Forecasting

BUS 735: Business Decision Making and Research

1

1.1 Goals and Agenda

Goals and Agenda

Learning Objective	**Active Learning Activity**
Learn how to identify regularities in time series data	Lecture / Excel Example.
Learn popular univariate time series forecasting methods	Lecture / Excel Example.
Learn how to use regression analysis for forecasting	Lecture / Excel Example.
Practice what we learn.	In-class exercise.
More practice.	Read Chapter 15, Homework exercises.
Assess what we have learned	Quiz??

2 Time Series Analysis

2.1 Example Data

Working with Example Data

- Dataset: Total number of Mining, Logging, and Construction employees (in thousands) in the La Crosse area (obtained from Bureau of Labor Statistics website, http://www.bls.gov).
- To plot the data, we need to convert it to a single column:
  1. First generate observation numbers 1 through 116
  2. Figure out what row the observation is in: `=int((obs-1)/12)+1`
  3. Figure out what column the observation is in: `=mod((obs-1),12)+1`
  4. Pick out the right observation: `=offset([top_corner],row,col)`
- Create dates: 2000.0 through 2010.58.

Graphing Example Data

- In Excel: `Insert`, `Line`, `Line with markers`.
- Right click on data, select `Select Data`.
- Remove all the nonsense there.
- Select `Add`.
- Type "Employment" in `Series Name`. Select data for `Series Values`.
- Click `Edit` under `Horizontal Axis Values`.
- Select dates.

2.2 Time Series Characteristics

Time Series Characteristics

- **Trend:** gradual, long-term movement of the data in a positive or negative direction.
- **Cycle:** repetitive up-and-down movement of the data, each "cycle" need not be the same length or magnitude.
- **Seasonal pattern:** up-and-down movement of the data that can be predicted quite accurately by the time of the year.
- **Random variations:** movements in the data that are otherwise unpredictable.

2.3 Forecasting Time Series

Time Series Analysis

- **Time series analysis:** use of statistical methods to uncover trend, cyclical patterns, and seasonal patterns; and using this information to forecast future outcomes for a variable.
- **Univariate time series:** using many observations of only the variable of interest to forecast that variable.
- **Multivariate time series:** using one or more related variables to help forecast variable of interest.
  - New housing sales may also help predict construction employment.
  - National price measures for costs of building materials may also help predict construction employment.
- BUS 735: Focus on univariate time series analysis.

3 Time Series Methods

3.1 Smoothing Methods

Moving Average

- **Naïve forecast:** Forecast for tomorrow is what happened today.
  - Often used to measure usefulness of other time series forecasts.
- **Moving average:** uses several recent values to forecast the next period's outcome.

$$MA_{t,q} = \frac{1}{q} \sum_{i=1}^{q} x_{t-i}$$

  - x_t denotes the value of the variable at time t,
  - $MA_{t,q}$ denotes the Moving Average forecast for time t, using the most recent q periods.

Moving Average Properties

- Moving average lag length:
  - Longer lag lengths cause forecast to react more quickly/slowly to recent changes in the variable.
  - Longer lag lengths cause forecast to be more smooth/volatile.
- Performs (forecasting accuracy) best with data that has
  - No pronounced cyclical or seasonal variation.
  - No long-term trend.

Weighted Moving Average

- **Weighted moving average:** like a moving average, but larger weights are assigned to more recent observations.

$$WMA_{t,q} = \sum_{i=1}^{q} w_i x_{t-i}$$

  - w_i is the weight given to the observation that occurred i periods ago.
  - $\sum_{i=1}^{q} w_i = 1$
  - Typically, $w_i > w_{i+1}$.
- More recent observations are viewed as more informative.

Exponential Smoothing

- **Exponential smoothing:** Averaging method using all previous data, but puts weights larger weight on the more recent observations.

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

- F_t is the forecast for period t.
- x_{t-1} is the value of the variable in the previous time period, $t-1$.
- Useful for capturing information on recent seasonal or cyclical patterns (but with a lag).
- $\alpha \in [0,1]$ is the smoothing parameter.
  - When α is larger, more weight is given to most recent observations.

