Introduktion till statistik för statsvetare

Mikael Möller

"Det finns inget så praktiskt som en bra teori"

November 2011
Let us take a look at the Swedish GDP series on a quarterly basis and in fixed prices.
Economic time series what are they?

The series of GDP has the following characteristics of economic time series:

- It is a collection of measurements of the same kind.
- The time between measurements are evenly spaced. In this case a quarter.
- The series contains a trend
- The series contains a periodicity
- The series contains irregularities
- The series contains intervention effects
Contents of this part of the course

How do we analyze such a series? We will here discuss three methods to do that. The last two are discussed in more advanced courses.

1. Moving averages also called linear filters
2. Component models and their decomposition
3. Exponential smoothing and forecasting models
4. Autoregressive Integrated Moving Averages models (ARIMA)
5. X12-ARIMA and TRAMO/SEATS

SCB uses the TRAMO/SEATS method for analyzing time series.
Definition of a time series

We will look at data that is measured in chronological order and for this data we make the following

**Definition**
An observed time series \( \{y_t\}_{t=1}^n \) is a chronological sequence of measurements where the time between any two consecutive measurements is constant.

Our main concern with time series is to make a forecast.

That’s all!

When analyzing a time series we may also find hidden patterns. Then we are happy because these patterns give us a deeper understanding of data.
Five basic steps in a time series analysis

In general, to make a forecast we need to

1. Define the problem
   How will the forecast be used? Who requires the forecast?
   How does the forecasting function fit into the organization?

2. Gather information
   a) numerical data and b) accumulated (expertise) knowledge.

3. Make a preliminary (exploratory) analysis
   Are there consistent patterns? Make plots! Is there a trend?
   Make plots! Is seasonality important? Make plots!

4. Choose and fit models
   Try different models. Sometimes a simple model is better
   than a sophisticated model.

5. Use and evaluate the forecasting model
   Save last measurements for evaluation.
Moving averages are weighted arithmetic means over a part of the time series

\[ \bar{y}_{t,k} = \frac{1}{k} \sum_{i=t+1}^{t+k} y_i \quad t = 0, 1, 2, \ldots, n - k \]

\[ = \sum_{i=t+1}^{t+k} \frac{1}{k} y_i = \sum_{i=t+1}^{t+k} a_{i-t} y_i \]

\[ \{ i := i - t \} = \sum_{i=1}^{k} a_i y_{t+i} \quad t = 0, 1, 2, \ldots, n - k \]

where \( a_{i-t} = \frac{1}{k} \) and \( \sum_{i=1}^{k} a_i = 1 \).
Moving averages – what are they good for

The simple moving averages have the following properties – depending on the choice of the $a_i$'s.

- They smooth data
- They have an extinction effect i.e. certain parts of the time series are mapped to 0.
- They map lines into lines
- They change/do not change time

(We show this with an example)
Moving averages – an example

For \( k = 4 \) and \( a_i = \frac{1}{4} \)

1. Given the time series: 5, 9, 2, 6, 7, 4, 5, 7, 3 and 8. Calculate \( \bar{y}_{t,k} \). What do we get and what problem do we have?

2. What happens if the time series has a period of 4?

3. What happens if \( y_t = \alpha + \beta t \)?

Answer the same questions when

\[
a_1 = \frac{1}{8}, \ a_2 = \frac{2}{8}, \ a_3 = \frac{2}{8}, \ a_4 = \frac{2}{8}, \ a_5 = \frac{1}{8}
\]
Smoothing

\[ \bar{y}_{0,4} = \frac{5+9+2+6}{4} = 5.50 \]
\[ \bar{y}_{1,4} = \frac{9+2+6+7}{4} = 6.00 \]
\[ \bar{y}_{2,4} = \frac{2+6+7+4}{4} = 4.75 \]
\[ \bar{y}_{3,4} = \frac{6+7+4+5}{4} = 5.50 \]
\[ \bar{y}_{4,4} = \frac{7+4+5+7}{4} = 5.75 \]
\[ \bar{y}_{5,4} = \frac{4+5+7+3}{4} = 4.75 \]
\[ \bar{y}_{6,4} = \frac{5+7+3+8}{4} = 5.75 \]
Period

