Prediction from Quasi-Random Time Series

Lorenza Saitta
Dipartimento di Informatica
Università del Piemonte Orientale
Alessandria, Italy

The Problem

- Data
  - ORACLE DataBase structure:

    < Client, #Order, Product, Quantity, Value, Lit/Euro >

    Each order may contain more than one product, or the same product with different quantities and prices
  - Time series representing the daily sells of 16,265 products, grouped into 1,004 categories
- Goal
  - From years 1999, 2000, 2001 --> Predict year 2002

Pre-Processing

- Impossible to handle each product or each category separately

  --> Necessity to group/select products/categories

- Impossible to handle groups of products/categories for each client

  --> Necessity to group clients

Product Selection

- For each year :
  - Order products according to decreasing revenue sharing
  - Order products according to decreasing sold quantity
  - Select select products that globally cover 80% of revenue (sold quantity)

Grouping Products by Revenue Sharing

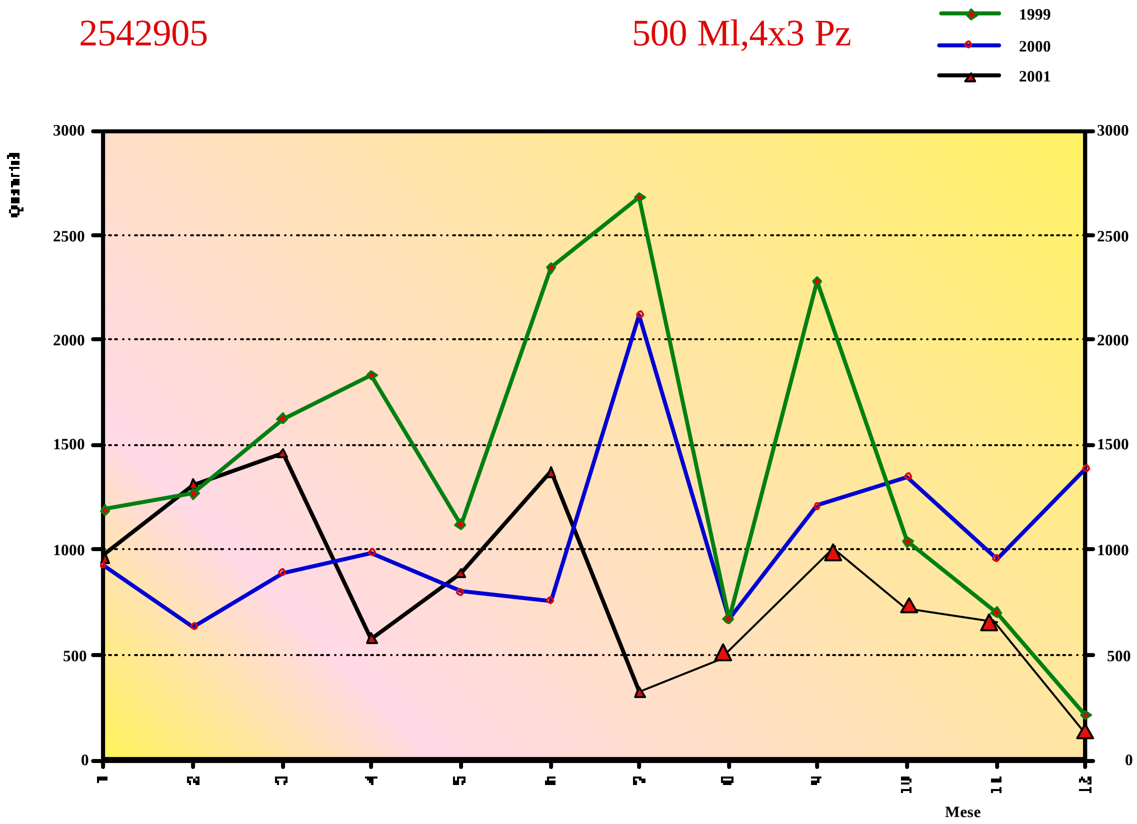
- Selection of those products that globally cover 80 % of revenue -> 53 Products

{3668995786, 9954462001, 5.3693}
{3179208029, 9950062610, 10.0219}
{2908553900, 9953071610, 14.2783}
{2833938896, 9950242825, 18.4256} { Value, Code, Cumulative function}
{2357611762, 9951252610, 21.8758}
{2111514418, 9952991610, 24.9659}
{2029888070, 9953091620, 27.9365}
…………
{2331985486, 9956521850, 79.6994}
{2325944507, 9952742830, 80.1763}

Annual Data

- Grouping by “days” or “weeks” produces apparently random time series
- Grouping starts being meaningful at the “month” level
- Necessity of correct alignment among years

