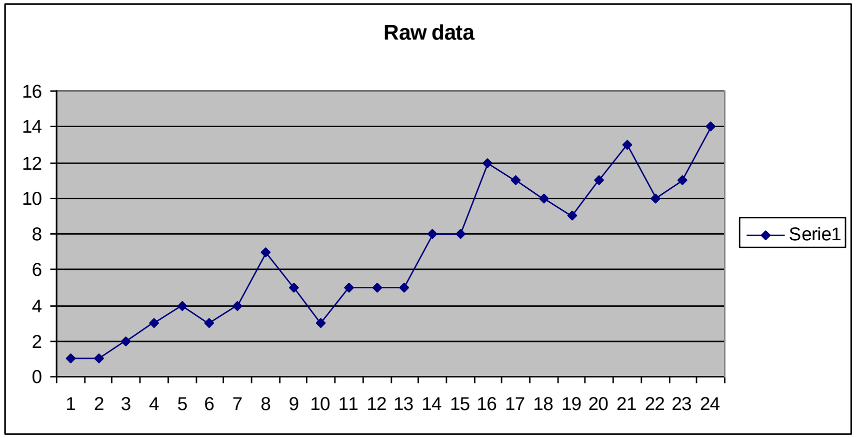
  1 1 2 3   4 3 4 7   5 3 5 5   5 8 8 12   11 10 9 11 13 10 11 14
- Make a 4-quarterly average ((½*1+1+2+3+½*4)/4=2.1) for T*

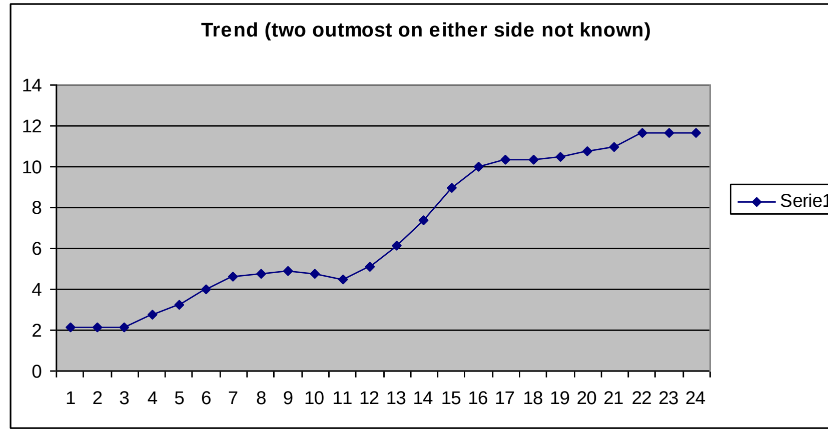
  - - 2.1 2.8   3.2 4 4.6 4.8   4.9 4.8 4.5 5.1 … 11.6 - -
- Five year moving average

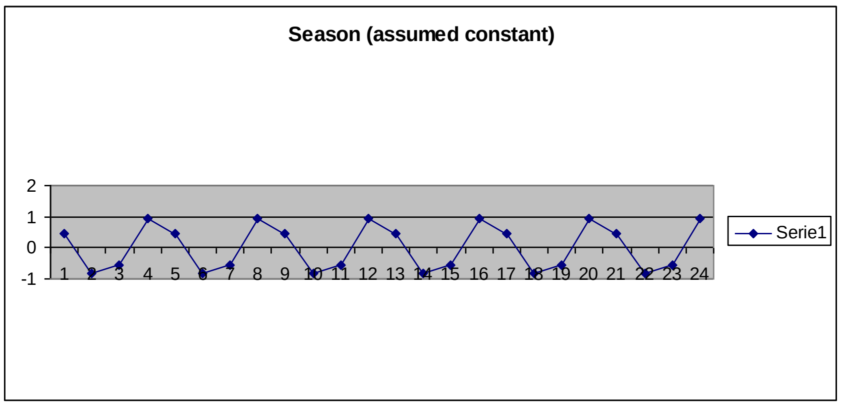
  …   - - - -   5.0 5.2 5.6 7.6   7.6 6.8 7.4 9.8 - - - - …
- Five year mean of trend, Quarterly average
- …   - - - -   - - 6.15 6.7   7.1 7.6 - -   - - - -   …
- Take the difference to get S*

