ODTUG Kscope15

HOLLYWOOD, FLORIDA
JUNE 21-25, 2015

Forecasting, Prediction Models, and Times Series Analysis with Oracle Business Intelligence and Analytics

ODTUG Kscope 15

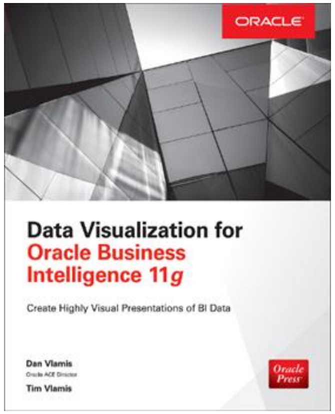
Dan Vlamis and Tim Vlamis
Vlamis Software Solutions
816-781-2880
http://www.vlamis.com

Presentation Agenda

- Understanding classification and forecasting (predictions)
- Use of Geneva Forecasting engine in Oracle OLAP
  - Holt-Winters and time series
  - Parameter choices
- ARIMA forecasting algorithm in R
  - Use Oracle R Enterprise
- Use of time dimension and time series functions in OBI

Vlamis Software Solutions

- Vlamis Software founded in 1992 in Kansas City, Missouri
- Developed more than 200 Oracle BI systems
- Specializes in ORACLE-based:
  - Data Warehousing
  - Business Intelligence
  - Data Mining and Predictive Analytics
  - Data Visualization
- Expert presenter at major Oracle conferences
- www.vlamis.com (blog, papers, newsletters, services)
- Co-authors of book “Data Visualization for OBI 11g”
- Co-author of book “Oracle Essbase & Oracle OLAP”
- Oracle University Partner
- Oracle Gold Partner



ORACLE EDUCATION RESELLER

ORACLE Gold Partner
Specialized
Oracle Business Intelligence Foundation Suite 11g

vlamis SOFTWARE SOLUTIONS

Dan and Tim Vlamis

Dan Vlamis

ORACLE ACE Director

- Founded Vlamis Software Solutions in 1993
- 25+ years in business intelligence, dimensional modeling
- Oracle ACE Director
- Developer for IRI (expert in Oracle OLAP and related)
- BA Computer Science Brown University

Tim Vlamis

ORACLE ACE

- 25+ years experience in business modeling and valuation, forecasting, and scenario analyses
- Oracle ACE
- Instructor for Oracle University's Data Mining Techniques and Oracle R Enterprise Essentials Courses
- Professional Certified Marketer (PCM) from AMA
- Adjunct Professor of Business Benedictine College
- MBA Kellogg School of Management (Northwestern University)
- BA Economics Yale University

vlamis™ SOFTWARE SOLUTIONS

Forecasting Today

- Predictions are the holy grail of BI systems and initiatives.
- Most all corporations have need for forecasting.
- Typical forecasting systems
  - Are stand alone or from ERP (not integrated to BI system)
  - Tend to use straight line or heuristic calculations.
  - Not always integrated into the business.
  - Are often tied directly to the budgeting process
- High level of angst surrounding forecasts.

Forecasting Should…

- Should be integrated with rest of BI system.
- Should be another series of measures that are revealed in the context of historic information.
- Should be a part of the Common Enterprise Model.
- Should have visibility across functional areas and roles in corporations
- Should leverage most powerful calculation tools (database and BI system)
- Ideally adjusted based on an integrated view across corporate functions (marketing, operations, finance, etc.).

Forecasting Methodologies

- Rule-based heuristic (last period, last period +5%, etc.)
- Cross-sectional methodologies (point in time)
- Time series (time sequenced data series)
- Mixed models
- Averages (moving, weighted, etc.)
- Linear and Non-linear regressions (line fitting)
- Transforms, projections, min/max

Methodologies for Today

- OLAP Geneva Forecasting Engine
  - Holt Winters for time series
- Oracle R Enterprise
  - ARIMA
- ODM Classification and Regression (overview)
- OBIEE Time Series Functions (overview)

OLAP Geneva Forecasting Engine

- The Geneva Forecasting Engine is set of programs that have been implemented into many popular demand forecasting systems.
- Oracle OLAP has integrated the Geneva Forecasting engine.
- It offers a set of algorithms and control functions for automatically generating forecasts.

OLAP DML Commands for GFE

- FCOPEN function -- Creates a forecasting context.
- FCSET command -- Specifies the forecast characteristics.
- FCEXEC command -- Executes a forecast and populates Oracle OLAP variables with forecasting data.
- FCQUERY function -- Retrieves information about the characteristics of a forecast or a trial of a forecast.
- FCCLOSE command -- Closes a forecasting context.

