Statistical Literacy for Managers Analyzing Time Series Data

Anders Wallgren, Britt Wallgren[1]
Örebro University and Statistics Sweden

Abstract

Time-series analysis differs from other parts of statistical science – instead of estimating quantitative parameters as regression coefficients or population totals the scope is how to get a good picture of qualitative patterns of the time-series studied. Our experiences from enterprises show that monthly data are often misunderstood or not used in the most efficient way. Even though managers as a rule are clever people with a university background they lack the statistical literacy necessary.

National Statistical Agencies (NSI:s) produce and publish thousands of time-series. The NSI:s have a great responsibility to promote statistical literacy among their users so that these time-series are interpreted in a correct way. The universities are responsible for teaching statistical literacy. The students should get the literacy necessary for their future jobs. According to our experience, the teaching of times-series analysis is not at all sufficient in volume, and is often technical only.

This paper can be used as a synopsis for a course in applied time-series analysis where statistical literacy is a central issue. A number of cases are presented that are relevant to persons that consume or produce monthly or quarterly reports as a part of their jobs.

Key Words: Administrative data, statistical graphics, time-series analysis, statistical information systems, teaching of statistics.

1. Introduction

Between 1968 and 1989 we were teaching statistics to students of economics and business administration. Between 1985 and 2005 we were active as business consultants at a number of manufacturing enterprises. During the period 1989-2009 we worked at the Department of Research and Development at Statistics Sweden. Experiences from these different fields have given us insight into how important statistical literacy is for the implementation of good statistical practice regarding statistical time-series.

2. Cross section data and time-series data

According to our opinion, times-series data are more important to large groups of users of statistical information than cross section data. A distinctive feature of time-series analysis is also that the same time-series data are regularly reported monthly or quarterly to the same group of users – within an enterprise, monthly reports on sales and profits are used

[1] ba.statistik@telia.com

by the managers for monitoring and decision making. Within central government the monthly and quarterly time-series produced by the National Statistical Institute are analyzed and used to monitor the economic development of the country. In this way time-series reporting constitutes an information system with a stable relation between those who report and those who use the information. In Chart 1 below, we compare cross section data and time-series data. If the tables with cross section data are interesting, then the charts with the corresponding time-series will be even more interesting.

Chart 1. Cross section data compared with time-series data

Cross section data:

Sweden

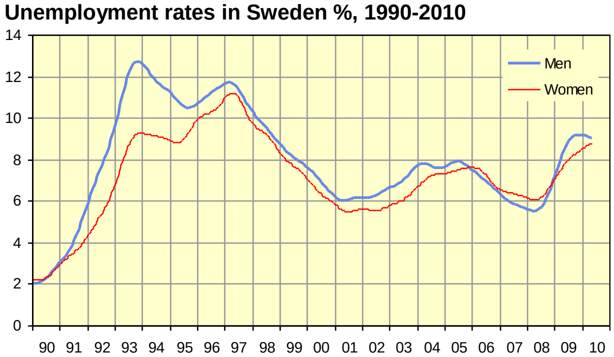
May 2010	Unemployment rate
Men	9.1 %
Women	8.6 %
Both sexes	8.9 %

Enterprise X

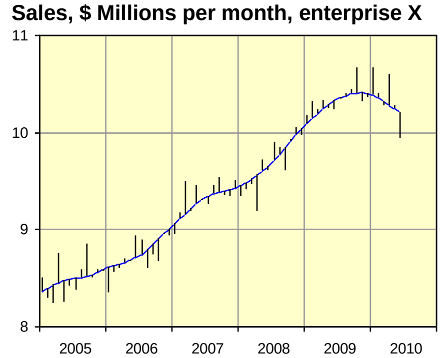
June 2010	$ Millions
Sales	9.946
Costs	9.012
Profit	0.934

Time-series data:

