New Models and Methods for Time Series for Time Series Analysis in Big Data Era

Dr. Xialu Liu
Management Information Systems Department
San Diego State Univeristy

Big Data Era


BIG
DATA
ALSO
TOOLS
DISK
TARGET
APPLIED
SENSOR
MAY
EXAMPLES
SHARED
DEFINITION
CURRENT
MOVING
RECONSIDER
WITHIN
EVERY
PRACTITIONERS
CAPTURE
ZETTABYTES
THOUGHT
BUSINESS
INTERNET
DISTRIBUTED
SETS
GENOMICS
COMPLEXITY
DESCRIBING
RADIO-FREQUENCY
MANAGEMENT
TERABYTES
CASE
ABILITY
PETABYTES
INCLUDE
ELAPSED
SYSTEMS
TOLERABLE
FORMS
SIZE
STORAGE
MPP
PARALLEL
SAN
QUALITIES
SINCE
MASSIVELY
GROW
CONNECTOMICS
ORGANIZATIONS
SET
RELATIONAL
USE
LARGER
CONTINUES
SOCIAL
INDEXING
TENS
COMPLEX
CITATION
ANALYTICS
NOW
BIOLOGICAL
PROCESSING
UBIQUITOUS
TYPES
CAPACITY
BURIED
WORLD'S
SOLID
HUNDREDS
NETWORKS
DATABASES
AMOUNT
DESKTOP
RECORDS
CURRENTLY
GARTNER
WORKING
OPPORTUNITIES
SEARCH
DIFFICULTY

Data

Data is everywhere.

Data

Data is everywhere.

- How good is this business school?

Data

Data is everywhere.

- How good is this business school?
  Graduate starting salary

Data

Data is everywhere.

- How good is this business school?
  Graduate starting salary
- How good is the movie?

Data

Data is everywhere.

- How good is this business school?
  Graduate starting salary
- How good is the movie?
  IMDb rating, box office

Data

Data is everywhere.

- How good is this business school?
  Graduate starting salary
- How good is the movie?
  IMDb rating, box office
- How popular is this topic?

Data

Data is everywhere.

- How good is this business school?
  Graduate starting salary
- How good is the movie?
  IMDb rating, box office
- How popular is this topic?
  Number of tweets

Data is everywhere.

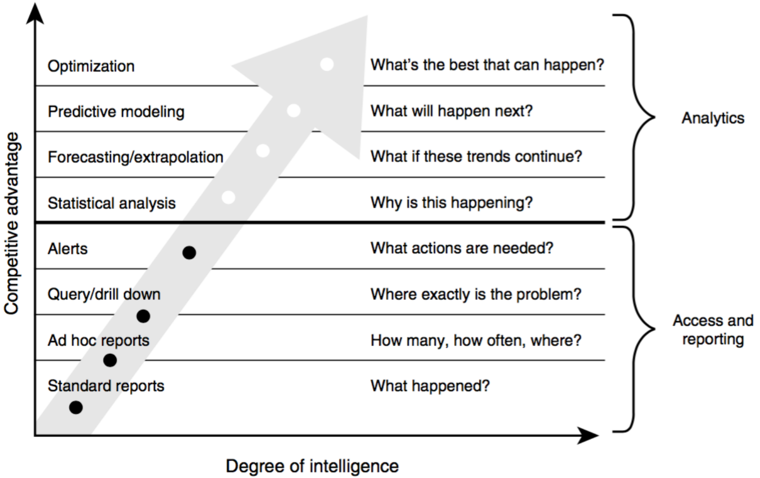
- How good is this business school?
  Graduate starting salary
- How good is the movie?
  IMDb rating, box office
- How popular is this topic?
  Number of tweets
- How much does the man love this woman?

Data

Data is everywhere.

- How good is this business school?
  Graduate starting salary
- How good is the movie?
  IMDb rating, box office
- How popular is this topic?
  Number of tweets
- How much does the man love this woman?
  The carat of the diamond

Level of Intelligence



-Gorman, M. F., Klimberg, R. K. (2014). Benchmarking academic programs in business analytics. *Interfaces.* **44(3)**, 329-341.

