QM353: Business Statistics

Chapter 8
Analyzing and Forecasting Time-Series Data

Chapter Goals

After completing this chapter, you should be able to:

- Identify the components present in a time series
- Develop and explain basic forecasting models
- Apply trend-based forecasting models, including linear trend, nonlinear trend, and seasonally adjusted trend
- Use smoothing-based forecasting models, including single and double exponential smoothing

Examples of Forecasting

- Governments forecast unemployment, interest rates, and expected tax revenues for policy purposes
- Marketing executives forecast demand, sales, and consumer preferences for strategic planning
- College administrators forecast enrollments to plan for facilities and for faculty recruitment
- Retail stores forecast demand to control inventory levels, hire employees and provide training

Categories of Forecasting

- Qualitative forecasting techniques
  - Based on statistical methods for analyzing historical data
- Qualitative forecasting techniques
  - Based on expert opinion and judgment
  - NOT a gut feel or an unsubstantiated opinions

Developing a Forecasting Model

- Steps in forecast modeling (see Chapter 15):
  - model specification
  - model fitting
  - model diagnosis
- Goal: use the simplest available model that meets forecasting needs to provide good forecasts for future performance

Forecasting Horizon

- Forecasting horizon is the lead time necessary (or available) to develop the forecasting model
- Intermediate term – less than one month
- Short term – one to three months
- Medium term – three months to two years
- Long term – two years or more

- Forecasting period: the unit of time for which forecasts are to be made
- Forecasting interval: the frequency with which new forecasts are prepared

Time-Series Analysis

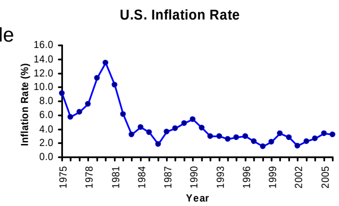
- The process for using past measurements to generate forecasts for the future
- Numerical data obtained at regular time intervals
- The time intervals can be annually, quarterly, daily, hourly, etc.
- Example:

Year:	2003	2004	2005	2006	2007
Sales:	75.3	74.2	78.5	79.7	80.2

Time Series Plot

A time-series plot is a two-dimensional plot of time series data

- the vertical axis measures the variable of interest
- the horizontal axis corresponds to the time periods



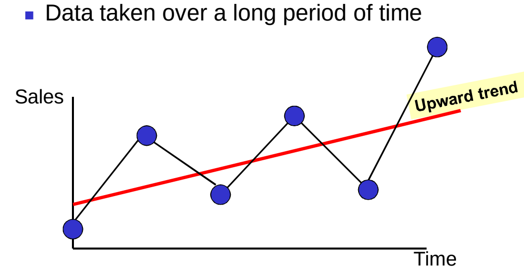
Time-Series Components

Time-Series

- Trend Component
- Seasonal Component
- Cyclical Component
- Random Component

Trend Component

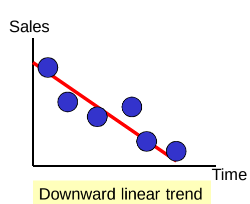
- Long-run increase or decrease over time (overall upward or downward movement)
- Data taken over a long period of time

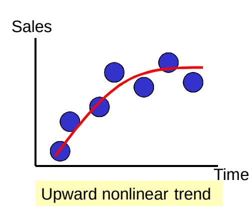


Trend Component

(continued)

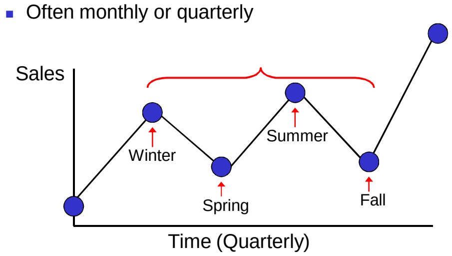
- Trend can be upward or downward (recall Chapters 13-14)
- Trend can be linear or non-linear (recall Chapter 15)





Seasonal Component

- Short-term regular wave-like patterns (repeating)
- Observed within 1 year
- Often monthly or quarterly



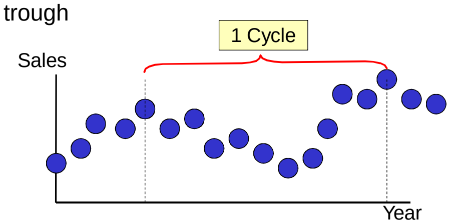
Seasonal Component

(continued)

- The pattern itself repeats throughout the time series
- The shortest period of repetition is the recurrence period
  - The recurrence period will be 1 year at MOST
- Examples:
  - Increase in visits to the doctor in the Fall and Winter, decrease in the Spring and Summer
  - Seasonal fluctuation in retails sales around various holidays (Christmas, Mother's Day, etc.)

