AMATYC
38th Annual Conference

Jacksonville, FL

Nov. 8, 2012

Brian Smith, McGill University

brian.smith@mcgill.ca

Time Series Analysis

All You Need is Arithmetic

Introduce Time Series Early!

The analysis of time series data is usually not introduced at an early stage of the mathematics or statistics curriculum .

Modern introductory statistics courses tend to rush to probabilistic applications involving risk and confidence.

Why Teach Time Series in the Mathematics Curriculum?

1. Teaching time series requires only basic arithmetic skills – e.g. summing, averaging, ratios, …

2. Time series decomposition is one of the most useful applications of statistics by industry and government.

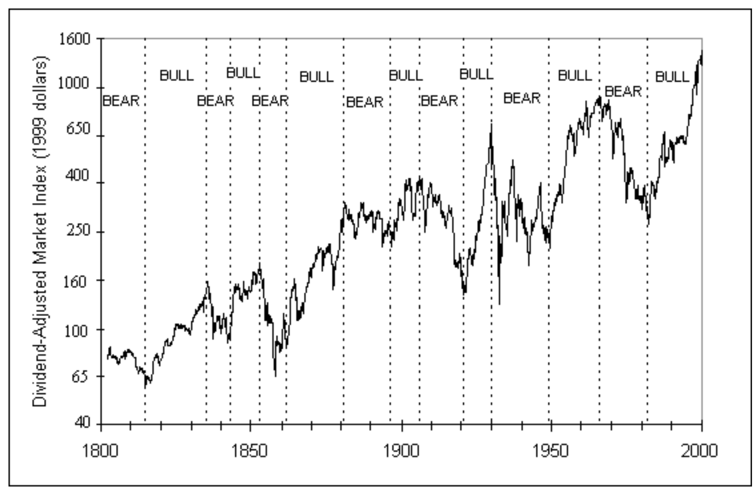
3. Time Series analysis quickly leads to the practical applications of (1) forecasting and (2) seasonal adjustment of data.

What is a Time Series?

A **time series** is a sequence of numbers collected at regular intervals over a period of time. In general we will assume equally spaced time intervals, such as yearly, monthly, weekly, or daily.

Examples of Time Series

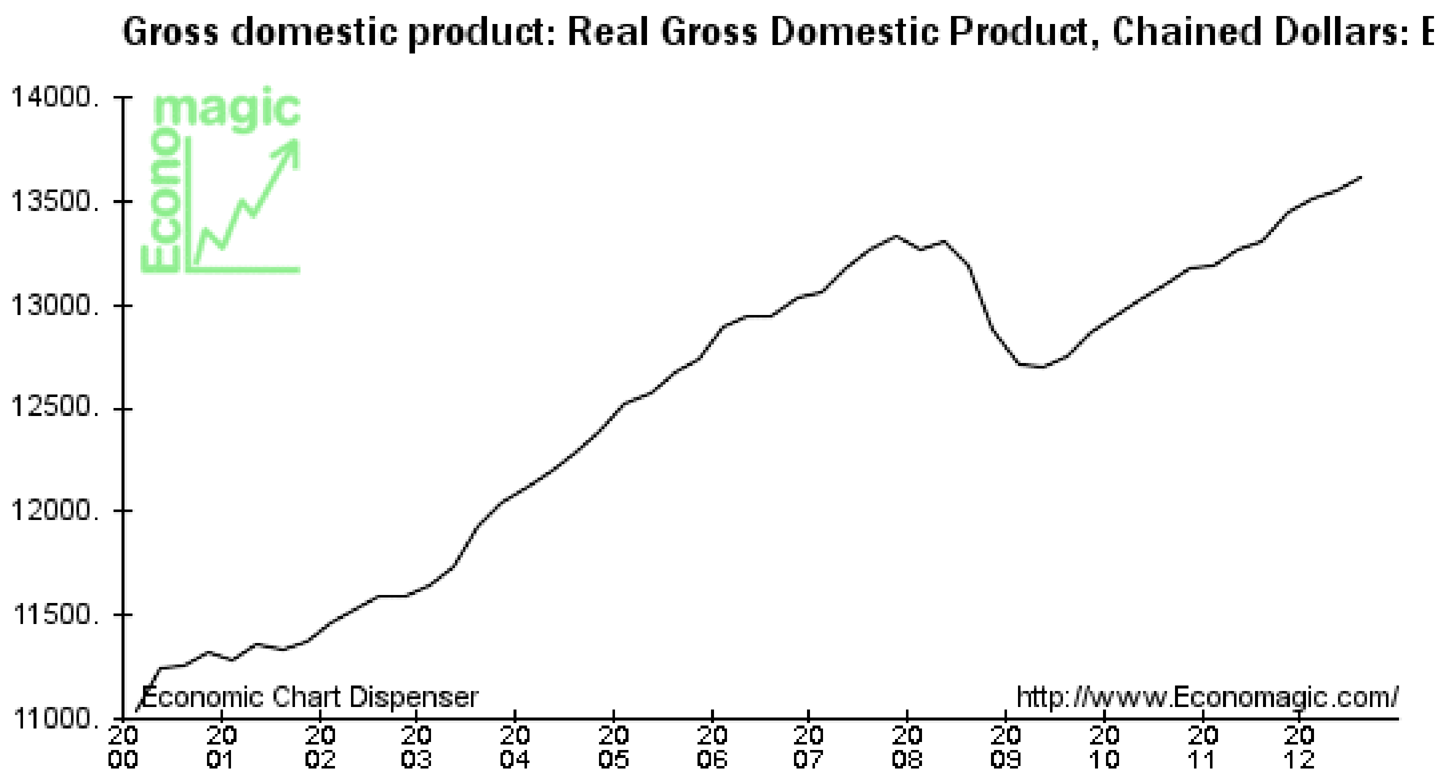
Examples of time series are the daily closing value of a stock price, the monthly sales of a computer, or the annual increase in population growth.