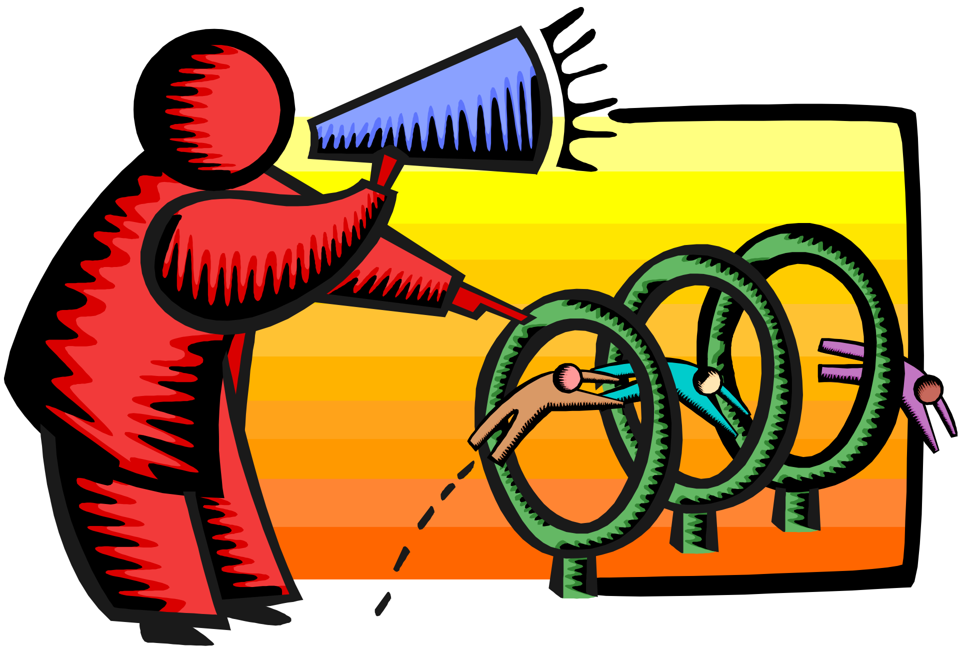
METHOD ‘method’

- **AUTOMATIC** best fit for the data. (Default)
- **LINREG** linear regression (y=a*x+b) is fitted to the data.
- **NLREG1** nonlinear regression x'=log(x) and y'=log(y) a polynomial model between x and y(y=c*x^a).
- **NLREG2** nonlinear regression x'=x and y'=ln(y) an exponential model between x and y(y=c*e^ax).
- **NLREG3** nonlinear regression x'=log(x) and y'=y a logarithmic model between x and y(y=a*log(x)+b).
- **NLREG4** nonlinear regression method ix'=1/x and y'=1/y an asymptotic curve (y=x/(a+bx)).
- **NLREG5** nonlinear regression method x'=x and y'=ln(y/(K-y)) an exponential asymptotic curve (y=cKe^ax/(1+ce^ax)).
- **SESMOOTH** single exponential smoothing method intended for short term forecasts of non-seasonal data.
- **DESMOOTH** double exponential smoothing method exponential smoothing is applied to both the series and the trend term.
- **CROSTON** Croston's Intermittent Demand method. used for intermittent data where more than half of the observations are zero
- **HOLT/WINTERS** “triple” exponential smoothing. used on seasonal data

Using “Holt-Winters”

- Triple “Exponential Smoothing” methodology
- Used for data suspected to be seasonal
- Needs multiple seasons
- Assumes regular periods
- Pre/post processing may be necessary (fiscal calendar 445, irregular holidays, “Black Swans”, outages, etc.)

OBIEE Demo of Time Series Decomposition



Exponential Smoothing

- Methodology for smoothing data and preferencing more recent periods when doing time series forecasts.
- Similar conceptually to a weighted moving average
- Weights decline according to an exponential function. $\{1, (1-\alpha), (1-\alpha)^2, (1-\alpha)^3, \ldots\}$
- Higher values give more weight to more recent periods

- Single (weighted average of most recent observation and the most recent smoothed statistic)
- Double (trend either up or down)
- Triple (period effect)

FCSET Parameters

- **ALLOCLAST** {YES|NO}
- ALPHA {MAX|MIN|STEP} decimal
- **APPROACH** {'APPAUTO'|'APPMANUAL'
- BETA {MAX|MIN|STEP} decimal
- COMPSMOOTH {YES|NO}
- CYCDECAY {MAX|MIN} decimal
- GAMMA {MAX|MIN|STEP} decimal
- **HISTPERIOD**S integer
- MAXFACTOR decimal
- **METHOD** 'method'
- MINFCFACTOR decimal
- MPTDECAY {MAX|MIN} decimal

NTRIALS integer

- **PERIODICITY** cycle-spec
- RATIO decimal
- **SMOOTHING** {YES|NO}
- TRANSFORM {'TRNOSEA'|'TRSEA'|'TRMPT'}
- TRENDHOLD {MAX|MIN|STEP} decimal
- WINDOWLEN integer

Alpha, Beta, Gamma Setting

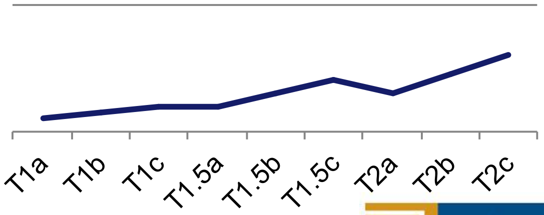
- Default Max is 0.3
- Default Min is 0.1
- Default Step is 0.1 (.05<= divisible value<=0.2)
- Greater value means nearer periods have more weight.
- Lower value means periods have more equal weight.