Suppose \( y_{t+4} = y_t \) we then have

\[
\bar{y}_{0,4} = \frac{y_1 + y_2 + y_3 + y_4}{4} = \bar{y}_{0,4}
\]

\[
\bar{y}_{1,4} = \frac{y_2 + y_3 + y_4 + y_5}{4} = \frac{y_2 + y_3 + y_4 + y_1}{4} = \bar{y}_{0,4}
\]

\[
\bar{y}_{2,4} = \frac{y_3 + y_4 + y_5 + y_6}{4} = \frac{y_3 + y_4 + y_1 + y_2}{4} = \bar{y}_{0,4}
\]

\[
\bar{y}_{3,4} = \frac{y_4 + y_5 + y_6 + y_7}{4} = \frac{y_4 + y_1 + y_2 + y_3}{4} = \bar{y}_{0,4}
\]

\[
\vdots = \vdots
\]

Hence a period of 4 is extinguished
Suppose that $y_t = \alpha + \beta t$ then

$$
\bar{y}_{0,4} = \frac{(\alpha + \beta) + (\alpha + 2\beta) + (\alpha + 3\beta) + (\alpha + 4\beta)}{4} = \alpha + \frac{1 + 2 + 3 + 4}{4} \beta
$$

$$
= \alpha + 2.5\beta
$$

$$
\bar{y}_{1,4} = \frac{\alpha + 2\beta + \alpha + 3\beta + \alpha + 4\beta + \alpha + 5\beta}{4} = \alpha + \frac{2 + 3 + 4 + 5}{4} \beta
$$

$$
= \alpha + 3.5\beta
$$

$$
:\vdots = :\vdots
$$
In general

\[ \bar{y}_{t,4} = \frac{\alpha + (t + 1) \beta + \alpha + (t + 2) \beta + \alpha + (t + 3) \beta + \alpha + (t + 4) \beta}{4} \]

\[ = \alpha + \frac{(t + 1) + (t + 2) + (t + 3) + (t + 4)}{4} \beta \]

\[ = \alpha + (t + 2.5) \beta \]

We see that time is half-shifted.
New filter

To have comparability in time we construct a new moving average

\[
\bar{y}_{t,4} = \frac{\bar{y}_{t,4} + \bar{y}_{t+1,4}}{2} = \frac{\alpha + (t + 2.5) \beta + \alpha + (t + 3.5) \beta}{2} \\
= \alpha + (t + 3) \beta
\]

In general we have the new filter

\[
\bar{y}_{t,4} = \frac{y_{t+1} + y_{t+2} + y_{t+3} + y_{t+4}}{4} + \frac{y_{t+2} + y_{t+3} + y_{t+4} + y_{t+5}}{4} \\
= \frac{y_{t+1} + 2y_{t+2} + 2y_{t+3} + 2y_{t+4} + y_{t+5}}{8}
\]

This moving average has the weights

\[
a_1 = \frac{1}{8}, a_2 = \frac{2}{8}, a_3 = \frac{2}{8}, a_4 = \frac{2}{8}, a_5 = \frac{1}{8}
\]
Problems with moving averages

- They need information from the past and the future
- They may induce regularities where there are none
Common components in an economic time series

The usual model for an economic time series is

\[ y_t = T_t + S_t + C_t + I_t \]

where

\[ T_t = \text{trend term} \]
\[ S_t = \text{seasonal term} \]
\[ C_t = \text{business cycle term} \]
\[ I_t = \text{irregular term} \]

and to these terms we may also add calendar effects but we assume that they have been taken care of.
A theoretical example

Define the following terms

\[ T_t = 1 + 0.1t \]
\[ S_t = \sin \left( \frac{\pi}{2} t \right) \text{ period } 4 \]
\[ C_t = 2 \sin^2 \left( \frac{\pi}{34} t \right) \text{ period } 17 \]
\[ I_t = 0 \]