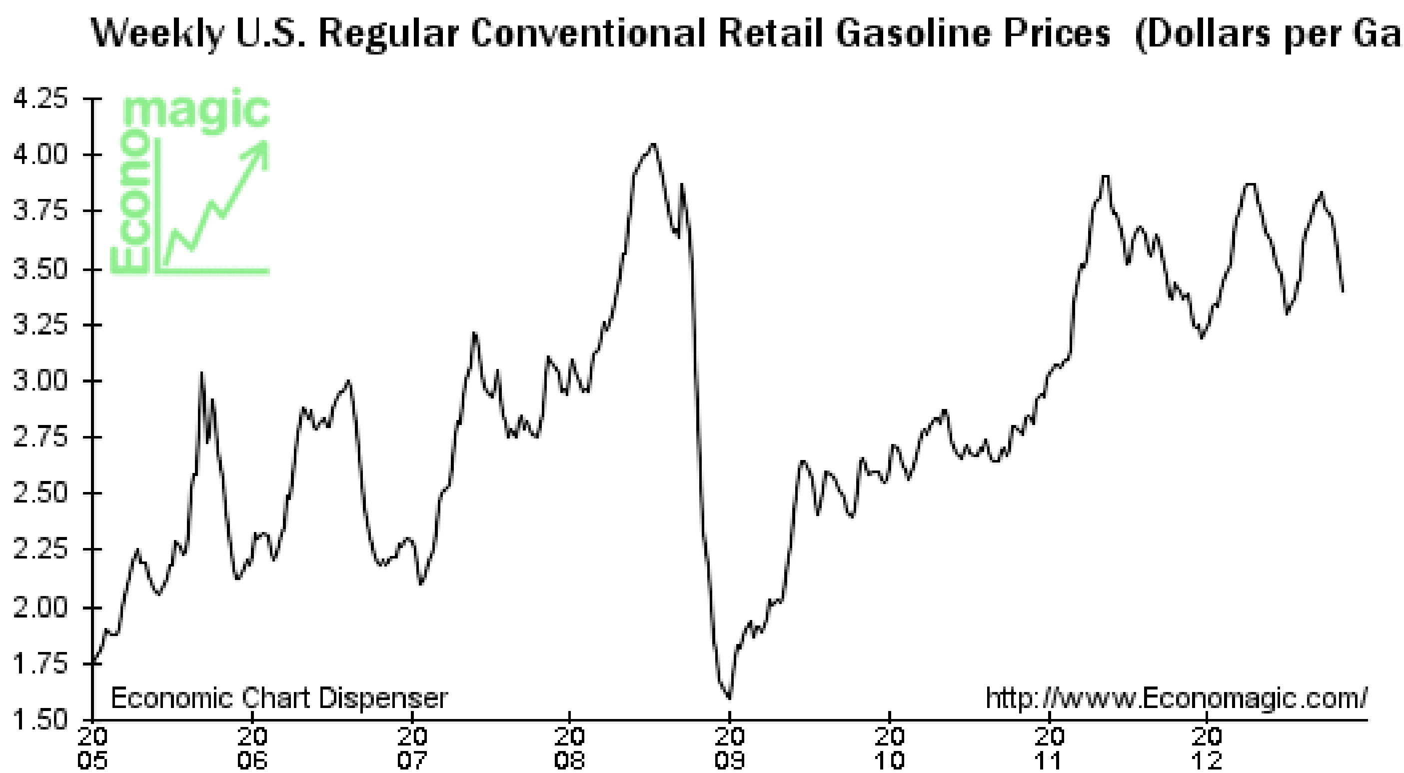
Apple Inc: 1 year Chart

Economagic – a site for economic time series data

Ten-Year GDP chart for U.S.



Weekly Gasoline prices 2005 - 2012



Time Series Analysis

Time Series Analysis consists of a variety of methodologies for identifying patterns in the series. We will present some of these procedures, stressing both the simplicity of the mathematics and the statistical importance of the results.

Description of the data

http://www.census.gov/epcd/ec97/def/44814.HTM

NAICS 44814: Family Clothing Stores

This industry comprises establishments primarily engaged in retailing a general line of new clothing for men, women, and children, without specializing in sales for an individual gender or age group. These establishments may provide basic alterations, such as hemming, taking in or letting out seams, or lengthening or shortening sleeves.

Data Set

The data set we will use to demonstrate the time series analysis consists of ten years of US retail sales for family clothing stores. This data set is identified by the US Bureau of the Census as NAICS 44814. The following extract is taken from the census bureau website.