Recommendations

- Be careful of accepting the APPAUTO setting
- Be aware of Embedded total time dimensions
- Match HISTPERIODS with PERIODICITY for best results
- PERIODICITY cycle-spec is hierarchical from higher grain to lower
    - Ex {52,7} 52 weeks in a year, 7 days in a week
    - Ex {4,13,7} 4 quarters in a year, 13 weeks in a quarter, 7 days in a week
    - Ex {12} 12 months in a year
  - Months are challenging to incorporate with other periods

Case Study Using Oracle OLAP

- Forecasted values from Oracle OLAP made no sense
- Client trying to use Best Fit – complicates study because don't know what method chosen
- Avoid tendency to inherit mistakes
- Problem in "HISTPERIODS" parameter
  - Solution: set HISTPERIODS to number of data points
- Problem in forecasting on hierarchical dimension – 12 month periods, 1 year period throwing off forecast
  - Solution: LIMIT TIME TO TIMELEVEL 'PERIOD'
- 4-4-5 "periods" artificially inflate every 3rd period
- Added 3rd year – average of 2 years

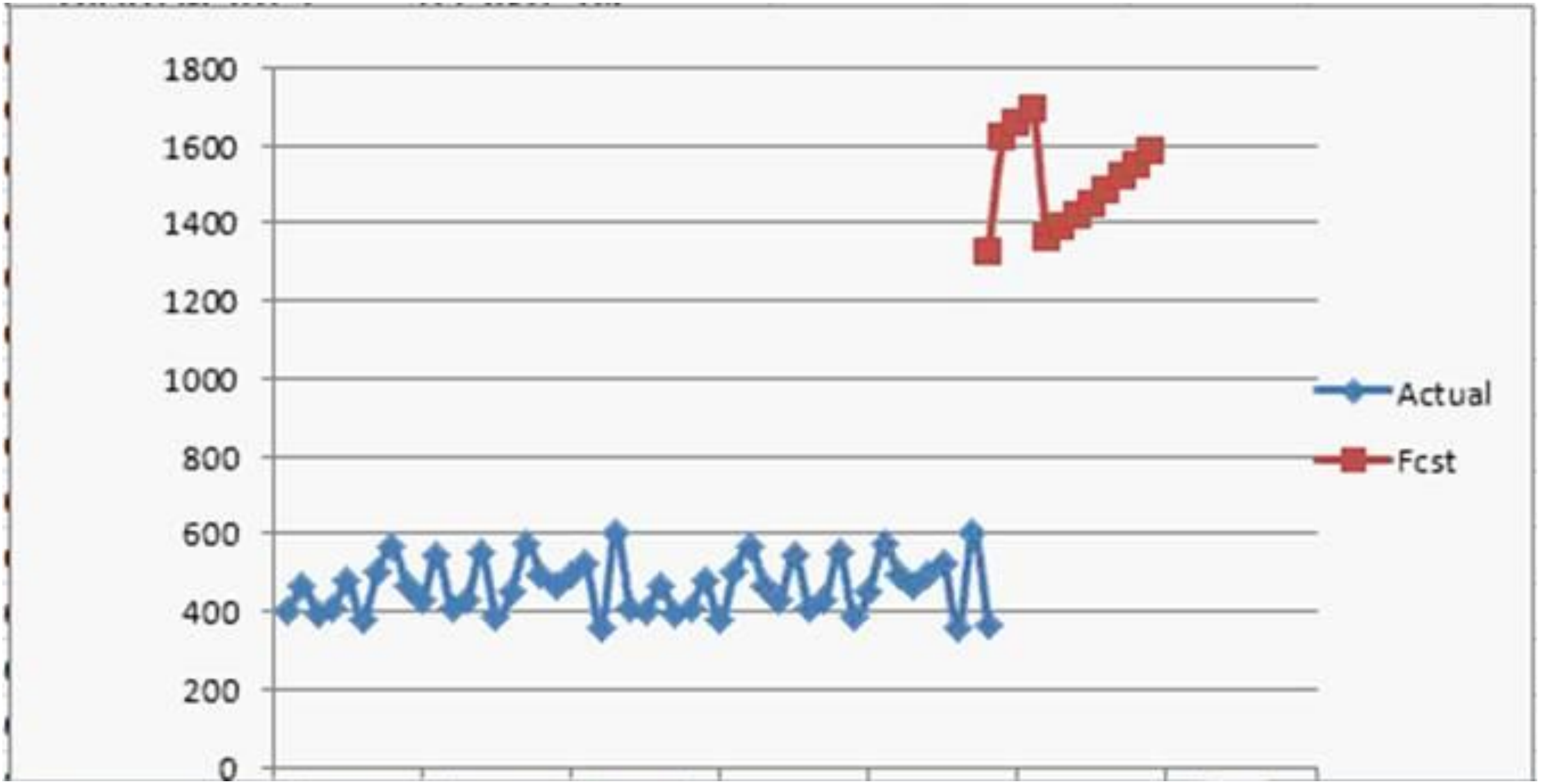


Example OLAP DML Forecast Program

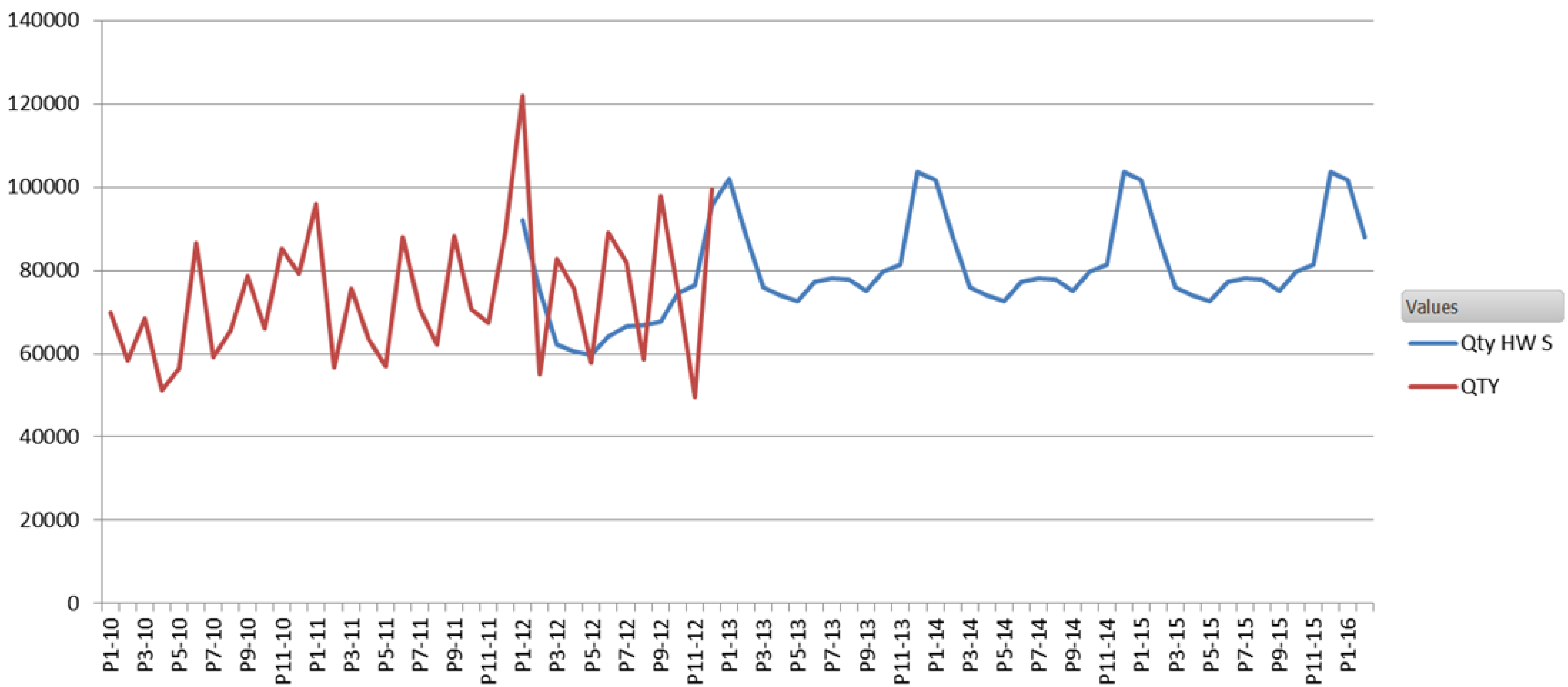
```
vrb _handle  int

“ Removed error handling and definition of temporary variables such as DJOFCST2_C_SEASONAL
LIMIT DJOFCST2_C_MEASURE_DIM TO 'QTY_HW'

_handle = FCOPEN('MyForecast')

limit djotime_d2 to djotime_d2_levelrel eq 'PERIOD'
SORT DJOTIME_D2 a DJOTIME_D2_END_DATE
"Set forecast parameters for 'best fit'
fcset _handle method 'HOLT/WINTERS' APPROACH 'APPMANUAL' SMOOTHING 'YES' MAXFCFACTOR 10.0 TRANSFORM 'TRSEA' -
     periodicity 12 histperiods 36 BETA MAX 0.5

"Execute the forecast - save seasonal and seasonal smoothed into the variables just defined
FCEXEC _handle time DJOTIME_D2 INTO DJOFCST2_C_STORED -
       seasonal DJOFCST2_C_SEASONAL -
       smseasonal DJOFCST2_C_SMSEASONAL backcast DJOFCST2_C_QTY

ALLSTAT
"Close the forecast
FCCLOSE _handle

update
commit

return
```