Unemployment rates in Sweden %, 1990-2010



Sales, $ Millions per month, enterprise X



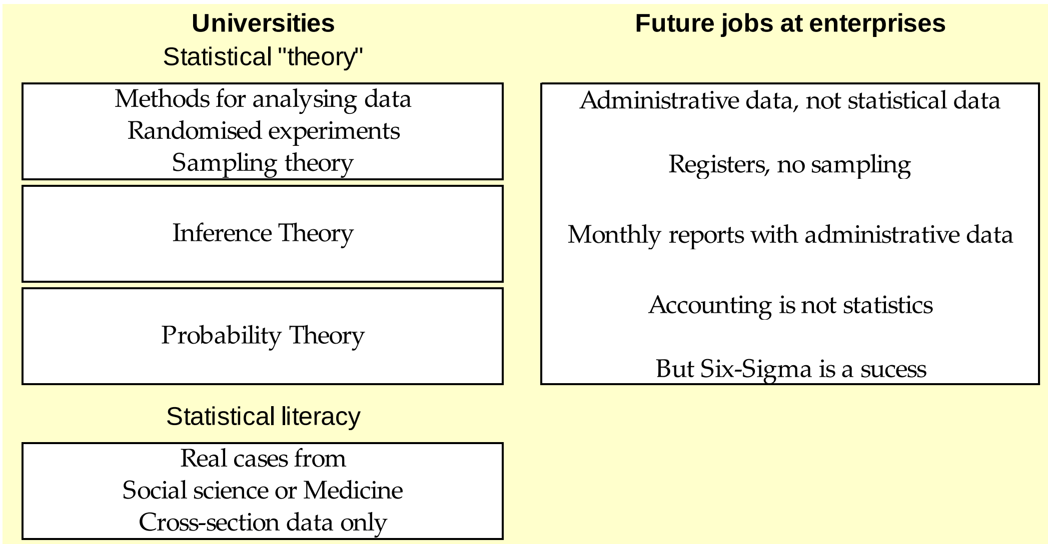
Statistical time-series analysis differs from other parts of statistical science – instead of estimating *quantitative parameters* as regression coefficients or population totals the scope is how to get a good picture of *qualitative patterns* of the time-series under study. Even if the scales on the y-axes in Chart 1 above were taken away, the time-series patterns are still interesting and interpretable. The economic crisis in Sweden in the beginning of the nineties is very clear and even without a scale on the y-axis we can see that the Swedish labour market now has started to recover from the Lehman-Brothers-crisis. In the chart describing Enterprise X, we can interpret the time-series pattern – sales are going down and it is perhaps time to reduce the production capacity and cut costs. Due to the fact that qualitative time-series patterns are important you need statistical literacy of a special kind to understand time-series data.

3. The role of time-series at universities

Despite the fact that the understanding of time-series is very important for many users of statistical information, university education does not reflect this importance. Courses in statistical science are dominated by cross section data, e.g. very few students at the master programmes in Sweden take courses in time-series analysis. Many students of engineering and business administration take courses in statistics, but there is as a rule no applied time-series analysis that is relevant for their future jobs as managers.

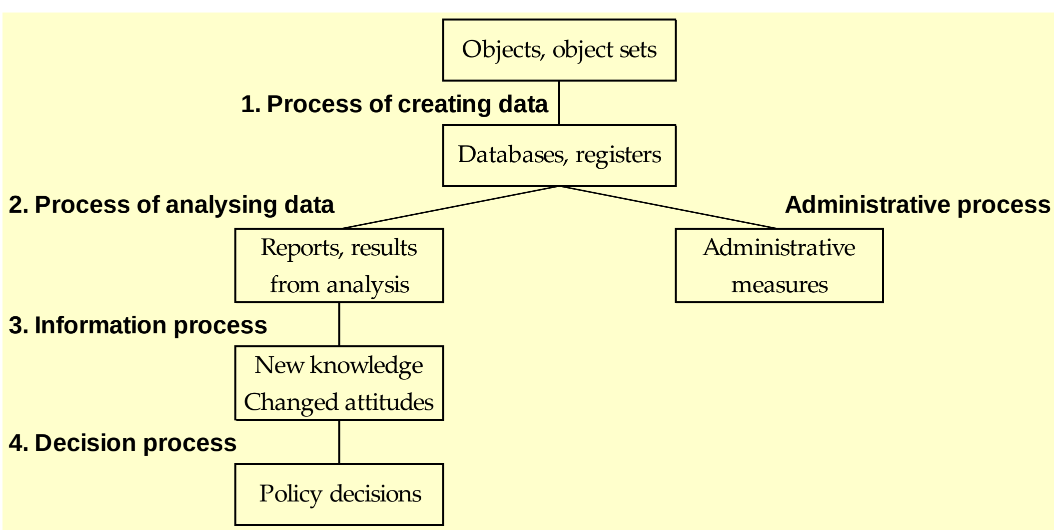
There is also a gap between the teaching of statistical theory that is mathematical in nature and the teaching of statistical literacy that is not mathematical. The teaching of statistical literacy as a rule builds on good case studies and should promote a deeper understanding of non-mathematical statistical principles. These case studies should be important for the students' areas of interest.

Chart 2. The gap between university education and the students' future jobs



Courses in probability, statistical inference and courses that build on this theory dominate statistical education. But in their future jobs the students of business administration or engineering will not meet data of the kind that is discussed during courses in inference or sampling. Instead they will meet administrative data that is stored in data warehouses. No sampling is used; instead the main sources of statistical information within a manufacturing enterprise are the registers with administrative data that are created during the production, sales and accounting processes. Statistics on production, sales and costs are based on what we call *register-based surveys* which are discussed in Wallgren (2007).