Cyclical Component

- Long-term wave-like patterns
- Regularly occur but may vary in length
- Often measured peak to peak or trough to trough



Cyclical Component

(continued)

- Recurrence period is longer than 1 year
- Sustained periods of highs and lows
- Cycles vary in length and intensity
- Examples:
  - Unemployment rates
  - Stock market indexes
  - New home sales

Random Component

- Unpredictable, random, "residual" fluctuations
- Will be present in virtually all situations
- Due to random variations of
  - Nature
    - Devastating tornado hits a manufacturing facility
  - Accidents or unusual events
    - Unexpected closing of a large employer in a community
- "Noise" in the time series
  - No discernable pattern

Trend-Based Forecasting

- Estimate a trend line using regression analysis

Year	Time Period (t)	Sales (y)
2003	1	20
2004	2	40
2005	3	30
2006	4	50
2007	5	70
2008	6	65

- Use time (t) as the independent variable:

$$\hat{y} = b_0 + b_1 t$$

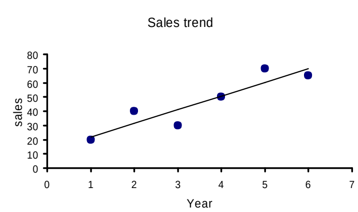
Trend-Based Forecasting

(continued)

Year	Time Period (t)	Sales (y)
2003	1	20
2004	2	40
2005	3	30
2006	4	50
2007	5	70
2008	6	65

- The linear trend model is:

$$\hat{y} = 12.333 + 9.5714\,t$$



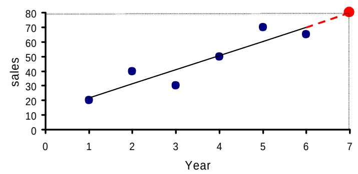
Trend-Based Forecasting

(continued)

- Forecast for time period 7:

Year	Time Period (t)	Sales (y)
2003	1	20
2004	2	40
2005	3	30
2006	4	50
2007	5	70
2008	6	65
2009	7	??

$$\hat{y} = 12.333 + 9.5714\,(7)$$
$$= 79.33$$



Comparing Forecast Values to Actual Data

- The forecast error or residual is the difference between the actual value in time t and the forecast value in time t:
- Error in time t:

$$e_t = y_t - F_t$$

Two Common Measures of Fit

Measures of fit are used to gauge how well the forecasts match the actual values (model diagnosis)

- MSE (mean squared error)
  - Average squared difference between y_t and F_t
- MAD (mean absolute deviation)
  - Average absolute value of difference between y_t and F_t
  - Less sensitive to extreme values
- RMSE (root mean square error)
  - Square root of MSE

MSE vs. MAD

Mean Square Error

$$MSE = \frac{\sum (y_t - F_t)^2}{n}$$

Mean Absolute Deviation

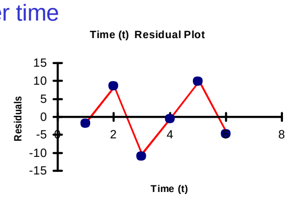
$$MAD = \frac{\sum | y_t - F_t |}{n}$$

where:

y_t = Actual value at time t
F_t = Predicted value at time t
n = Number of time periods

Autocorrelation

- Autocorrelation is correlation of the error terms (residuals) over time



- Here, residuals show a cyclic pattern, not random
- Also called serial correlation

- Violates the regression assumption that residuals are random and independent

True Forecasts

- True forecasts are gauged by how well it forecasts future values – not how well it fits historical data
- To determine if the trend model produced a true forecast, you have to wait until the future time actually arrives
  - Can use split samples
- Forecast bias
  - Positive – underforecast
  - Negative – overforecast

Nonlinear Trend Forecasting

- A nonlinear regression model can be used when the time series exhibits a nonlinear trend
- One form of a nonlinear model:

$$y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \varepsilon$$

- Compare R^2 and s_ε to that of linear model to see if this is an improvement
- Can try other functional forms to get best fit

Finding Seasonal Indexes

Ratio-to-moving average method:

- Begin by removing the seasonal and irregular components (S_t and I_t), leaving the trend and cyclical components (T_t and C_t)
- To do this, we need moving averages

Moving Average: averages of consecutive time series values

Multiplicative Time-Series Model

- Used primarily for forecasting
- Allows consideration of seasonal variation
- Observed value in time series is the product of components

$$y_t = T_t \times S_t \times C_t \times I_t$$

where
T_t = Trend value at time t
S_t = Seasonal value at time t
C_t = Cyclical value at time t
I_t = Irregular (random) value at time t