Forecasts Did Not Make Sense

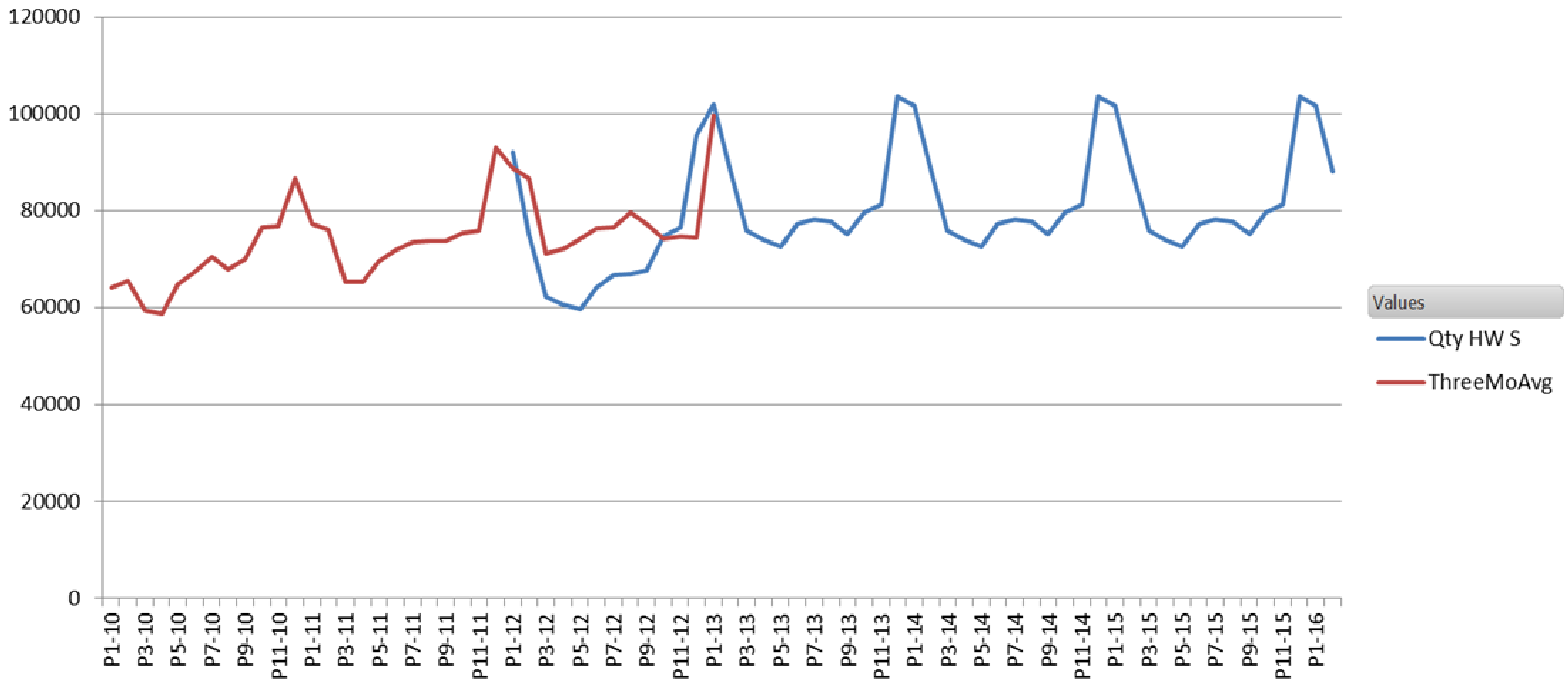
Forecasts Did Not Make Sense



Holt-Winters Forecast After Fix



Holt-Winters Vs. 3-Mo Moving Avg



Common Transformations

- Use average value per period to eliminate differences among periods (especially months)
- Shorter periods can reveal interesting patterns (e.g. average daily sales rather than average

Essbase @TREND

- Includes single, double, and triple exponential smoothing techniques.
- Includes linear and non-linear regression option.
- Does not include an auto-choice function.
- Non-linear regression transforms must be manually applied.

- Many other transform, calculation, and modeling capabilities in Essbase.

ARIMA

- Autoregressive Integrated Moving Average
- Powerful algorithm for series analysis and prediction
- Three parameters (p, d, q)
  - Auto regression (how reliant series values are on previous series values). AR(0) is white noise.
  - Integrated (degree of AR differencing, Random Walk)
  - Moving average (smoothing function)
- ARIMA (1,0,0) = AR(1)
- ARIMA (1,0,1) = ARMA (1,1)
- Large number of potential models
- Know the name Rob Hyndman for ARIMA in R

https://www.otexts.org/fpp/

Stationarity

- Processes with no growth related to time.
- Random walks are stationary.
- Necessary to difference non-stationary series before applying ARMA models. (ARIMA handles this through the "Integrated" term "d"of the (p, d, q) model parameters.)

Non-Seasonal ARIMA (p, d, q)

- $\phi(B)(1 - B^d)\gamma_t = \mathrm{c} + \theta(B)\varepsilon_t$
- $\{\varepsilon_t\}$ is a white noise process with 0 mean and variance σ^2.
- B is a backshift operator
- $\phi(z)$ is a polynomial of order p
- $\theta(z)$ is a polynomial of order q

Seasonal ARIMA $(p, d, q)(P, D, Q)_m$

- $\Phi(B^m)\phi(B)(1-B^D)\left(1-B^d\right)\gamma_t = \mathrm{c} + \Theta(B^m)\theta(B)\varepsilon_t$
- $\{\varepsilon_t\}$ is a white noise process with 0 mean and variance σ^2.
- B is a backshift operator
- $\Phi(z)$is a polynomial of order p
- $\Theta(z)$is a polynomial of order q

Forecast() package in R

Includes methods:

- ets()
- auto.arima()
- Arima()
- arima()
- HoltWinters()
- StructTS()