Now plot \( T_t, S_t \) and \( C_t \) in turn.
A theoretical example (forts)

We get the following figures for the three components above

\[ T_t \]  \[ S_t \]  \[ C_t \]

Here \( T_t \) is the trend, \( S_t \) is the seasonality with a period of four and \( C_t \) is the business cycle. Compare the GDP figure on the first slide.
A theoretical example (forts)

Let us move away from zero

\[ T_t = 1 + 0.1t \]

\[ S_t = \sin \left( \frac{\pi}{2} t \right) + 2 \]

\[ C_t = 2 \sin^2 \left( \frac{\pi}{34} t \right) + 1 \]

Plot the sum \( T_t + S_t + C_t \), \( T_t \times S_t \times C_t \) and \( \ln (T_t \times S_t \times C_t) \).
A theoretical GDP plot

Figure: $T_t + S_t + C_t$
The moving average tool

Now in the exercise above you showed that the function

\[ G_5(y_t) = \frac{y_{t-2} + 2y_{t-1} + 2y_t + 2y_{t+1} + y_{t+2}}{8}, \]

where \( t = 3, 4, \ldots, n - 2 \), has the following properties

\[ G_5(T_t) = G_5(1 + 0.1t) = 1 + 0.1t, \]
\[ G_5(S_t) = 0 \text{ if period 4,} \]

and most certainly you agree on that \( G_5(C_t) = G_5(2 \sin^2 \frac{\pi}{34} t) \) is a complicated expression.

The important thing here is that

\[ C_t \approx G_5 \left( 2 \sin^2 \frac{\pi}{34} t \right) \]
How to do a decomposition

So the moving average $G_5$ has the following impact on our time series, with linear trend and quarterly period,

$$G_5 (y_t) \approx T_t + C_t$$

From this follows that

$$\hat{S}_t = y_t - G_5 (y_t)$$

and hence we have the season.

Next do an ordinary linear regression on $y_t - \hat{S}_t$ to find the trend. This trend is denoted $\hat{T}_t$. Lastly we find the business cycle

$$\hat{C}_t = y_t - \hat{S}_t - \hat{T}_t$$

$y_t$ is now decomposed into trend, season and business cycle.
How to do a decomposition (forts)

If there had been an irregular component $I_t$ then we get

$$\hat{S}I_t = y_t - G_5(y_t)$$

and we do the following smoothing of the season: Add all first quarters for all years and take the mean. Do the same for the other quarters. That is; compute

$$\bar{s}_i = \frac{1}{m} \sum_{j=0}^{m-1} \hat{S}I_{4j+1} \quad i = 1, 2, 3, 4, \ m = \text{no years},$$

Next, for $k = 1, 2, \ldots, m$, recalculate the estimated season as

$$\hat{S}_{i+k} = \bar{s}_{i+k} - \frac{1}{4} \sum_{j=1}^{4} \bar{s}_j$$
How to do a decomposition (forts)

The trend is calculated as before and we get the business cycle with irregular component

$$\widehat{Cl}_t = y_t - \widehat{S}_t - \widehat{T}_t$$

To get rid of the irregular component we apply the filter

$$\widehat{C}_t = \frac{\widehat{Cl}_{t-1} + \widehat{Cl}_t + \widehat{Cl}_{t+1}}{3}$$

and finally we get

$$\widehat{I}_t = y_t - \widehat{S}_t - \widehat{T}_t - \widehat{C}_t$$
Decomposition example

Decompose the GDP series
Introduktion till tidsserier
Composition – decomposition methods
Exponential smoothing and forecasting models

Simple exponential smoothing

In time series analysis we are concerned with the problem

\[ y_t = f(\cdot) + \epsilon_t \]

and we wish to find the \( f(\cdot) \) to be able to make forecasts. The methods of linear filtering and decomposition are not good for forecasting. We lose data at the end (and the beginning) of the time series. Their main strength is that they give an insight into the structure of the time series. They give us information about \( f(\cdot) \). Exponential smoothing is a set of techniques that both give smoothing and forecasts.
Simple exponential smoothing

The simplest of the exponential smoothing techniques is: simple exponential smoothing.