Moving Averages

- Used for smoothing
- Series of arithmetic means over time
- Result dependent upon length of period chosen for computing means
- To smooth out seasonal variation, the number of periods should be equal to the number of seasons
  - For quarterly data, number of periods = 4
  - For monthly data, number of periods = 12

Moving Averages

(continued)

- Example: Four-quarter moving average
  - First average:

$$\text{Moving average}_1 = \frac{Q1 + Q2 + Q3 + Q4}{4}$$

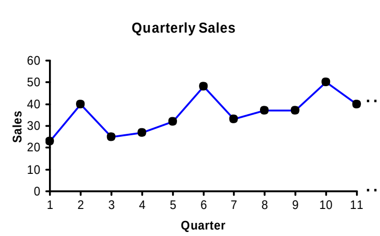
  - Second average:

$$\text{Moving average}_2 = \frac{Q2 + Q3 + Q4 + Q5}{4}$$

  - etc…

Seasonal Data

Quarter	Sales
1	23
2	40
3	25
4	27
5	32
6	48
7	33
8	37
9	37
10	50
11	40
etc…	etc…



Calculating Moving Averages

Quarter	Sales
1	23
2	40
3	25
4	27
5	32
6	48
7	33
8	37
9	37
10	50
11	40

etc…

Average Period	4-Quarter Moving Average
2.5	28.75
3.5	31.00
4.5	33.00
5.5	35.00
6.5	37.50
7.5	38.75
8.5	39.25
9.5	41.00

$$2.5 = \frac{1+2+3+4}{4}$$

$$28.75 = \frac{23+40+25+27}{4}$$

- Each moving average is for a consecutive block of 4 quarters

Centered Moving Averages

- Average periods of 2.5 or 3.5 don't match the original quarters, so we average two consecutive moving averages to get centered moving averages

Average Period	4-Quarter Moving Average
2.5	28.75
3.5	31.00
4.5	33.00
5.5	35.00
6.5	37.50
7.5	38.75
8.5	39.25
9.5	41.00

etc…

Centered Period	Centered Moving Average
3	29.88
4	32.00
5	34.00
6	36.25
7	38.13
8	39.00
9	40.13

Calculating the Ratio-to-Moving Average

- Now estimate the $S_t \times I_t$ value
- Divide the actual sales value by the centered moving average for that quarter
- Ratio-to-Moving Average formula:

$$S_t \times I_t = \frac{y_t}{T_t \times C_t}$$

Calculating Seasonal Indexes

Quarter	Sales	Centered Moving Average	Ratio-to-Moving Average
1	23		
2	40		
3	25	29.88	0.837
4	27	32.00	0.844
5	32	34.00	0.941
6	48	36.25	1.324
7	33	38.13	0.865
8	37	39.00	0.949
9	37	40.13	0.922
10	50	etc…	etc…
11	40	…	…
…	…	…	…

Example:

$$0.837 = \frac{25}{29.88}$$

Calculating Seasonal Indexes

(continued)

	Quarter	Sales	Centered Moving Average	Ratio-to-Moving Average
	1	23		
	2	40		
Fall	3	25	29.88	0.837
	4	27	32.00	0.844
	5	32	34.00	0.941
	6	48	36.25	1.324
Fall	7	33	38.13	0.865
	8	37	39.00	0.949
	9	37	40.13	0.922
	10	50	etc…	etc…
Fall	11	40	…	…
	…	…	…	…

Average all of the Fall values to get Fall's seasonal index

Do the same for the other three seasons to get the other seasonal indexes

Interpreting Seasonal Indexes

- Suppose we get these seasonal indexes:

Season	Seasonal Index
Spring	0.825
Summer	1.310
Fall	0.920
Winter	0.945

$\Sigma = 4.000$ -- four seasons, so must sum to 4

- Interpretation:

Spring sales average 82.5% of the annual average sales

Summer sales are 31.0% higher than the annual average sales

etc…

Deseasonalizing

- The data is deseasonalized by dividing the observed value by its seasonal index

$$T_t \times C_t \times I_t = \frac{y_t}{S_t}$$

- This smooths the data by removing seasonal variation

Deseasonalizing

(continued)

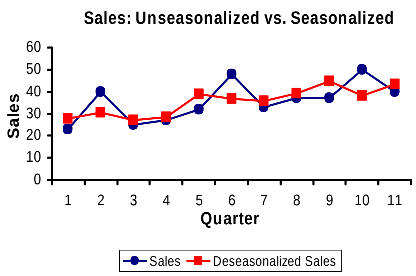
Quarter	Sales	Seasonal Index	Deseasonalized Sales
1	23	0.825	27.88
2	40	1.310	30.53
3	25	0.920	27.17
4	27	0.945	28.57
5	32	0.825	38.79
6	48	1.310	36.64
7	33	0.920	35.87
8	37	0.945	39.15
9	37	0.825	44.85
10	50	1.310	38.17
11	40	0.920	43.48
…		…	…