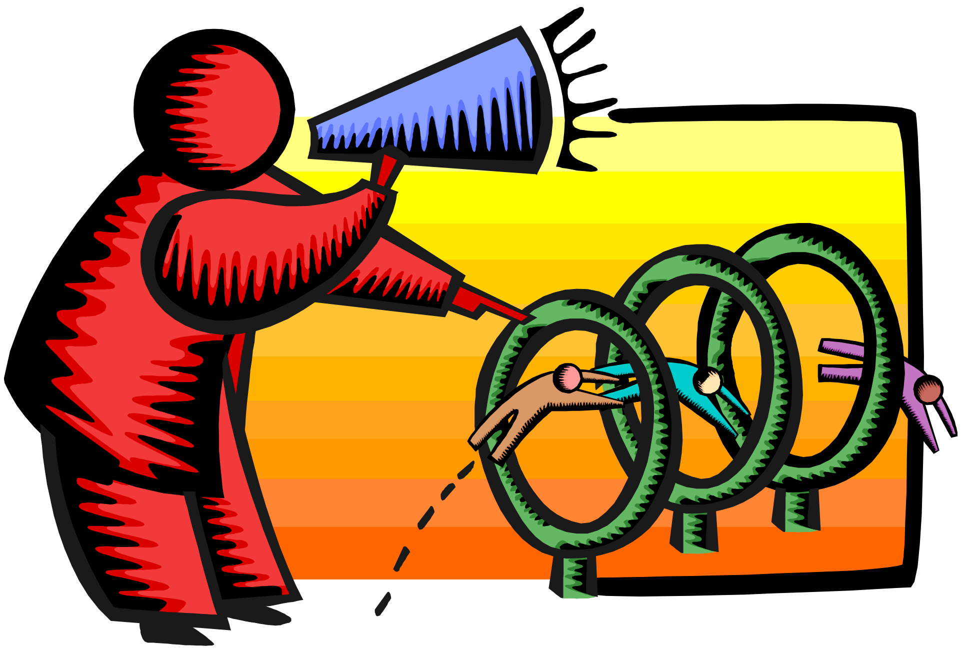
Produces

- Simple forecasting
- Auto chooses best model (smallest AIC)
- Choose the model yourself
- Somewhat limited; use Arima()
- Exponential smoothing (seasonal)
- Maximum likelihood fit (ARIMA 0,2,2)

Choosing an ARIMA model

- Auto.arima can be used for model choice.
- Manual model choice requires hypothesis testing and evaluation of results.
- Use minimum AIC to chose best model
  - $AIC = -2log(L) + 2(p + q + P + Q + k)$
  - Compare AIC values to each other, absolute values carry no meaning

arima Demo



ARIMA vs. Holt-Winters

- Holt-Winters can be used for series that are seasonal and have a trend. (require order 2 differencing in ARIMA)
- Model selection can be complex in ARIMA and auto.arima selection may not be well understood.
- ARIMA best for stationary data series.
- ARIMA very powerful, but more to learn.
- Initial values more important in ARIMA (can have a big effect on predictions depending on model selected.)
- ARIMA provides confidence intervals

Time Series Functions in OBI 11g

- Very powerful, accessible capability
- Time dimension must be designated
- Query results must be exact to pull from cache
- Can be “expensive” in processing
- Make sure that unique keys are defined at each level (“Jan13” rather than “Jan”)

AGO function

- Defines a time-based offset
- Can nest multiple AGO statements (same level)
- Ago(<<Measure>>, <<Level>>, <<Number of Periods>>)
- Measure is a fact such as sales.
- Level is an optional term, default is set by the grain of the query (BY clause) or is specified in repository for level based measures.
- Number of periods is an integer specifying the offset value.

TODATE

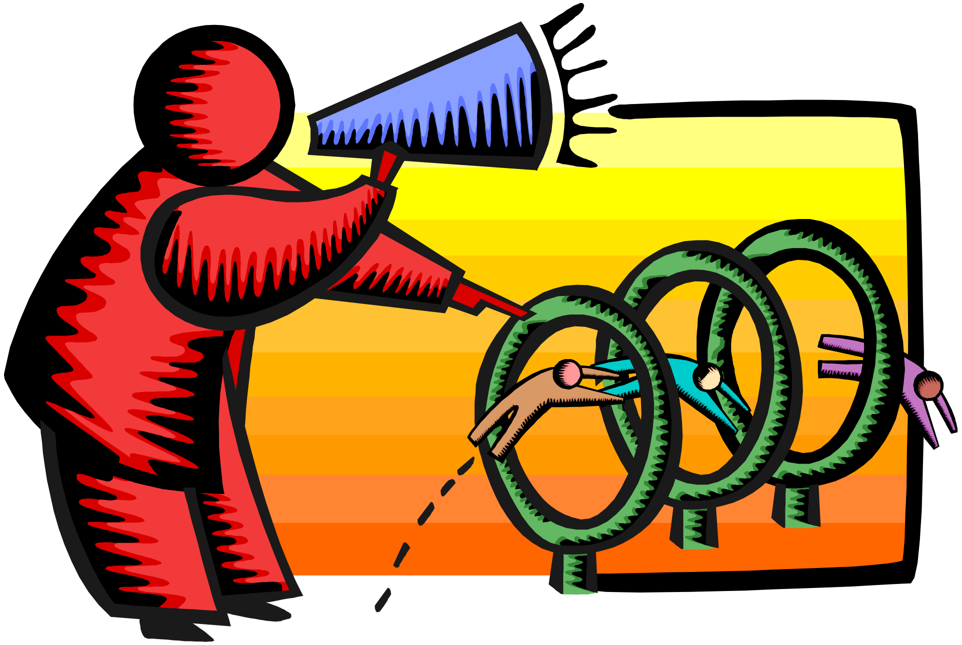
- Time-based aggregation function.
- Calculates based on starting value to current.
- Can nest with AGO (same level)
- ToDate(<<Measure>>, <<Level>>)
- Measure is a fact such as sales
- Level is the time grain such as year or month