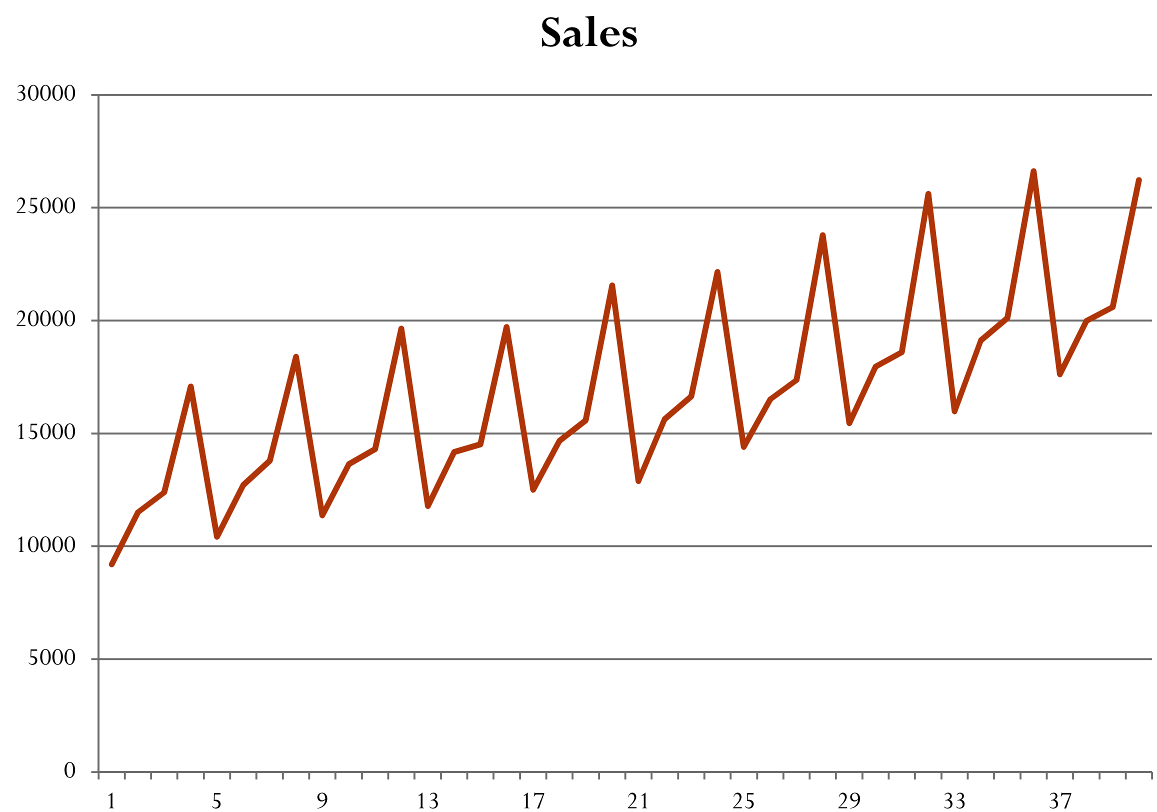
Ten Years of Data

We will use the ten years of data from 1998 to 2007. The actual data set consists of monthly sales figures, but for simplicity we have aggregated the data into quarterly values. The data are shown below in Excel format. Values are shown in millions, so that the entry of 9202 for the first quarter of 1992 represents $9,202,000,000. We see that quarterly sales for US family clothing stores are in billions of dollars.

Excel data file

	A	B	C	D	E	F	G
1	Year	Quarter	Sales ($Millions)		Year	Quarter	Sales ($Millions)
2	1998	1	9202		2003	1	12890
3		2	11505			2	15637
4		3	12387			3	16643
5		4	17075			4	22152
6	1999	1	10422		2004	1	14396
7		2	12714			2	16516
8		3	13795			3	17366
9		4	18402			4	23790
10	2000	1	11361		2005	1	15452
11		2	13632			2	17965
12		3	14297			3	18589
13		4	19638			4	25612
14	2001	1	11771		2006	1	15970
15		2	14174			2	19126
16		3	14511			3	20107
17		4	19709			4	26617
18	2002	1	12505		2007	1	17624
19		2	14666			2	19987
20		3	15578			3	20597
21		4	21556			4	26221

Time Series Plot



Identifying Components of the Series

Visual inspection identifies two predominant patterns

- an upward trend (sales appear to be increasing over time)
- sharp upward and downward movements in the series - these are seasonal variations

Analyzing the seasonal pattern

There is a large peak in every fourth quarter, indicating that sales are highest at the end of the year – i.e. in months October, November, December.)

This high point is followed by a sharp drop which shows that sales in the first quarter (January, February, March) fall off dramatically.

Components of the Model

T = Trend – the long term tendency of the data to increase or decrease over time

C = Cyclical - possible business cycles of several years duration

S = Seasonal – variations of less than a year's duration, typically quarterly or monthly

I = Irregular or random events, such as strikes, hurricanes, earthquakes, etc.

The Family Clothing Store Data

Our family clothing store sales data shows strong trend and seasonal factors with minimal cyclical or irregular components; therefore initially we will concentrate on identifying trend and seasonal patterns.

The Multiplicative Decomposition Model

We assume that a multiplicative model is appropriate, i.e. the time series values Y can be written in the form