Example:

$$27.88 = \frac{23}{0.825}$$

etc…

Unseasonalized vs. Seasonalized



Seasonal Adjustment Summarized

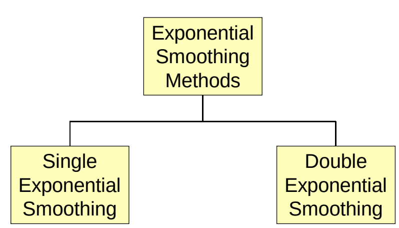
1. Compute each moving average
2. Compute the centered moving averages
3. Isolate the seasonal component by determining the ratio-to-moving average values
4. Determine seasonal indexes and normalize if necessary
5. Deseasonalize the time series
6. Develop trend line using deseasonalized data
7. Develop unadjusted forecasts using trend projection
8. Seasonally adjust the forecasts

Using Dummy Variables for Seasonality

- Can incorporate the seasonal component using dummy variables in a regression model
- Example for 4 seasons:
  - Let x_1 = 1 if winter, 0 if not winter
  - Let x_2 = 1 if spring, 0 if not spring
  - Let x_3 = 1 if summer, 0 if not summer
  - (Fall is the default season)
- Model:

$$F_t = \beta_0 + \beta_1 t + \beta_2 x_1 + \beta_3 x_2 + \beta_4 x_3 + \varepsilon$$

Forecasting Using Smoothing Methods



Used when there is no pronounced trend in the data

The goal is to "smooth out" the irregular component

Exponential Smoothing

- Assumes most recent data is more indicative of possible future values
- Current observations can be weighted more heavily than older observations
- The forecast developed reflect the current data more
- Good for short term forecasting and for time series that are not seasonal

Single Exponential Smoothing

- A weighted moving average
  - Weights decline exponentially
  - Most recent observation weighted most
- Used for smoothing and short term forecasting
- Easy to update

Single Exponential Smoothing

(continued)

- The weighting factor is ~
  - Subjectively chosen
  - Range from 0 to 1
  - Smaller ~ gives more smoothing, larger ~ gives less smoothing
- The weight is:
  - Close to 0 for smoothing out unwanted cyclical and irregular components
  - Close to 1 for forecasting

Exponential Smoothing Model

Single exponential smoothing model

$$F_{t+1} = F_t + \alpha(y_t - F_t)$$

or:

$$F_{t+1} = \alpha y_t + (1-\alpha)F_t$$

where:

F_{t+1} = forecast value for period t + 1
y_t = actual value for period t
F_t = forecast value for period t
α = alpha (smoothing constant)

Exponential Smoothing Example

Suppose we use weight ~ = 0.2

Quarter (t)	Sales (y_t)	Forecast from prior period	Forecast for next period (F_{t+1})
1	23	NA	23
2	40	23	(0.2)(40)+(0.8)(23)=26.4
3	25	26.4	(0.2)(25)+(0.8)(26.4)=26.12
4	27	26.12	(0.2)(27)+(0.8)(26.12)=26.296
5	32	26.296	(0.2)(32)+(0.8)(26.296)=27.437
6	48	27.437	(0.2)(48)+(0.8)(27.437)=31.549
7	33	31.549	(0.2)(33)+(0.8)(31.549)=31.840
8	37	31.840	(0.2)(37)+(0.8)(31.840)=32.872
9	37	32.872	(0.2)(37)+(0.8)(32.872)=33.697
10	50	33.697	(0.2)(50)+(0.8)(33.697)=36.958
etc…	etc…	etc…	etc…

$F_1 = y_1$ since no prior information exists

$F_{t+1} = \alpha y_t + (1-\alpha)F_t$

Sales vs. Smoothed Sales

- Seasonal fluctuations have been smoothed
- NOTE: the smoothed value in this case is generally a little low, since the trend is upward sloping and the weighting factor is only 0.2

Exponential Smoothing in Excel

- Use: Data / data analysis / exponential smoothing
  - The “damping factor” is (1 - ~)

Mean Absolute Percent Error

$$MAPE = \frac{\sum \frac{|y_t - F_t|}{y_t}(100)}{n}$$

where:

y_t = Value of time series in time t

F_t = Forecast values for time period t

n = Number of periods of available data

Chapter Summary

- Discussed the importance of forecasting
- Addressed component factors present in the time-series model
- Described least square trend fitting and forecasting
  - linear and nonlinear models
- Performed smoothing of data series
  - moving averages
  - single and double exponential smoothing