PERIODROLLING

- Defines a period of time contextually
- Performs an operation across a specified set of query grain periods
- PeriodRolling(<<Measure>>, <<Starting Period Offset>>, <<Ending Period Offset>>, <<[Hierarchy]>>)
- Measure is a fact such as sales
- Starting Period Offset is an integer value, use a minus sign (“-2” means 2 periods ago)
- Ending Period Offset defines the end of the period, use a zero for current period
- Hierarchy is an optional setting to specify which time hierarchy to use such as “fiscal”
- Use “unbound” for starting period offset to calculate total from beginning
- PeriodRolling uses either the query level grain of “measure” or the measure level for “measure” if it has been set in the Admin tool.

vlamis™ SOFTWARE SOLUTIONS

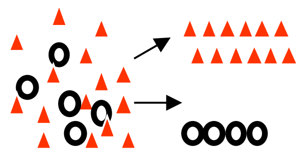
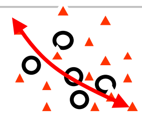
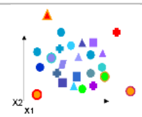
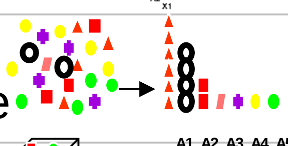
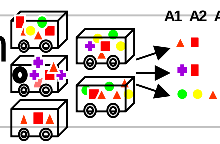
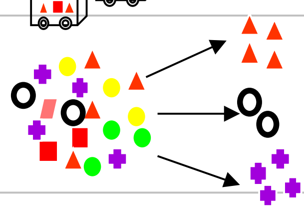
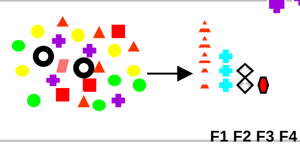
Oracle BI Trend Demo



Oracle Data Mining

- Oracle Data Mining is an option for the Enterprise Edition of the Oracle Database.
- A collection of APIs and specialized SQL functions.
- Includes a large number of specialized algorithms and built-in procedures.
- Makes use of many built-in capabilities of the Oracle Database
- ODM typically refers to “Oracle Data Mining”

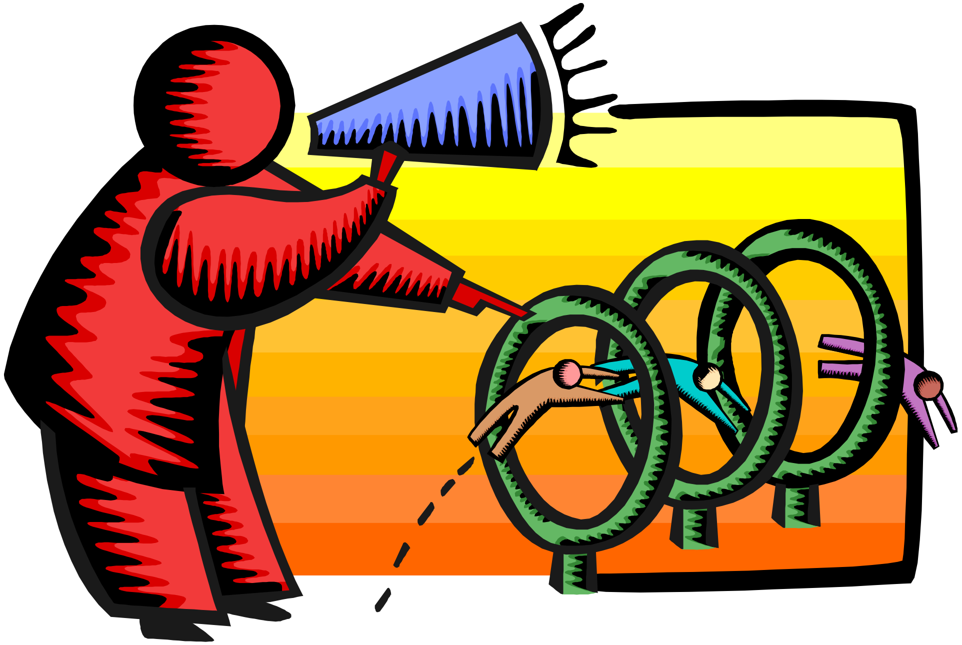
Oracle Data Mining Algorithms

Problem	Algorithm	Applicability
Classification 	Logistic Regression (GLM) Decision Trees Naïve Bayes Support Vector Machine	Classical statistical technique Popular / Rules / transparency Embedded app Wide / narrow data / text
Regression 	Multiple Regression (GLM) Support Vector Machine	Classical statistical technique Wide / narrow data / text
Anomaly Detection 	One Class SVM	Fraud Detection
Attribute Importance 	Minimum Description Length (MDL)	Attribute reduction Identify useful data Reduce data noise
Association Rules 	Apriori	Market basket analysis Link analysis
Clustering 	Hierarchical K-Means Hierarchical O-Cluster	Product grouping Text mining Gene and protein analysis
Feature Extraction 	NMF	Text analysis Feature reduction