$$Y = T \times C \times S \times I$$

Measuring Trend

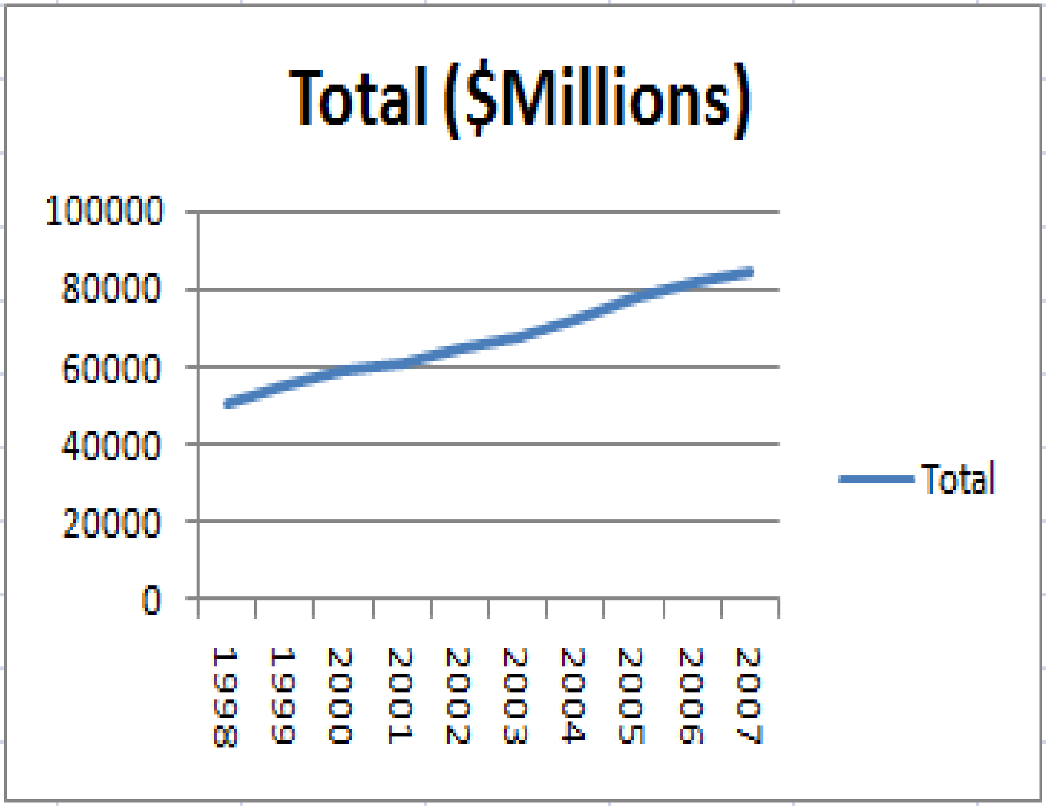
Our first operation is to calculate the annual totals. Clearly, since the total incorporates the values from all four quarters there is no way to distinguish between the seasons. Hence the annual total will not show the seasonal variation.

The first total is given by

$$T_1 = 9202 + 11505 + 12387 + 17075 = 50169$$

Graph of Annual Totals

Year	Total
1998	50169
1999	55333
2000	58928
2001	60165
2002	64305
2003	67322
2004	72068
2005	77618
2006	81820
2007	84429

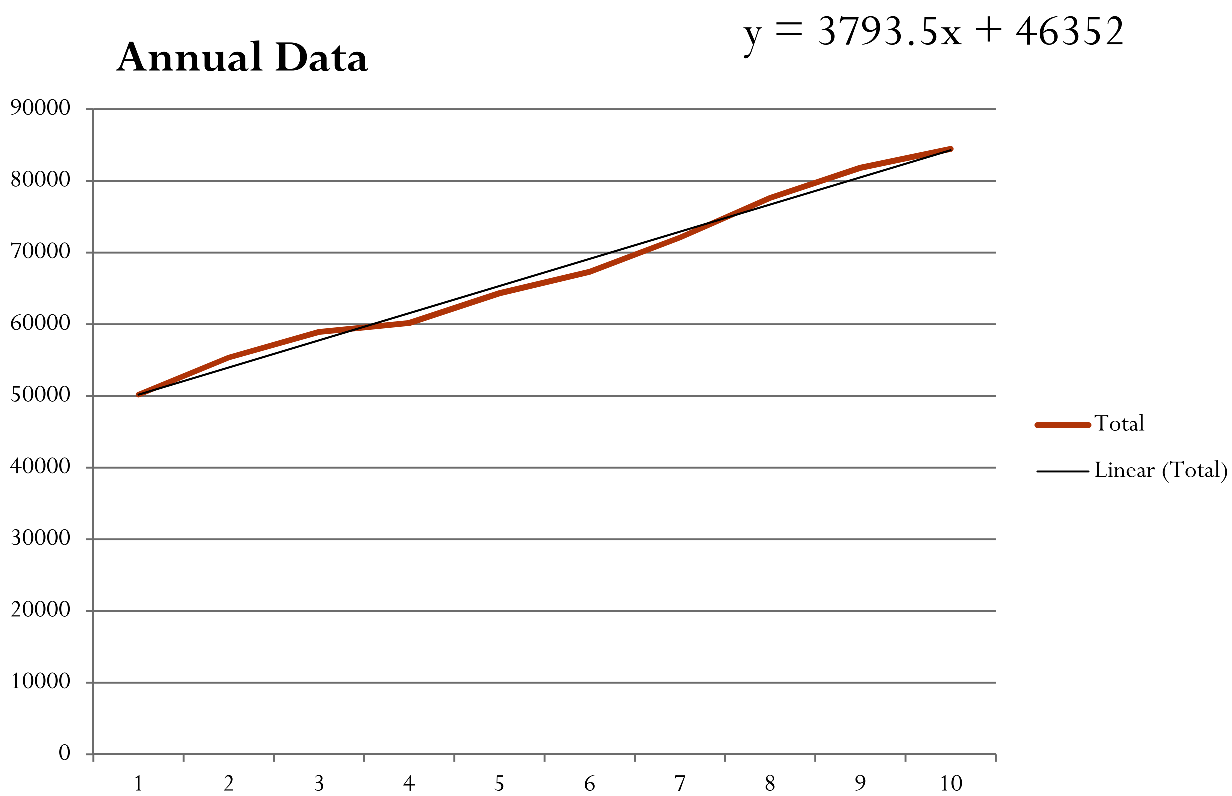


Linear Trend

Note that the graph now shows an upward trend that is almost linear. If the same pattern continues for future years we can use the slope of the straight line to predict future trend values.