Chart 3. The structure of a combined administrative and statistical information system



At an enterprise systems of this kind are created by the IT-staff to meet the managers' demands for statistical information. As a rule, no statisticians with experience of this kind of applications are consulted. As a consequence there can be important statistical possibilities of the system that are overlooked. Quality issues can also be overlooked, and bad estimates and misleading ways of describing data can be used. Lack of statistical literacy is often the cause of these shortcomings of the statistical part of the system.

Statistical literacy and the statistical paradigm are two closely related factors. It is only when you look upon a profit and loss statement as a statistical report that you start to use your statistical skills when you read it. We think that the academic paradigms often are too narrow. Teachers of statistics have a paradigm based on probability and statistical inference and have in many cases difficulties when they want to combine theory with the professional future of their students. Also, teachers of accounting or informatics cannot relate their subjects to statistical science.

4. The role of time-series at a manufacturing enterprise

A manufacturing enterprise that produces thousands of articles and has hundreds of customers creates huge amounts of data during the processes of production, sales and accounting. But how should these data be analysed? We have found that in many enterprises, the process of analysing data is the weakest part of the information system illustrated in Chart 3 above. Let us have a look on a monthly profit and loss report – how is this kind of report used by the managers? Chart 4 below is an example of how a monthly report can look like – monthly values are compared with the corresponding month previous year. This is an inefficient way of using available data – of the last 13 observations only two are used. Accumulated sums for different periods are also compared. This will delay information on changes in the series.

Chart 4. Profit and loss statement for enterprise X, June 2010, $ Millions

	Outcome		**Budget**		**Previous year**	
	June 2010	Jan-Jun 2010	June 2010	Jan-Jun 2010	June 2009	Jan-Jun 2009
Sales	20	112	24	123	19	117
Costs	18	99	18	104	17	105
Profit	2	13	6	19	2	12

A central issue for the CEO at an enterprise is to decide if the production capacity should be increased or decreased. The time aspect is fundamental – how can managers understand how the enterprise is developing? For a number of important time-series describing sales, costs, profits etc they want know:

a) Is it going up or down?
b) How much is it going up/down?
c) When did the change happen?
d) Why did the change happen?

A course in applied time-series analysis can start here. We will here give a number of examples that can be used for teaching statistical literacy regarding time-series data. The examples are chosen so that they are relevant for students that are planning to get jobs at manufacturing enterprises.

4.1 Is it going up or down?

The following three charts illustrate three pitfalls in time-series analysis:

– Chart 5a: Yearly data can be misleading as changes within years are masked.
– Chart 6a: Seasonal variation masks the trend in the monthly series.
– Chart 7a: Random variation masks the trend in the monthly series.

After discussion Charts 5b, 6b and 7b below are shown. Chart 5b shows how important quarterly data are as compared with yearly data. Also the importance of seasonal adjustment and trend estimation are shown in these charts. It is only Charts 5b, 6b and 7b that are suitable for decision-making. These charts are what we call needle-charts in Wallgren et al (1996) and show the estimated trend as a curve and the seasonally adjusted values at the ends of the needles.

Chart 5a. Is this series going up or down?

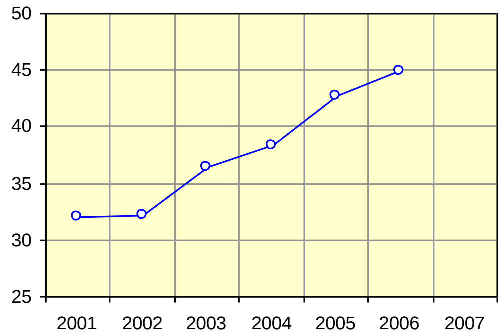


Chart 5b. It is going down!

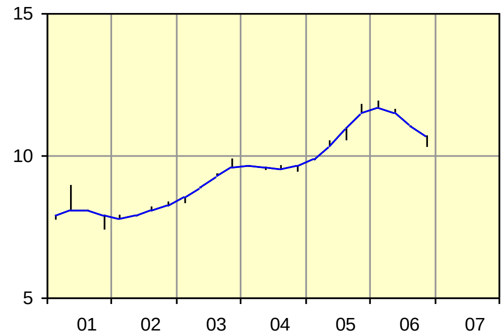


Chart 6a. Is this series going up or down?

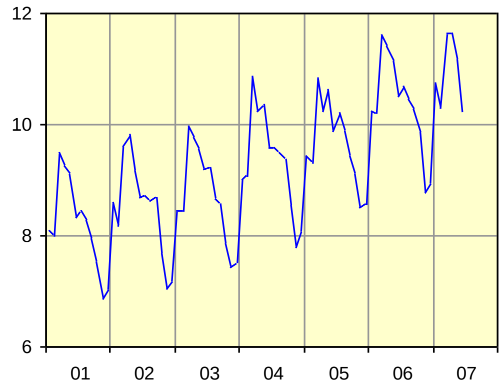


Chart 6b. It is going down!