Adjusted Exponential Smoothing

- **Adjusted exponential smoothing:** exponential smoothing that is adjusted to incorporate information on a *long-term trend*.

$$AF_t = F_t + T_t$$

  - AF_t is the adjusted exponential smoothing forecast.
  - F_t is the regular exponential smoothing forecast.
  - T_t is the latest estimate of the trend.
- Trend is computed by,

$$T_t = \beta(F_t - F_{t-1}) + (1-\beta)T_{t-1}$$

  - $\beta \in [0,1]$ is a trend weighting parameter.
- Trend formula allows for changing trend throughout the data.

3.2 Regression Methods

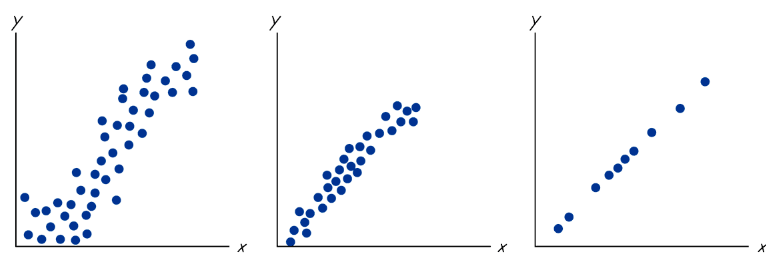
Regression

- Regression line: equation of the line that describes the linear relationship between variable x and variable y.
- Need to assume that one variable causes another.
  - x: *independent* or *explanatory* variable.
  - y: *dependent* or *outcome* variable.
  - Variable x can influence the value for variable y, but not vice versa.

Regression Model Examples

- How does housing demand affects construction employment?
  - x_i: housing demand (independent variable, aka explanatory variable).
  - y_i: construction employment (dependent variable, aka outcome variable).
- Seasonal adjustment: how does winter season affect construction employment?
  - Dummy variable: $x_i = 1$ if winter, $x_i = 0$ otherwise).
  - y_i: construction employment.
- Be careful!
  - Construction demand
  - Construction worker pay

Positive linear correlation



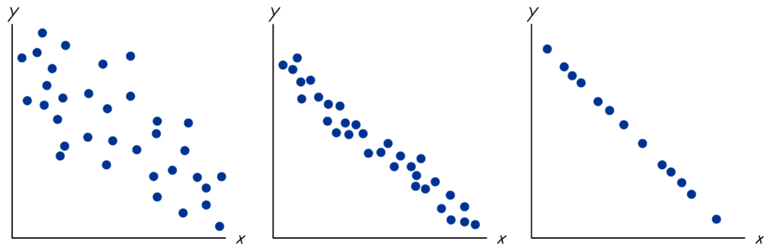
(a) Positive correlation between x and y

(b) Strong positive correlation between x and y

(c) Perfect positive correlation between x and y

- Regression analysis will find equation for best fitting line for positively related variables.
- Stronger correlation, better forecast accuracy.
- Perfectly positively correlated??

Negative linear correlation



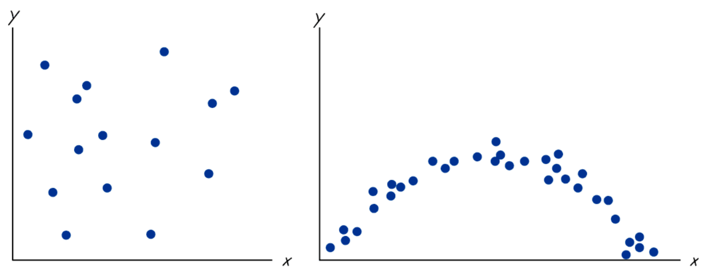
(d) Negative correlation between x and y

(e) Strong negative correlation between x and y

(f) Perfect negative correlation between x and y

- Regression analysis will find equation for best fitting line for negatively related variables.
- Stronger correlation, better forecast accuracy.
- Perfectly negatively correlated??