To derive this technique we use the following line of reasoning:
For a given time series \( \{y_t\}_{t=0}^{n-1} \) we want a forecast, \( F_n \), at time \( n \). Now imagine that we had a forecast at time \( n-1 \) and then found the observed value \( y_{n-1} \). Then a reasonable value for \( F_n \) would be

\[
F_n = F_{n-1} + \alpha (y_{n-1} - F_{n-1})
\]

That is we correct the old forecast, \( F_{n-1} \), with a fraction (\( \alpha \)) of the difference between the observed value \( y_{t-1} \) and the forecast \( (F_{n-1}), (y_{n-1} - F_{n-1}) \). To use this method we need to find a value of \( \alpha \) where \( 0 \leq \alpha \leq 1 \).
Simple exponential smoothing (forts)

Rewrite the equation

$$F_n = \alpha y_{n-1} + (1 - \alpha) F_{n-1}$$

Iterate

$$F_n = \alpha y_{n-1} + (1 - \alpha) F_{n-1}$$
$$= \alpha y_{n-1} + (1 - \alpha) \left( \alpha y_{n-2} + (1 - \alpha) F_{n-2} \right)$$
$$= \alpha y_{n-1} + (1 - \alpha) \alpha y_{n-2} + (1 - \alpha)^2 F_{n-2}$$
$$= \alpha y_{n-1} + (1 - \alpha) \alpha y_{n-2} + (1 - \alpha)^2 \left( \alpha y_{n-3} + (1 - \alpha) F_{n-3} \right)$$
$$= \alpha y_{n-1} + (1 - \alpha) \alpha y_{n-2} + (1 - \alpha)^2 \alpha y_{n-3} + (1 - \alpha)^2 F_{n-3}$$

Hence

$$F_n = \alpha \sum_{k=0}^{n-1} (1 - \alpha)^k y_{n-k} + (1 - \alpha)^n F_0$$
Simple exponential smoothing forecast

As a forecast at time $n + m$, made at time $n$, we take the last known forecast

$$F_{n+m} = F_n \quad m = 1, 2, 3, \ldots$$

Why the last forecast? This follows from

$$F_{n+1} = \alpha y_n + (1 - \alpha) F_n \approx \alpha F_n + (1 - \alpha) F_n$$

$$= F_n \text{ (because } F_n \approx y_n)$$

$$F_{n+2} = \alpha y_{n+1} + (1 - \alpha) F_{n+1} \approx \alpha F_{n+1} + (1 - \alpha) F_{n+1} = F_{n+1}$$

$$= F_n \text{ (because } F_{n+1} \approx y_{n+1})$$

and so on. Here we impute $F_n$ for $y_n$ and $F_{n+1}$ for $y_{n+1}$ because these are the best estimates available.
Simple exponential smoothing forecast interval

For the computed forecast $F_{n+m}$ we have the following symmetric forecast interval with confidence degree 95%

$$F_{n+m} \pm \lambda_{0.025} s \sqrt{1 + (m - 1) \alpha^2}.$$

It is really not possible to construct a forecast interval – we do not have a statistical model. We only have an iterating procedure.

But it may be shown that simple exponential smoothing is a special form of the statistical model ARIMA(0, 1, 1).

Hence it is relevant to construct a confidence interval for the forecast of simple exponential smoothing.
When is simple exponential smoothing useful

When the time series is slowly changing and its mean is fairly stable.

When you need to update thousands of individual series at the same time.

Example

In a big city there are thousands of newsstands. With aid of simple exponential smoothing it is quite easy to get an estimate of how many papers to deliver to a particular newsstand.

The main model where we use simple exponential smoothing is

\[ y(n) = \mu(n) + \epsilon \]

where \( \mu(n) \) is slowly changing.