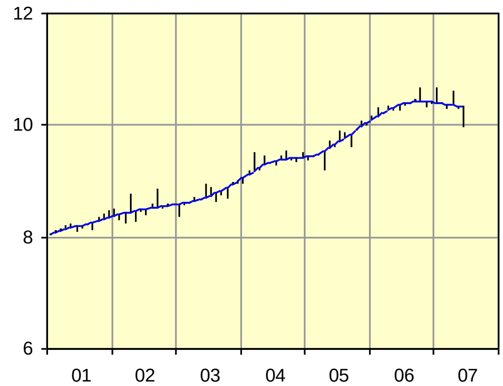


Chart 7a. Is this series going up or down?

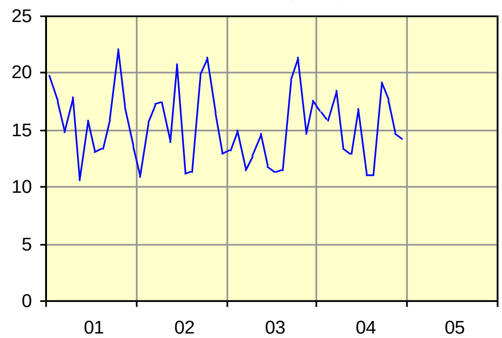
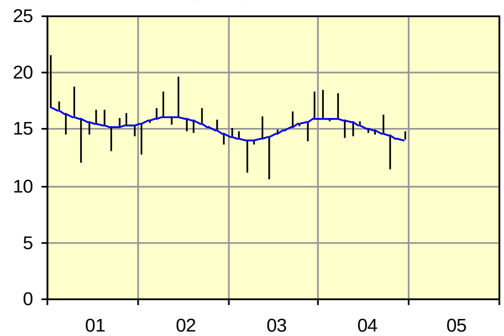


Chart 7b. It is going down!



4.2 How much is it going up/down – When did the change happen?

The next issue that should be discussed is how to measure the rate of change – how much are these series going down? The most common way to describe rate of change in a monthly or quarterly series is to compare with the same month or quarter preceding year. However, very few people understand that this measure is bad.

The examples in Chart 8 explain this. The time scale goes from December year 1 to January year 3 in Chart 8a and 8b. If we compare January year 3 with the values for January year 2 there is zero change. That measure is misleading, and if we instead compare December year 2 with December year 1, this measure says "up" in Chart 8a and "down" in Chart 8b which is completely misleading. In these simple examples it is easy to see that this measure is bad, but statistical reporting in business, media and also of official statistics use this measure to a very large extent.

Chart 8a. No change, is that true?

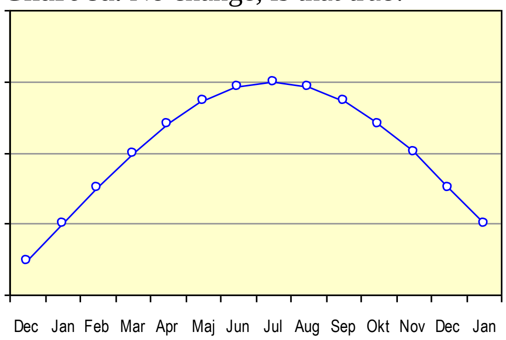


Chart 8b. No change, is that true?

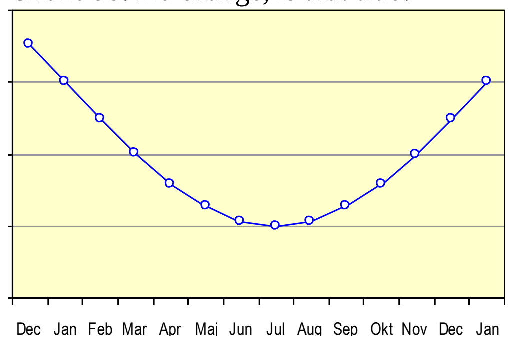


Chart 8c. The 2008 crisis, when did employment start to go down?

		Employed thousands Original values	Rate of change thousands per year	Employed thousands Trend values	Rate of change thousands per year
2008	Jan	4 390.6	41.7	4 487.4	57.2
2008	Feb	4 435.7	98.7	4 491.8	52.5
2008	Mar	4 433.6	104.3	4 495.6	46.2
2008	Apr	4 449.6	56.7	4 498.5	34.7
2008	May	4 480.2	47.7	4 499.8	15.8
2008	Jun	4 579.0	79.1	4 499.2	-7.7
2008	Jul	4 692.6	32.3	4 496.3	-34.9
2008	Aug	4 569.8	5.1	4 490.9	-64.4
2008	Sep	4 464.2	33.8	4 483.3	-91.7
2008	Oct	4 477.8	19.2	4 473.8	-114.4
2008	Nov	4 444.0	-30.7	4 462.8	-131.6
2008	Dec	4 405.2	-22.2	4 450.9	-142.6
2009	Jan	4 371.9	-18.7	4 438.6	-147.6

Data from the Swedish Labour Force Survey in Chart 8c can be used to compare different measures of change. If we use the original values and compare with the same month preceding year we se the first sign of turning down in November 2008. If we instead use the trend values and compare with the trend value for the previous month, then we see the first sign of turning down in June 2008, six months earlier. ARIMA-models can also be used as a method to describe the time-series pattern. The extrapolated values at the end of the series give a good description of the present rate of change.