We next illustrate methods to estimate the linear trend.

Straight line Tend for Annual Data



Using Excel we obtain the linear regression equation $y = 3793.5x + 46352$

Approximating the trend line using a linear equation

Year (1998 = 1)	Total
1	50169
2	55333
3	58928
4	60165
5	64305
6	67322
7	72068
8	77618
9	81820
10	84429

$$\text{Slope } m = \frac{y_2 - y_1}{x_2 - x_1} = \frac{84429 - 50169}{10 - 1} = 3806.7$$

$$y - y_1 = m(x - x_1)$$

$$y - 50169 = 3806.7(x - 1)$$

$$y = 3806.7x + 45323.3$$

Compare with regression equation $y = 3793.5x + 46352$

Trend Forecast for 2008

We use the trend line $y = 3793.5x + 46352$. Since we coded t= 1 for 1998 it follows that t = 11 for 2008.

Substituting t = 11 we get y = 3793.5(11) + 46352 =88080.5

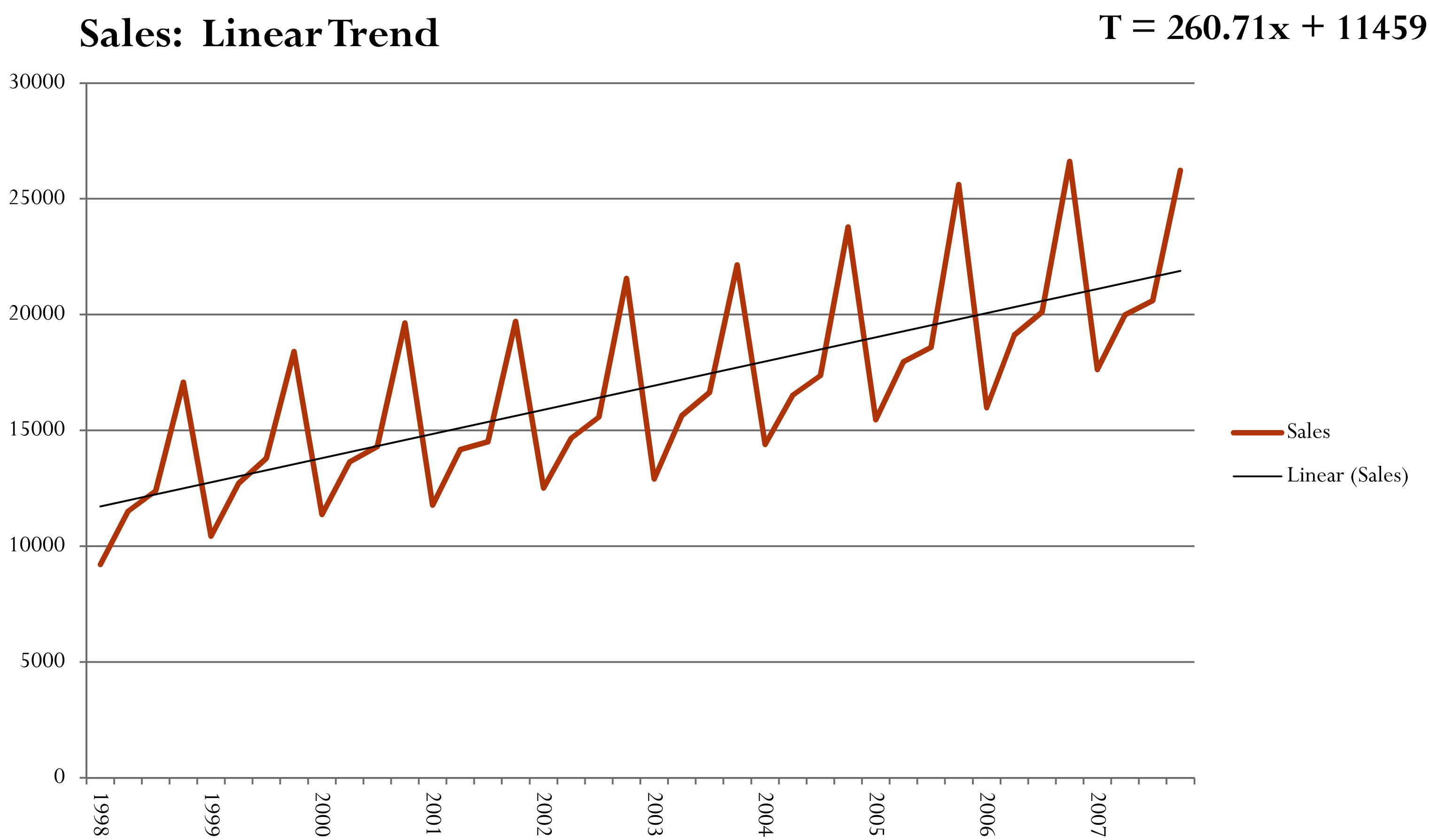
We forecast total sales for 2008 to be $88,080,500,000.

Linear Trend

If the same pattern continues for future years we can use the slope of the straight line to predict future trend values.

However, a forecast based on trend alone will predict annual totals but will not help us to forecast for each of the four seasons.

Using Excel to find the Linear Trend for quarterly data



Seasonal Indexes

Next, we need to find a way to measure the seasonal variation; we accomplish this by calculating seasonal indices (indexes) for each of the four quarters.