Time Series Data

Time series

- A sequence of data points observed over time
- Successive data points are usually expected o be dependent
- Examples: the daily weight, the daily return of stocks, the three-minute traffic volume, historical financial data of a firm

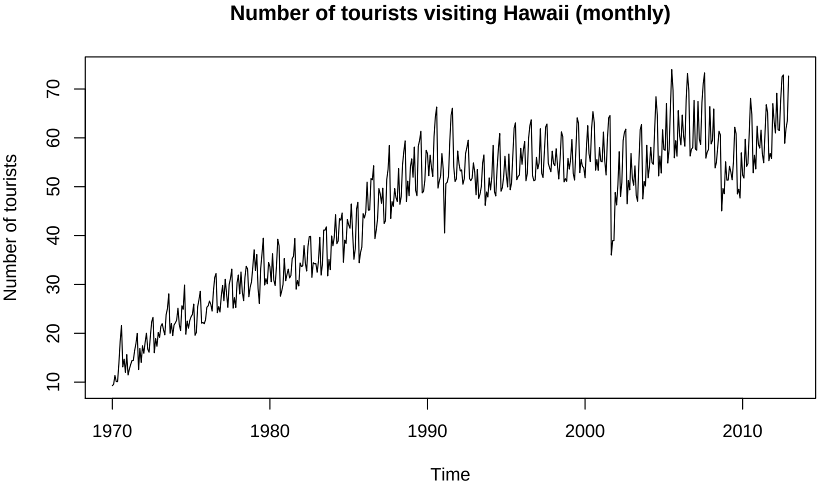
Why to study time series?

- Understanding the underlying mechanism driving the process (why is this happening?)
- Forecasting (what will happen next?)

Time Series Data

- Traditional time series analysis deals with scalar or vector observations
  - Linear: ARIMA and seasonal ARIMA models
- In big data era, information available becomes massive and complex. My research focuses on time series data in
  - nonlinear dynamics
  - functional form
  - high dimension

Hawaii Tourism Data: 1970-2012



We re-scale the data by dividing 10^5.
Source: Hawaii Visitors Bureau.

Hawaii Tourism Data

Questions: how to understand and predict the number of tourists? why is it so important?

- Tourism is the largest single source of the state GDP, representing about \$14 billion, 21% of its entire economy
- Tourism contributed \$1.5 billion in total state tax revenue in 2013
- Most local service industries rely heavily on tourism , for example, airlines, hotels, casinos, shopping malls, theaters
- It is important for state budget and for supply chain management of local firms

Hawaii Tourism Data: Seasonal ARIMA Models

Seasonal ARIMA model:

$$(1 - \phi_1 B)(1 - \phi_{12} B^{12})(1 - B^{12}) X_t = \varepsilon_t,$$

where B is backshift operator, i.e. $B^q X_t = X_{t-q}$.

Define time series Y_t by taking a seasonal difference,

$$Y_t = (1 - B^{12}) X_t = X_t - X_{t-12}.$$

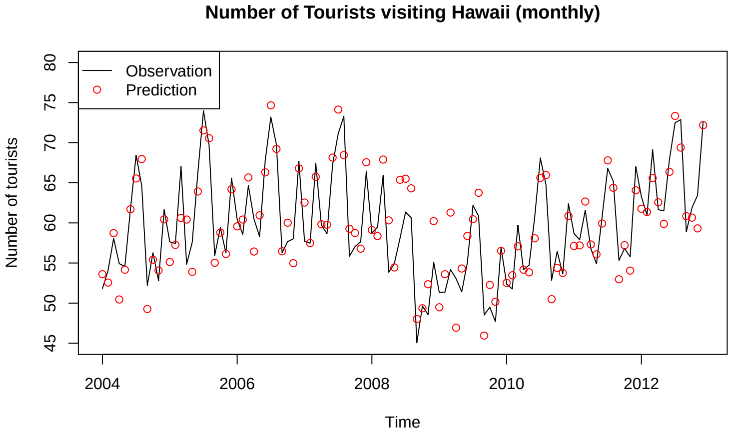
Then

$$(1 - \phi_1 B)(1 - \phi_{12} B^{12}) Y_t = \varepsilon_t$$

$$Y_t = \phi_1 Y_{t-1} + \phi_{12} Y_{t-12} + \phi_1 \phi_{12} Y_{t-13} + \varepsilon_t.$$