4.3 When did the change happen – Why did the change happen?

A very common way of describing trend is to compute accumulated sums for the last 12 months. In charts these sums are located at the last month in the period for the sum. Charts of this kind we have seen in many enterprises. If we divide with 12 this corresponds to moving averages that are not centred.

In the beginning of our first case as business consultants we saw that at the enterprise we were working together with, they usually described their sales with charts similar to Chart 9a below that is based on data up to July 1985. They had no price index for their sales prices so all charts were in current prices. At the first meeting with the managers we presented Chart 9b and our presentation caused great astonishment. We had to explain the price index we hade computed and the difference between our X11-trends and their accumulated sums. The fact that accumulated sums delay signals with about six months was considered sensational. The arrows in Chart 9a-b illustrate these delays.

Chart 9a.
Accumulated sums/12 at *current* prices

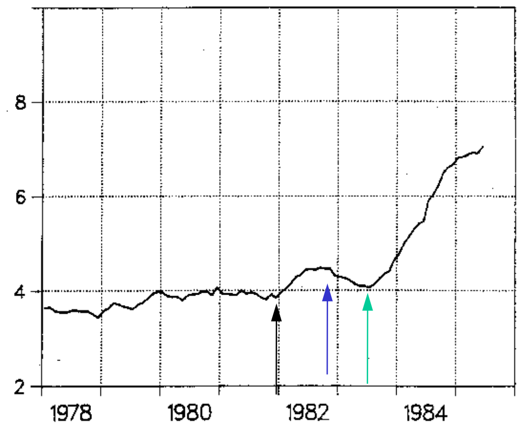


Chart 9b.
Estimated trend at *fixed* prices

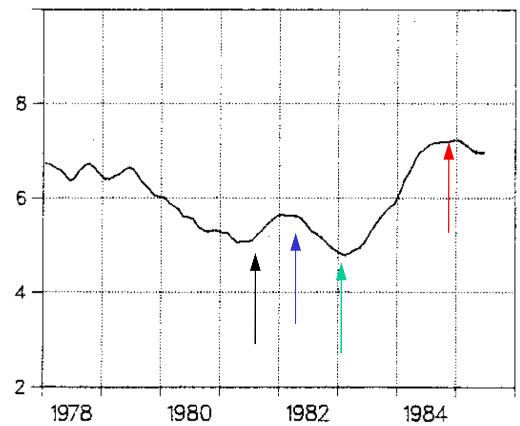


Chart 9c. Sales of one commodity

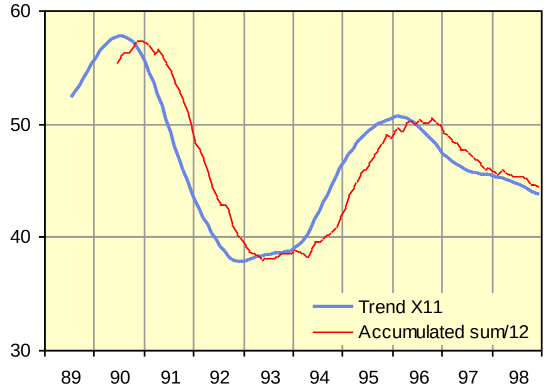


Chart 9c describes sales for one commodity and the delay is very clear.

In Chart 9a sales are in current prices. This means that changes can be caused by changed volume and/or changed prices. We always compute price indexes and sales at fixed prices. In this way we can describe prices only in one chart and volume changes only in charts like Chart 9b above.

Price indexes are important for economic and business statistics. For this reason we as a rule started our work for a client with creating a register-based survey that generated monthly price indexes for different commodity groups and markets. The information we needed for this were stored in the client's computer system. By matching the customer register and the article register with the register with all sales transactions these price indexes were computed every month. This price information was considered very important by the clients.

4.4 Why did the change happen?

As young students we were both fascinated by courses that were designed to teach us statistical literacy. Spurious correlations and confounding factors were explained with mathematically simple cross-tabulations. However, the examples were all based on cross section data and mainly medical cases.

It is important to show the students that these principles are also valid for their future jobs. In Chart 10a we have an example of spurious correlation – product group 1 seems to be more profitable than group 2. When we control for market in Chart 10b we see that within each market, product group 2 has a higher profit margin.