We proceed as follows. Consider the first year, 1998. The annual total for that year is 50169. If there was absolutely no seasonal difference then all four quarters would have the same value, namely, $50169 \div 4 = 12542.25$. The next Excel screen shows the original quarterly data in column C and the average values in column D:

Original quarterly data in column C and the annual average values in column D:

	A	B	C	D
1	Year	Quarter	Sales ($Millions)	Average
2	1998	1	9202	12542.25
3		2	11505	12542.25
4		3	12387	12542.25
5		4	17075	12542.25

Calculating Seasonal Effects

In the first quarter we compute the ratio $\frac{9202}{12542.25} \times 100 = 73.37$

This tells us that sales in the first quarter are 73.37% of the average for the quarter. Stated differently, we can say that sales in the first quarter are 26.63% below the average. Performing the same calculation for the remaining quarters of 1998 we get:

	A	B	C	D	E
1	Year	Quarter	Sales ($Millions)	Average	Sales/Average
2	1998	1	9202	12542.25	73.37%
3		2	11505	12542.25	91.73%
4		3	12387	12542.25	98.76%
5		4	17075	12542.25	136.14%

Seasonal Effects

We conclude that for 1998, sales in the first quarter are 73.37% of the average, while sales in quarters 2, 3, and 4, are, respectively 91.73%, 98.76%, and 136.14%.

Evidently, the first quarter is the “worst” quarter with sales at only 73.37% of the average, whereas the fourth quarter is the “best” with a value 136.14%, i.e. 36.14% above average.

We next repeat this process for each of the 10 years.

The following table shows the percentages by season and by year:

Year	Q1	Q2	Q3	Q4
1998	73.37	91.73	98.76	136.14
1999	75.34	91.91	99.72	133.03
2000	77.12	92.53	97.05	133.30
2001	78.26	94.23	96.47	131.03
2002	77.79	91.23	96.90	134.09
2003	76.59	92.91	98.89	131.62
2004	79.90	91.67	96.39	132.04
2005	79.63	92.58	95.80	131.99
2006	78.07	93.50	98.30	130.12
2007	83.50	94.69	97.58	124.23
Mean	**77.96**	**92.70**	**97.59**	**131.76**

Seasonal Pattern

Reading down each column we see remarkable consistency. For example the last column shows that in each of the ten years sales in the fourth quarter are considerably higher than the annual average.

Of course not all years are identical, so we see some variation – in particular the value for 2007 is 124.23 indicating that the seasonal increase that year was not as high as in the previous years when it was over 30%. Sales analysts may want to look back and try to pinpoint the economic factors that led to that result in 2007.

The Seasonal Indexes

The bottom row in the table shows the mean value for each quarter over the ten years. We can call this mean value a ***seasonal index*** for the corresponding quarter.

The seasonal index of 77.96 for the first quarter indicates that, *on the average*, sales in that quarter will be approximately 78% of the average, or 22% below the average.

Deseasonalization (seasonal adjustment) of data

The seasonal factor can be misleading when we want to understand the underlying trend in a time series.

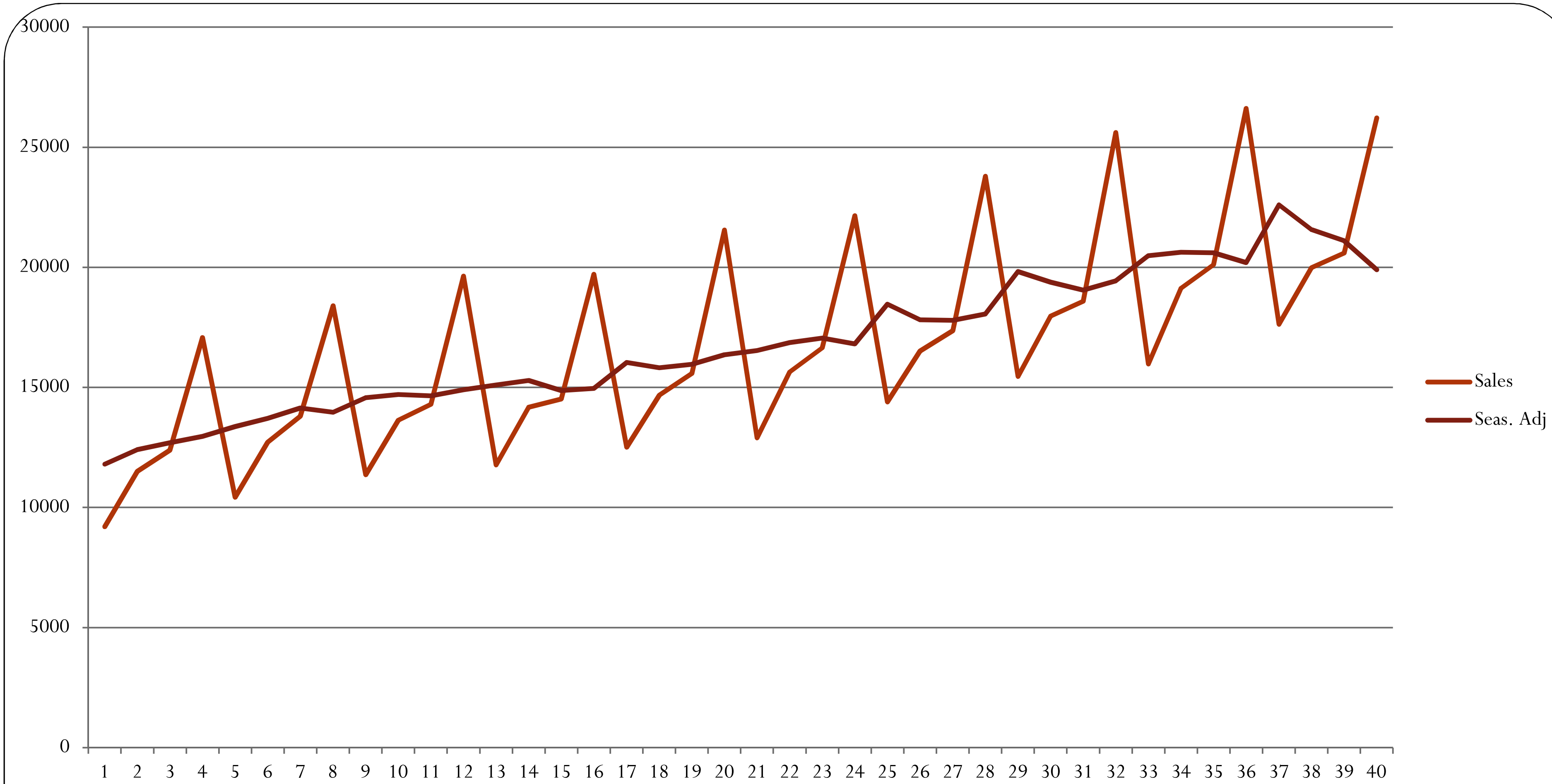
In the above example, the sales volume for the fourth quarter of 2003 is 22152. A newcomer in the business may be impressed that this is so much larger than for the three preceding quarters.