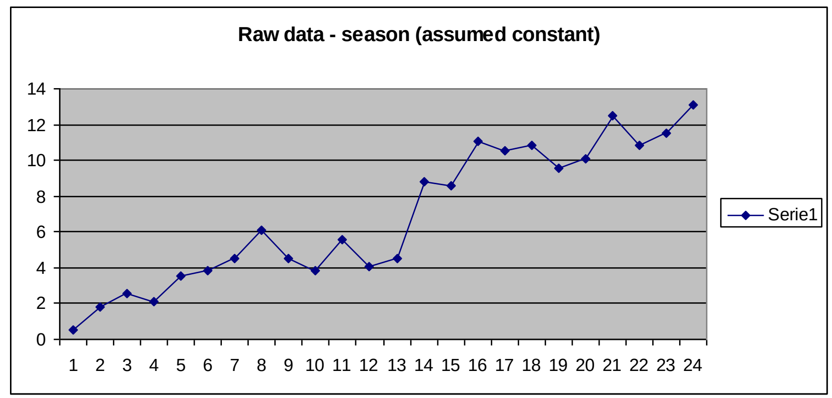
  …   - - - -   - - -0.55 0.9   0.5 -0.8 - -   - - - -   …
- The error term will be estimated as E*=Y-T*-S*

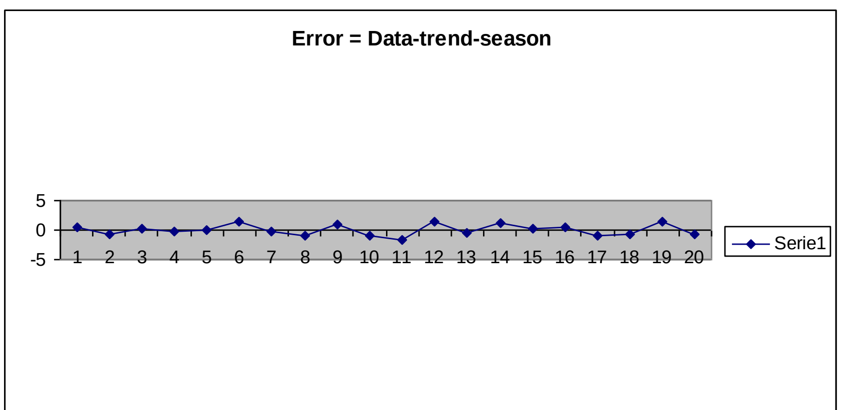
  - - - -   - - - -   0.5 1.2 0.4 -1.6 - - - -   - - - -


Raw data
0 2 4 6 8 10 12 14 16
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
Serie1


Trend (two outmost on either side not known)
0 2 4 6 8 10 12 14
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
Serie1


Season (assumed constant)
-1 0 1 2
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
Serie1


Raw data - season (assumed constant)
0 2 4 6 8 10 12 14
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
Serie1


Error = Data-trend-season
-5 0 5
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Serie1

A simple example: Model Y = T + S + E

- This was only one of many possible, but simple, methods
- But you can see the problem.
  - Trend was not estimated for the last half-year.
  - Season was assumed constant. Otherwise we would not be able to estimate the season for the two last years
  - Thus this adjustment technique will not work properly for the last two years.
  - But if we were not interested in the last years it works nicely
- There are several possible solutions. We will discuss three different ideas very shortly.

Three common approaches

- The exponential smoothing group. Remake all averages to cover only observed values but do it in an intelligent way
- The ARIMA group. Use ARIMA techniques to predict the series backwards and forwards so that you get a long enough series. Then use a simple adjustment technique on the predicted series.
- Dynamic time series models. Make a complete model for the whole series and estimate all components using the model.

The exponential smoothing - group

- Exponential smoothing $Y^*_t = \alpha Y_t + (1-\alpha)\, Y^*_{t-1}$. where α is a number between 0 and 1.

- Why does this work? What α to choose?