Chart 10a. Profit margins

Gross profit margin 2001		
	Product group 1	Product group 2
Total	**18.1%**	**16.7%**

Chart 10b. Profit margins by market

Gross profit margin 2001		
	Product group 1	Product group 2
Home market	19.4%	24.0%
Export	13.3%	15.4%
Total	**18.1%**	**16.7%**

These important statistical principles are also valid for time-series data, but we think that this fact is not mentioned during courses in statistics. In Chart 11a below both levels and trends are spurious due to a forgotten confounding variable – market. The spurious trend for product group 1 in Chart 11a is going up, but when we control for market in Chart 11b the two trends for product group 1 are going down.

Chart 11a. Profit margins by product group

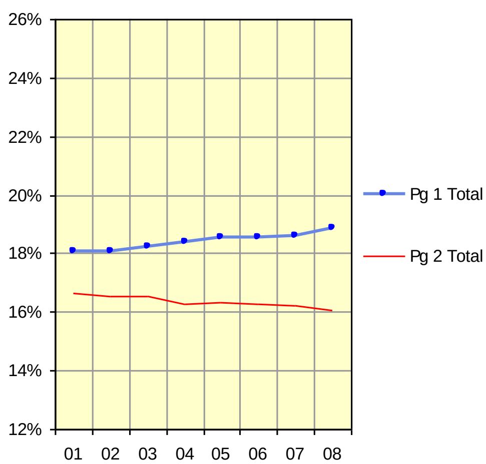
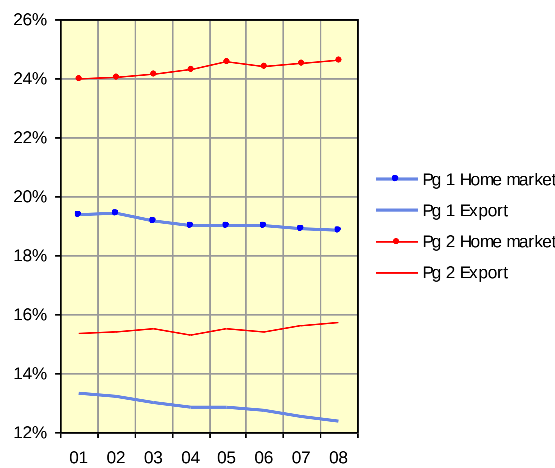


Chart 11b. Profit margins by product group and market



5. The role of time-series analysis at NSI:s

At National Statistical Institutes thousands of time-series are published every month and quarter. Time-series analysis is as a rule synonymous with seasonal adjustment and nothing more at NSI:s. This leads to that NSI:s publish what they think are correct and high quality data, but users misunderstand and draw wrong conclusions. We give here two examples with data from the Swedish Labor Force survey. In Chart 12 below, the number respective percent of employed persons of both sexes between 45-54 years is shown. It is a needle chart with estimated trend and seasonally adjusted values.

In Chart 12a there is a strong upward trend between 1987 and 1995. A conclusion can then be that this age group was not affected at all by the Swedish economic crisis in the beginning of the nineties. However, this conclusion is wrong – in Chart 12b we can see that the employment rate went down from 92% to 84% during the nineties. This means that this age group was clearly affected by the crisis. If a NSI only shows Chart 12a, most users will interpret the chart in a wrong way as they don't know that the number of persons belonging to this age group increased during the period. The employment trend in Chart 12a is confounded with the strong upward trend of the size of the age group.

Chart 12a. Employed in Sweden, both sexes 45-54 years, 1987-2009

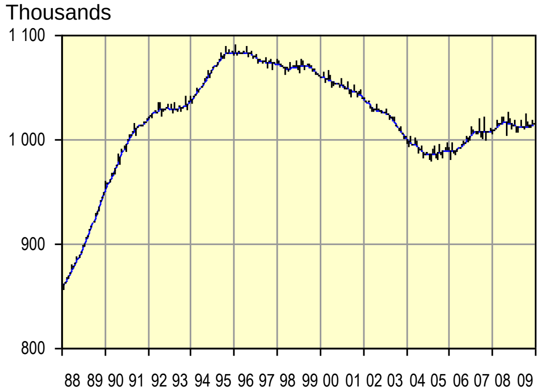
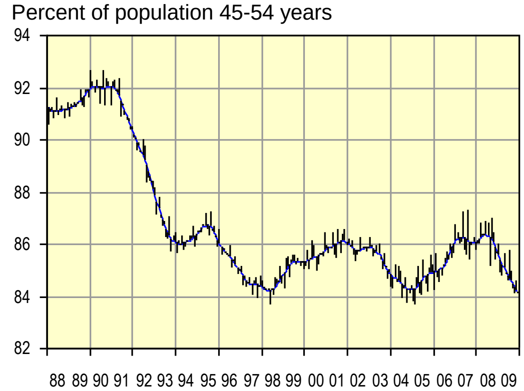


Chart 12b. Employed in Sweden, both sexes 45-54 years, 1987-2009



In chart 13, the proportion of the age group between 15-74 years that are active at the labor market is illustrated with needle charts. Chart 13a is what is published today, but the negative trend pattern in that chart is disturbed by the fact that the age-structure of the population 15-74 years changed during the period 2001-2010. In chart 13b we have controlled for age – if the population 15-74 years had the same age structure during all years 2001- 2010 then the trend would have been positive.