Interpreting a seasonally adjusted value

We deseasonalize the value by dividing out the seasonal index. Hence, the seasonally adjusted value for the fourth quarter of 2003 is actually 22152/1.3176 = 16812.51.

The interpretation of this value is that, were it not for the surge in sales expected as a result of the seasonal effect, the underlying value of sales is only $16,812,510,000, so that the reported value of $22,152,000,000 is not as impressive as it first appears.

We note that this is actually less than the deseasonalized value for the previous period, suggesting that when you remove the seasonal effect there is actually a decrease in the sales volume!



Seasonally Adjusted Data

The seasonally adjusted data is superimposed on the graph of the original data. Note how much smoother the adjusted data is.

The seasonal effect

	Y	S	Y/S = T.C.I
2003	12890	78.0	16534.1
	15637	92.7	16868.4
	16643	97.6	17054.0
	22152	131.8	16812.4

We observe that from the third to the fourth quarter of 2003 the original data increases from 16643 to 22152 but the deseasonalized value decreases from 17054 to 16812. This shows that the apparent gain is in fact solely due to the seasonal component and in fact the underlying series shows a decrease during that period.

Forecasting

A forecast based on trend alone will not be very useful if the seasonal effect is ignored. Therefore, we improve the forecast by first calculating the trend and then multiplying it by the seasonal index.

SOME BAD FORECASTS!

- *"Computers in the future may weigh no more than 1.5 tons."*

  Popular Mechanics, forecasting the relentless march of science, 1949

- *"I think there is a world market for maybe five computers."*

  Chairman of IBM, 1943

..even more bad forecasts

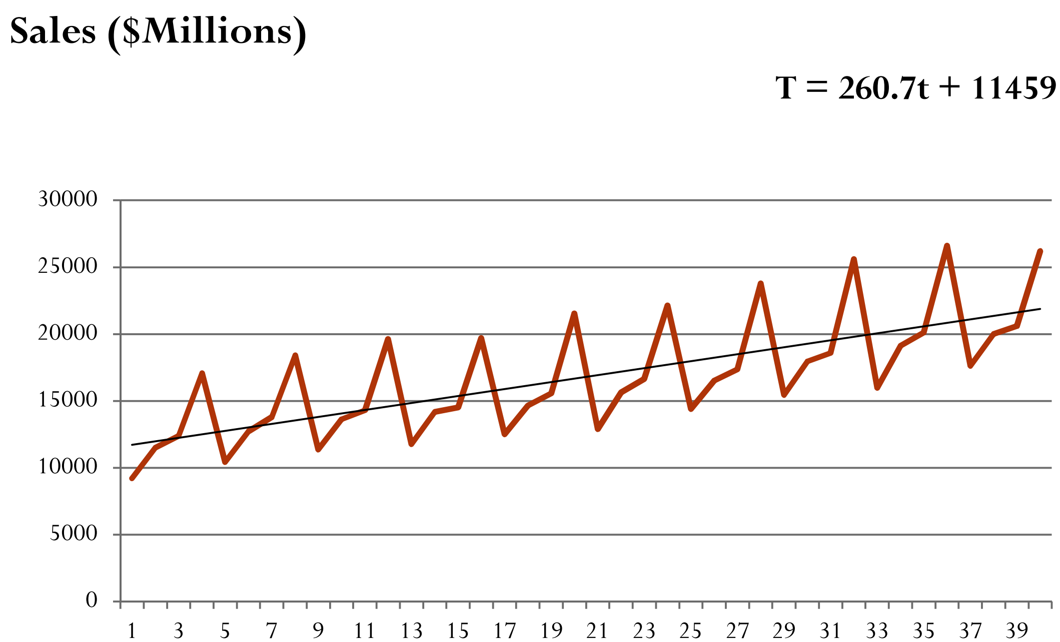
"*There is no reason anyone would want a computer in their home.*"

President, Chairman and founder of Digital Equipment Corp., 1977

"*Heavier-than-air flying machines are impossible.*"

President, Royal Society, 1895.

Time Series Plot with Linear Trend



The equation of the trend line is $T = 260.7t + 11459$, where $t = 1$ at the first quarter of 1998.