  We illustrate a by formulating a simple state space model: Innovation: $T_t = T_{t-1} + \eta_t$; Measurement: $Y_t = T_t + \varepsilon_t$.

$$Y^*_t - T_t = \alpha(Y_t - T_t) + (1-\alpha)(Y^*_{t-1} - T_{t-1}) - (1-\alpha)(T_t - T_{t-1}) =$$
$$= \alpha\varepsilon_t + (1-\alpha)(Y^*_{t-1} - T_{t-1}) - (1-\alpha)\eta_t$$

- Under equilibrium: $Var(Y^*_t - T_t) = (\alpha^2\sigma^2_\varepsilon + (1-\alpha)^2\sigma^2_\eta)/(1-(1-\alpha)^2)$. This is minimised for $\alpha = \sigma^2_\eta/(\sigma^2_\varepsilon+\sigma^2_\eta)$ (ABLUE, Asymptotically best linear unbiased estimate)

More complicated variants

- Exponential smoothing with a season

  $Y^*_t = \alpha(Y_t - S^*_{t-p}) + (1-\alpha)\, Y^*_{t-1}$

  (exponential smoothing for level)

  $S^*_t = \beta\,(Y_t - Y^*_t) + (1-\beta)\, S^*_{t-p}$

  (exponential smoothing for seasons)

- Holt-Winter's method without season

  (exponential smoothing with trend)

  $Y^*_t = \alpha Y_t + (1-\alpha)(Y^*_{t-1} + T_{t-1})$

  (Exponential smoothing. T is here the slope of the trend)

  $T^*_t = \gamma(Y^*_t - Y^*_{t-1}) + (1-\gamma)T^*_{t-1}$

  (exponential smoothing for trend)

- Holt-Winter's method with season …

The methods of the NSIs

(Use ARIMA models for the series and extend before adjusting)

- Large and widely used program packages.
  - Seats-Tramo is created by Maravall from the Spanish National Bank, (Standard for Statistics, Sweden)
  - X12-Arima by Dagum at Statistics, Canada (Standard in the US)
  - X11-Arima by Dagum at Statistics in Sweden (earlier version is included in SAS)

- Mostly used as black boxes.

- The principle is as above. Season + Error = Y – Moving average of past and future periods. But future are unknown.
- Estimate an ARIMA-process and predict future values using it. Use these predicted values as true ones in the moving average (or Fourier analysis).

- The choice of the the ARIMA-model will be extremely important. This is usually done, ad hoc – by looking at the the smoothed series or at the usual and the partial autocorrelationfunctions (ACF and PACF)

- In SEATS/Tramo there is a good automatic black box model for the parameter choices. When handling hundreds of time series like all branches in NACE at the two-digit-level, you must use an automatic procedure.

- The automatic version in X12-Arima is not quite as good according to a test at Statistics Sweden.

- X12-Arima is more traditional Arima modelling (in the way you may have learnt in a time series course), while SEATS-Tramo uses spectral density in the model identification phase.

Dynamic models – (more like standard theoretical statistics)

- Build models for everything, i.e. also for the latent parameters (like we did when we illustrated exponential smoothing). E.g.

  - Seasonal pattern: $S_{tp} = S_{tp-p} + \varepsilon_{stp}$
    $Cov(\varepsilon_{stp}) = \Sigma$ where $\Sigma_i \sigma_{ijp} = 0$
    (Here S_{tp} is a vector with all p components for the time period tp)
  - Trend $\Delta T_t = \Delta T_{t-1} + \varepsilon_{\Delta Tt}$ (Here t is a number)
  - ”Cycle” $C_t = \beta C_{t-1} + (1-\beta) C_{t-2} + \varepsilon_{\theta t}$
    (A stationary AR(2)), which is periodic for some parameters
  - $Y = C_t + T_t + S_t + \varepsilon_{Yt}$

- Estimate the parameters. Compute the BLUE predictors of all interesting components in the past and in the future

Transformations

- Economic positive variables are usually handled best as multiplicative, i.e. take the logarithm before analysing.
  - Multiplicative adjustment

- Other transformations like Box-Cox transformations exist and may be used if you have a good reasons

- One disadvantage, though, is that you sometimes want additivity (adjusted Trade balance should equal adjusted exports – adjusted imports)

Direct – indirect adjustment

- You have often many components e.g. in the sum BNP = Σ(all sectors). Should you
  - Adjust the total first and adjust the other components afterwards with the restriction on their sum (direct adjustment)
  - Adjust every copmonent first and then sum the adjusted components to make the adjusted total (indirect seasonal adjustment)

- There is no general agreement on the answer.
  - Those using simple smoothing techniques usually favour direct adjustment
  - Those in favour of ARIMA-models usually favour indirect if the time series models (sometimes p, q parameter) are similar but th**e**y favour direct if the models are completely different
  - Modellers adjust the series simultaneously using multivariate state space time series models. This gives efficient adjustment if the model is true but to the price of quite complicated calculations and the models may be false