Chart 13a. Labor force ratios, both sexes 15-74 years, percent

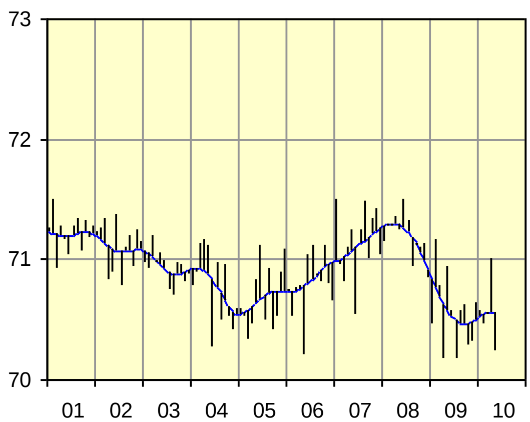
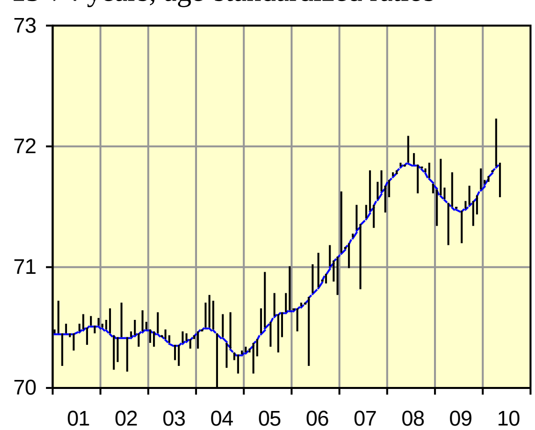


Chart 13b. Labor force ratios, both sexes 15-74 years, age-standardized ratios



6. Quality issues for time-series data

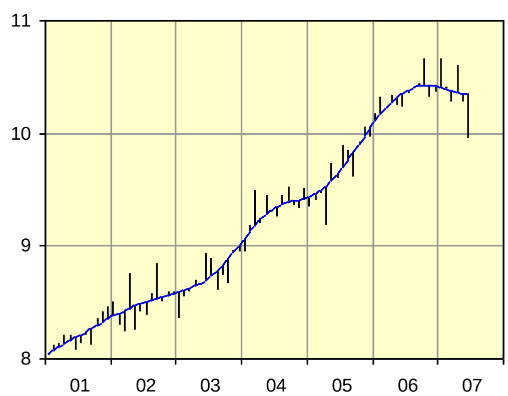
6.1 Random variation and statistical literacy

Understanding random variation is an important issue of statistical science. We have seen many cases where the random component in time-series has not been understood. At an enterprise this can lead to that short-term changes are misinterpreted – a normal change downward is interpreted as "Disaster" or a normal change upwards is interpreted as "The crisis is over". Changes in sales or profit are often related with strong passions "My strategy is a success/failure!" These passions make sometimes sound judgement difficult. In such cases statistical literacy can help. In time-series analysis we decompose an observed series into:

Observed value = Trend component + Seasonal component + Random component

Monthly and quarterly series should be decomposed according to this model and information about the standard error of the random component gives a good basis for the judgement of individual values and changes between values.

Chart 14a. Sales, $ Millions per month



The needle chart shows the trend component and the random component. The idea is that the needles should give an illustration of the quality of the time-series. Chart 14a shows that the last observation during 2007 can be interpreted as the sales last month were much lower than expected.

This judgement based on the chart should be combined with an analysis based on numerical values as the table in Chart 14b. The random component should be transformed into standardized residuals.