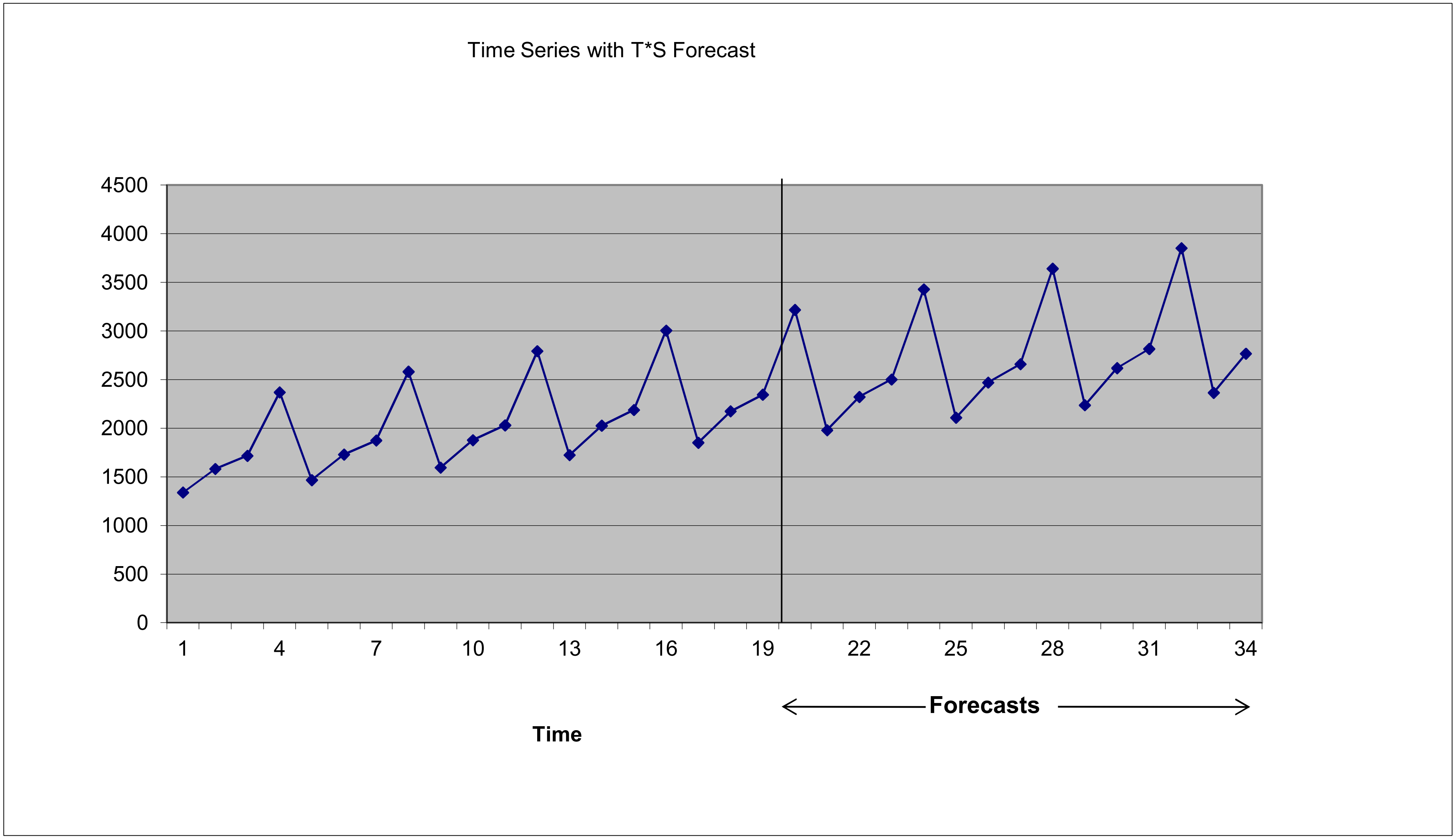
Trend x Seasonal Forecast

Since the data set consists of 40 values it follows that $t = 40$ for the fourth quarter of 2007. Hence, to forecast the trend one period ahead to the first quarter of 2008, we set $t = 41$ and compute $T = 260.7(41) + 11459 = 22147.7$.

This would be a serious overestimation of expected sales since for the first quarter the seasonal index implies that sales, on average, will be 77.96% of the trend. Therefore we adjust the forecast by multiplying the trend value by the seasonal index:

Hence, Forecast $= T \times S = 22147.7 \times 0.7796 = 17266.35$ or $17,266,350,000.

Graph of forecasts with trend and seasonal components



Long term forecasting

Forecasts for later time periods can be generated in the same fashion, though students should be warned of the dangers of extrapolation and the inevitable inaccuracies involved in long-term forecasting.

Conclusion

Introducing a module on time series analysis in an introductory mathematics or statistics course serves to reinforce basic arithmetic skills (addition, averaging, ratios) in the framework of a useful application that will familiarize students with an important statistical concept.

Seasonally adjusting data and forecasting future values are essential operations in business planning and in government reporting of population statistics. Our focus here is to show how these important statistical ideas can be introduced using only arithmetic computation that is accessible to students at the most introductory level.

Comments on Methodology

We note that the method used in this presentation for computing seasonal indices is not the standard ratio-to-moving average technique employed by professional forecasting models. The methods used here have been simplified to be readily available to students in an introductory mathematics course. While the methodology has been simplified the overall effect and interpretation remains unchanged.

For further reading see the bibliography on the next slide.

Bibliography

Makridakis, S., S.C . Wheelwright, and R.J. Hyndman (1998) *Forecasting: Methods and Applications*, 3rd ed., New York: John Wiley & Sons.

Hanke, J.E. and D.W. Wichern (2009) *Business Forecasting*, 9th ed., Upper Saddle River, N.J.: Prentice Hall.