No linear correlation



(g) No correlation between x and y

(h) Nonlinear relationship between x and y

- Panel (g): no relationship at all.
- Panel (h): strong relationship, but not a *linear* relationship.
  - Need to transform your x variable before proceeding.

Regression line

- Population regression line:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

  - The actual coefficients β_0 and β_1 describing the relationship between x and y are unknown.
  - ϵ_i: error term, since linear relationship between the x variables and y are not perfect.

- Use sample data to come up with an estimate of the regression line:

$$y_i = b_0 + b_1 x_i + e_i$$

  - e_i: residual = the difference between the predicted value $\hat{y}$ and the actual value y_i.

Interpreting the slope

- Interpreting the slope, b_1: amount the y is predicted to increase when increasing x by one unit.
- When $b_1 < 0$ there is a negative linear relationship. That is increasing x causes y to decrease.
- When $b_1 > 0$ there is a positive linear relationship. That is increasing x causes y to increase.
- When $b_1 = 0$ there is no linear relationship between x and y.

Multiple Regression

- Multiple regression line (population):

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_2 + ... + \beta_{k-1} x_{k-1} + \epsilon_i$$

- Multiple regression line (sample):

$$y_i = b_0 + b_1 x_{1,i} + b_2 x_2 + ... + b_k x_k + e_i$$

  - $k + 1$: number of parameters (coefficients) you are estimating.
- Example, could use multiple variables to forecast construction employment:
  - New housing sales, measure of costs of construction materials, population growth.
  - Dummy for winter, Dummy for spring, Dummy for summer, time period (captures trend).

Multicolinearity

- **Multicolinearity:** when two or more explanatory variables are closely related to one another.
- This makes distinguishing the causal effect of each variable difficult.
  - Example: using unemployment *and* consumer spending as explanatory variables for construction employment.

- When there is multicolinearity, one or more explanatory variables can accurately predict another explanatory variables.
  - Any one of these explanatory variables has little explanatory power over and above the others.
- Perfect multicolinearity: when one or more explanatory variables perfectly predicts another explanatory variables.
  - Regression analysis blows up.

3.3 Seasonal Adjustment

Seasonal Adjustment

- Previous methods capture information in recent movements, but not past seasonal fluctuations.
  - Example, we may realize that construction employment is always lowest in Jan, Feb, and March.
- **Seasonal factor:** percentage of a total year that occurs in a specific season:

$$S_k = \frac{D_k}{\sum D_k}$$

  - D_k is the sum of all values occurring in season k, for all years considered.
- Use your favorite forecasting method, forecast *years* only.
- For each season, multiply annual forecast by the seasonal factor.

Regression Seasonal Adjustment

- Run a regression.
  - Use trend as an explanatory variable.
  - Use seasonal dummies as explanatory variables.
- Note: avoid multicolinearity, choose 1 fewer dummy variables than total seasons.
- Coefficient on seasonal dummies: impact of the season over and above excluded season dummy.

4 Forecast Accuracy

4.1 Absolute Deviations

Forecast Accuracy

- Useful to compare forecasts from multiple techniques.
- **Mean absolute deviation (MAD):** average distance between the forecast value and the actual value.

$$MAD = \frac{1}{T} \sum_{t=1}^{T} \|x_t - F_t\|$$

- **Mean absolute percentage deviation:** measures the distance between the forecast and actual values as a percentage of the total values.

$$MAPD = \frac{\sum_{t=1}^{T} \|x_t - F_t\|}{\sum_{t=1}^{T} x_t}$$

4.2 Squared Deviations and Bias

Forecast Accuracy and Bias

- **Mean Squared Error (MSE):** instead of taking absolute value of differences, square them:

$$MSE = \frac{1}{T} \sum_{t=1}^{T} (x_t - F_t)^2$$

- Variance of forecasts (population formula):

$$VAR = \frac{1}{T} \sum_{t=1}^{T} (F_t - \bar{F}_t)^2$$

- **Bias:** when a forecast is persistently wrong, either in the positive direction or negative direction. Bias = $\sqrt{\text{MSE - VAR}}$
- **Root Mean Squared Error (RMSE) =** $\sqrt{\text{MSE}}$.