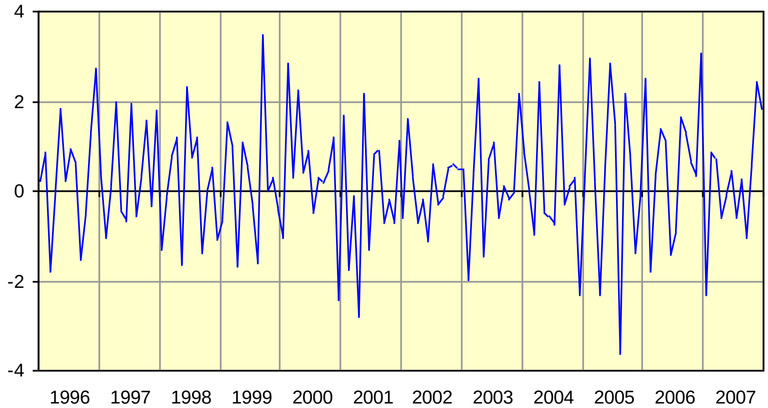
Chart 14b. Sales, $ Millions per month

		Observed values	Trend component	Seasonal component	Random component	Standardized residuals
2007	Jan	10.741	10.404	0.078	0.259	1.870
2007	Feb	10.293	10.390	-0.109	0.012	0.098
2007	Mar	11.642	10.375	1.373	-0.105	-0.747
2007	Apr	11.650	10.361	1.056	0.232	1.681
2007	May	11.164	10.351	0.895	-0.082	-0.579
2007	Jun	10.215	10.346	0.269	-0.400	-2.865

Analyzing and understanding the random component in time-series data is an issue that connects statistical literacy regarding time-series to probability theory and statistical inference.

Our next example illustrates the consequences when people ignore this aspect. Every month an index describing the volume of industrial production is produced and seasonally adjusted values are published. It is customary that the rate of change is described as the change between this month's seasonally adjusted index value and previous month's seasonally adjusted value. Chart 15 shows these monthly changes in the Swedish seasonally adjusted index of industrial production. This is a picture of almost pure random variation – there is no visible information on the business cycle.

Chart 15. Percent changes in Swedish industrial production volume
Change from previous month, percent per month, seasonally adjusted values



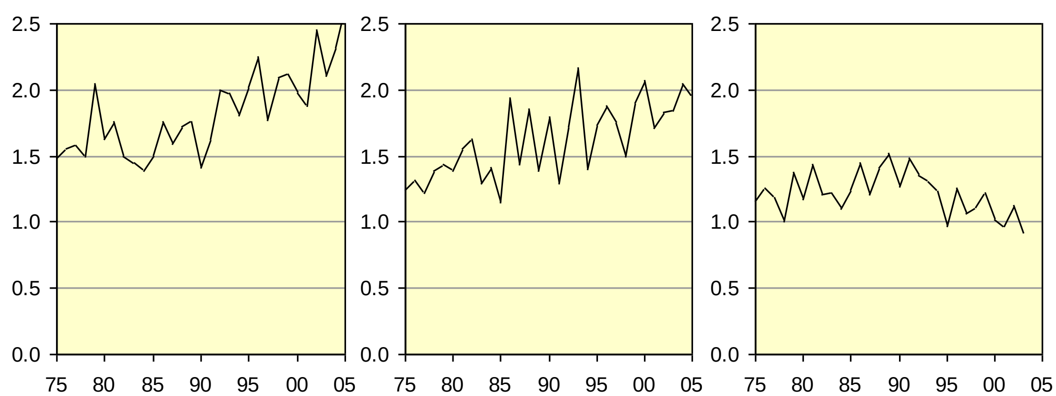
6.2 Evaluating sources

However, quality of data is not only a matter of random variation. It is also important to stress the importance of evaluating the quality of different sources. If you want to study the development of e.g. murder there are three different sources that are presented in the Statistical Yearbook of Sweden that can be used. These three sources are presented in Chart 16 below and the trends differ – in two sources the trend goes up but in one source the trend goes down.

To evaluate these sources the students should study how data has been generated in each source, if something has been written on the quality of theses sources by Statistics Sweden or by researchers in the field and then finally make their own judgement. In this case the source in Chart 16c is the best, this is discussed in Wallgren (2007, p 173).

Chart 16. Murder/manslaughter in Sweden **–** Number of persons per 100 000 inhabitants

16a. Reported to the police **16b.** Persons found guilty **16c.** Persons killed



7. Conclusions

Summing up our discussion leads to the following conclusions:

- Times-series data are more important than cross section data for managers and many others.
- Teaching of time-series analysis is often mathematical only, statistical literacy aspects of time-series data is often overlooked.
- Statistical description of time-series data is not trivial – if the trend goes up or down is an important and delicate issue.
- Statistical graphics is essential – to describe qualitative time-series patterns you need very good charts.
- Spurious correlations and confounding factors are equally important for time-series data as for cross section data.
- The random component is visible in time-series data. It must be analyzed and understood.
- Examples and cases should be relevant for the student's area of interest and their future jobs.

References

Wallgren, A., Wallgren, B., Persson, R., Jorner, U., Haaland, J. (1996): *Graphing Statistics & Data – Creating Better Charts.* Sage Publications 1996.

Wallgren, A. and Wallgren, B. (2007): *Register-based Statistics – Administrative Data for Statistical Purposes.* Wiley & Sons Ltd 2007.