Classification

- Prediction model for non-continuous information
  - Binary such as yes/no
  - Limited set (low/medium/high)
- Involves “supervised learning”
  - Prediction directed by a previously known dependent variable or “target” variable.
  - Commonly includes three phases:
    - Training
    - Testing
    - Scoring
- Results in predictive models that are applied to new data sets.
- In our example, we predict which prospects are likely to buy insurance.

Oracle Data Mining Demo



Oracle Test Drive

- Free to try out Oracle BI, Advanced Analytics and Big Data
- Go to www.vlamis.com/td
- Runs off of Amazon AWS
- Step-by-step exercises
- Test Drives for:
  - Oracle BI
  - Oracle Advanced Analytics
  - Big Data
- Once signed up, you have private instance for 3 hours
- Available now

BIWA SIG

Better Information Better Results

www.oraclebiwa.org

BIWA Summit 2016, Jan 26-28

Oracle HQ Conference Center

Business Intelligence, Warehousing and Analytics and Spatial

IOUG Special Interest Group

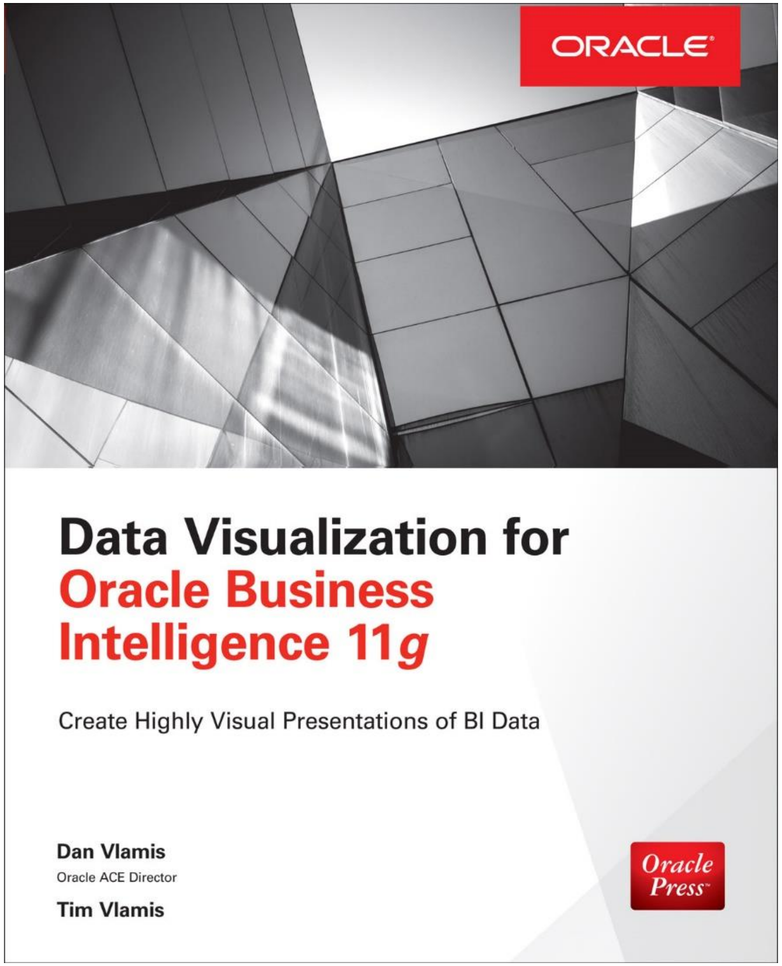
www.biwasummit.org

NoCOUG

IOUG independent oracle users group

Drawing for Free Book

- Add business card to basket or fill out card



Vlamis Kscope Presentations

Presenter	Session	Time	Title
Tim and Dan Vlamis	Session 1	Monday 8:30 - 9:30 AM	Forecasting, Prediction Models, and Time Series Analysis with Database Analytics and OBIEE
Dan and Tim Vlamis	Session 4	Monday 2:00 – 3:00 PM	Data Visualization for Oracle Business Intelligence 11g
Tim Vlamis and Michael Caskey	HOT-EPM	Tuesday 3:30 – 5:45 PM	Hands-on Training: Integrating Oracle Advanced Analytics into OBIEE Dashboards
Tim Vlamis and Michael Caskey	Session 11	Wednesday 8:30 - 9:30 AM	Starting Smart in Oracle Advanced Analytics
Mark Rittman, Alex Gorbachev and Tim Vlamis	Deep Dive	Thursday	Bringing Oracle Tools to Big Data

ODTUG Kscope15

HOLLYWOOD, FLORIDA

JUNE 21-25, 2015