- When the series are adjusted multiplicative the usual advice is to use direct adjustment

Criteria for good adjustment

- It is not easy to decide which adjustment method is the best (except for modelbased where standard rules like MSE can be used
- Three criteria should be mentioned
- Idempotency (if you adjust an already adjusted series nothing should happen). This does not hold for the exponential smoothing group or for simple methods based on averages. But it applies to methods based on dynamic models. The ARIMA-group lies in between
- Small changes when new observations are added. The adjusted value Y^*_t based on $Y_1 \ldots Y_t$ and $Y_1 \ldots Y_{t+1}$ should be similar.

- Small variations in the smoothed series e.g. in $\Sigma(Y^*_t - Y^*_{t-1})^2$ or $a\Sigma(Y^*_t - 2Y^*_{t-1} + Y^*_{t-2})^2 + b\Sigma(Y^*_t - Y_t)^2$ for some constants, a and b, when the effects of S, K and E have been removed in Y^*_t. Or just plot and check.

- Of course one can obtain a smoothed series by minimising this expression. This is called a Hodges-Prescott-filter. For large a it will very smooth almost a straight line and for large b the series will not be smoothed very much.

- When presenting a time series always tell which methods that have been used for seasonal adjustment and smoothing

- Researchers and many economic analysts usually want to use the raw data and do the adjustment themselves

- Smoothing always make the series look nicer and to have a better precision than the true precision. (But may thus be misleading)

Time series – Other aspects

Special problems with time series

- Many time series have a particular structure of the measurement errors. Which may be good to know. E.g.
- Labour force survey.
  - Every month 7/8 of the sample will be in the sample three months later and 1/8 are fresh. The auto-correlation function for the sampling error will be roughly r(t) =
    - 1 if t=0
    - ~(24-t)/24*r^t i t is divisible by 3 and less than 25 (r varies depending on what variuable you look at, but may for instance be around 0,95
    - 0 if t is not divisiable by 3 or larger than 23
  - Standard time series analysis (e.g. Arima) will find a spurious period of three months. (I have heard Swedish economists dicuss what the economic reason for this three months' period is.
- The Consumer price index changes the basket every new year.
  - If you look for that in your analysis you will find it.
- Seasonal adjustment and linkage may also create problems. E.g. an adjusted series is mostly much smoother than the true series should be. Estimated variances will thus be smaller and the economist will have easier to prove his hypothesis, since the estimated variances are smaller than the true variances.

Longitudinal studies

- Longitudinal studies is a collection term for studies where you follow the same units over time.
- The opposite is cross-sectional studies
- E.g. study the enterprises which have got regional localisation suppoert to see what happens to them and compare to similar firms which did not get support or who got some other type of support.
- Or study firms that grow several years in a row
- Longitudal and crossectional studies have different advantages.
- If you follow persons most people get an increase in the income. This does not mean that the whole population earns more, since their is a group of young persons getting their first job and another group who retires (or dies). The last group has usually higher salaries than the first group. ”The ecological fallacy”

The inspection paradox etc

- Suppose that you want to study the development of firms on the Swedish stock exchange. If you select the 10 biggest and study them for 5 years.
- Then you will probably underestimate the development. Why?

The inspection paradox etc

- Suppose that you want to study the development of firms on the Swedish stock exchange. If you select the 10 largest and study them for 5 years.
- Then you will probably underestimate the development. Why?
- The correct way would be to compare the ten largest with the 10 largest after 5 years. The largest firms can only loose their ranking.

The Inspection Paradox etc

- Select a random sample of in-patients at a certain hospital and ask them about how long their stay at the hospital have been and note when they leave the hospital
- Why will the average of these times be an overestimate of the average time in the hospital?

The Inspection Paradox etc

- Suppose that you want to study the amount of environmental investments during one year. You make a study on a random sample in September and asks many questions on September.
- Those who answered that they did an environmental investment in September are reinterviewed in a special survey and asked about all investments since last September.
- Why will this give a skew estimate of the total environmental investments?

The Inspection Paradox etc

- Select a random sample of all SJ-trains during a certain month and measure how much they were delayed.
- Why will an average of these delays be a misleading measure of how well SJ follows the time table?

Thank you for your attention