Hawaii Tourism Data: Seasonal ARIMA models

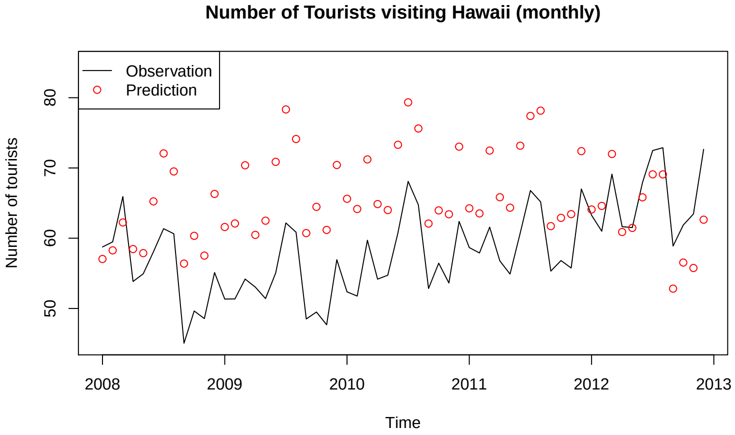
One month ahead out-sample forecasting from 2005 to 2012.



Mean squared error is 6.8475.

Hawaii Tourism Data: Seasonal ARIMA models

Four years ahead out-sample forecasting from 2008 to 2012.



Mean squared error is 89.8845.

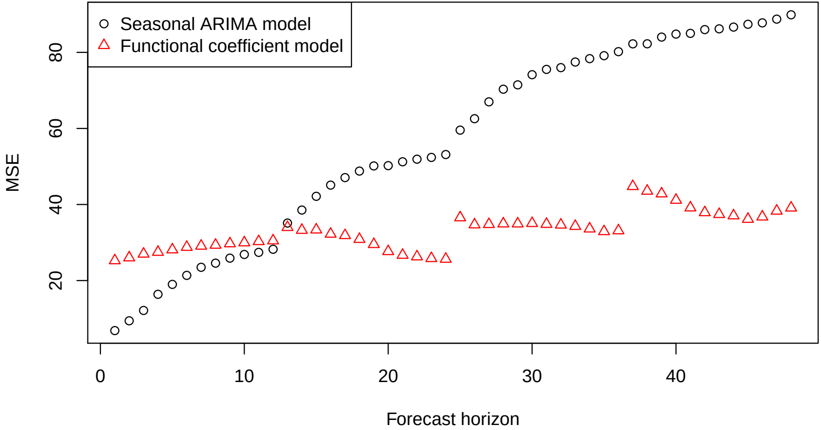
- Long term prediction does not work
- The impact of the past on the present changes, and nonliner dynamics exists
- Some exogenous variables may help

The growth rate of annual personal disposable income (PDI) of U.S. and Japan are added as $\{x_{1t}\}$ and $\{x_{2t}\}$, respectively, since U.S. and Japan contribute more than 80% of the tourists in Hawaii.

$$y_{tj} = [\alpha_0(s_t) + \beta_{0j}(s_t)] + [\alpha_1(s_t) + \beta_{1j}(s_t)]x_{1t} + [\alpha_2(s_t) + \beta_{2j}(s_t)]x_{2t} + e_{tj}$$

h is selected by generalized cross-validation.

Hawaii Tourism Data



Plots of mean squared error for our model and seasonal ARIMA model

Other applications

- Scalar time series data with seasonality and nonlinear dynamics
- Example: electricity consumption, airline traffic volume, etc

Why is volatility important?

- Volatility is a very popular topic in finance, and crucial for option pricing.
- Many hedging strategies depend on the volatility of assets.

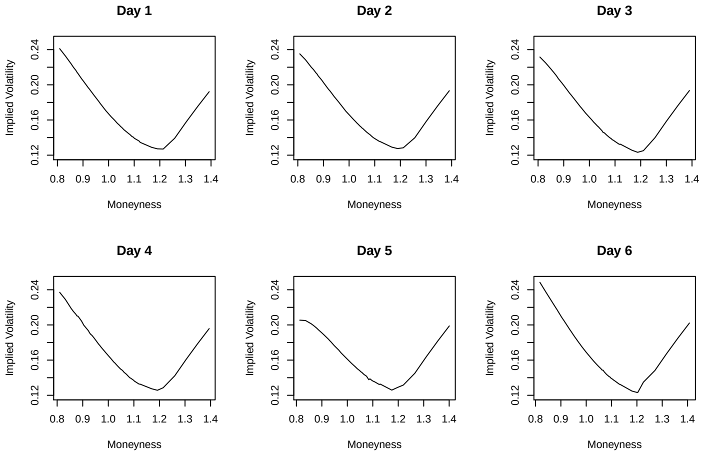
What is implied volatility?

- Black-Scholes model is the world's most well-known pricing model. Scholes won the Nobel Prize for this work in 1997.
- Volatility derived from Black-Scholes model is implied volatility
- Volatility smile: plot of implied volatility against moneyness (strike price/underlying asset price) yields a 'smile'

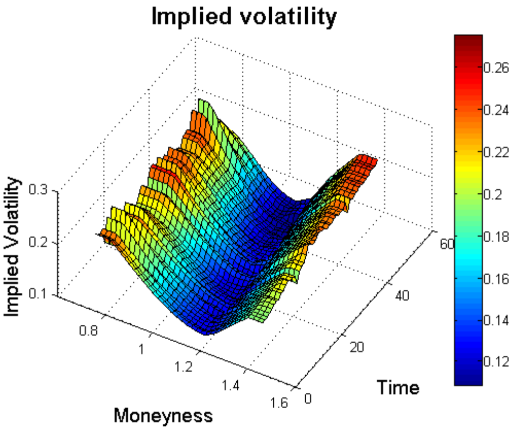
Volatility is treated as a function of moneyness, and our aim is to predict volatility curve.

Volatility Smiles

Daily implied volatilities of European call options of the S &P 500 index from July 9, 2004 to Sep 20, 2004. The expiration date is Dec 18, 2013. The strike prices range from 950 to 1550.



Volatility Smiles



Other Examples for Functional Time Series

Functional time series data are widely observed in many fields.

- Finance: yield curve
- Demography: age-specific mortality rate, birth rate
- Meteorology: temperature, participation, and cloud cover in a region

Convolutional FAR(p) Models

$$X_t(s) = \sum_{i=1}^{p} \int_0^1 \phi_i(s-u) X_{t-i}(u) du + \varepsilon_t(s) \text{ where } \ s \in [0,1]$$

- $\phi(\cdot)$ is defined on $[-1, 1]$.
  - Finite support for integration of estimated $\phi(\cdot)$.
  - $w_s(u) = \phi(s-u)$ is the weight function for s.
- Noise process $\{\varepsilon_t(\cdot), t = 1, \ldots, T\}$ is assumed to be i.i.d following an Ornstein-Uhlenbeck process

$$d\varepsilon_t(s) = -\rho \varepsilon_t(s) ds + \sigma dW_t(s)$$

with $\mathrm{Var}(\varepsilon_t(s)) = \sigma^2/2\rho$, $\mathrm{Cor}(\varepsilon_t(s_1), \varepsilon_t(s_2)) = \exp\{-\rho|s_1 - s_2|\}$.

Volatility Smiles Prediction

Table: Out-of-sample forecasting MSEs for different models

MSE1(CFAR model)	MSE2(FAR model)	MSE3(random walk)
$8.7134e-05$	$1.1010e-04$	$1.1938e-04$