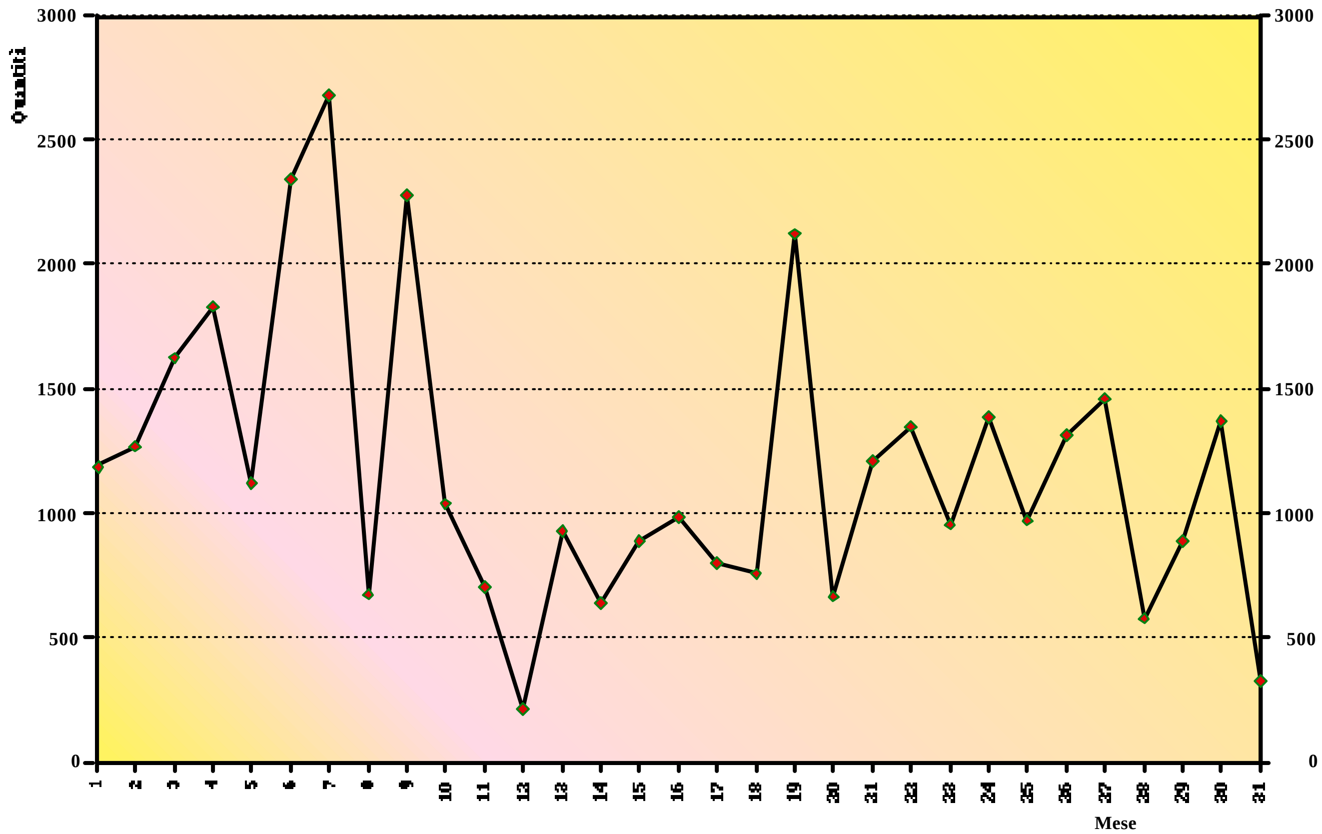
Yearly Comparison



Sequence over 31 Months

2542905 500 Ml, 4x3 Pz

Quantità



Overview of Analysis

- Single Products
  - General statistical characteristics
  - Yearly analysis (monthly, trimester data)
  - Three-year sequences
  - Time series analysis (trend, periodicity, oscillations, noise)
    - Additive Model $Y_t = T_t + C_t + S_t + R_t$
    - Multiplicative Model $Y_t = T_t * C_t * S_t * R_t$
  - Model of the selling process
- Products hierarchies

General Observation

- There is no substantial difference in behavior between the single and aggregated product analysis

Statistical Analysis

- Autocorrelation Coefficient (ACF)

  95% Significativity Plot

  (No significant correlations)
- Partial Autocorrelation up to 6 points (PACF)

  95% Significativity Plot

  (No significant correlations)
- Smoothing (Simple moving average, Spencer's and Henderson's weighted moving averages, EWMA, 3RSS)

  (Series are too short)

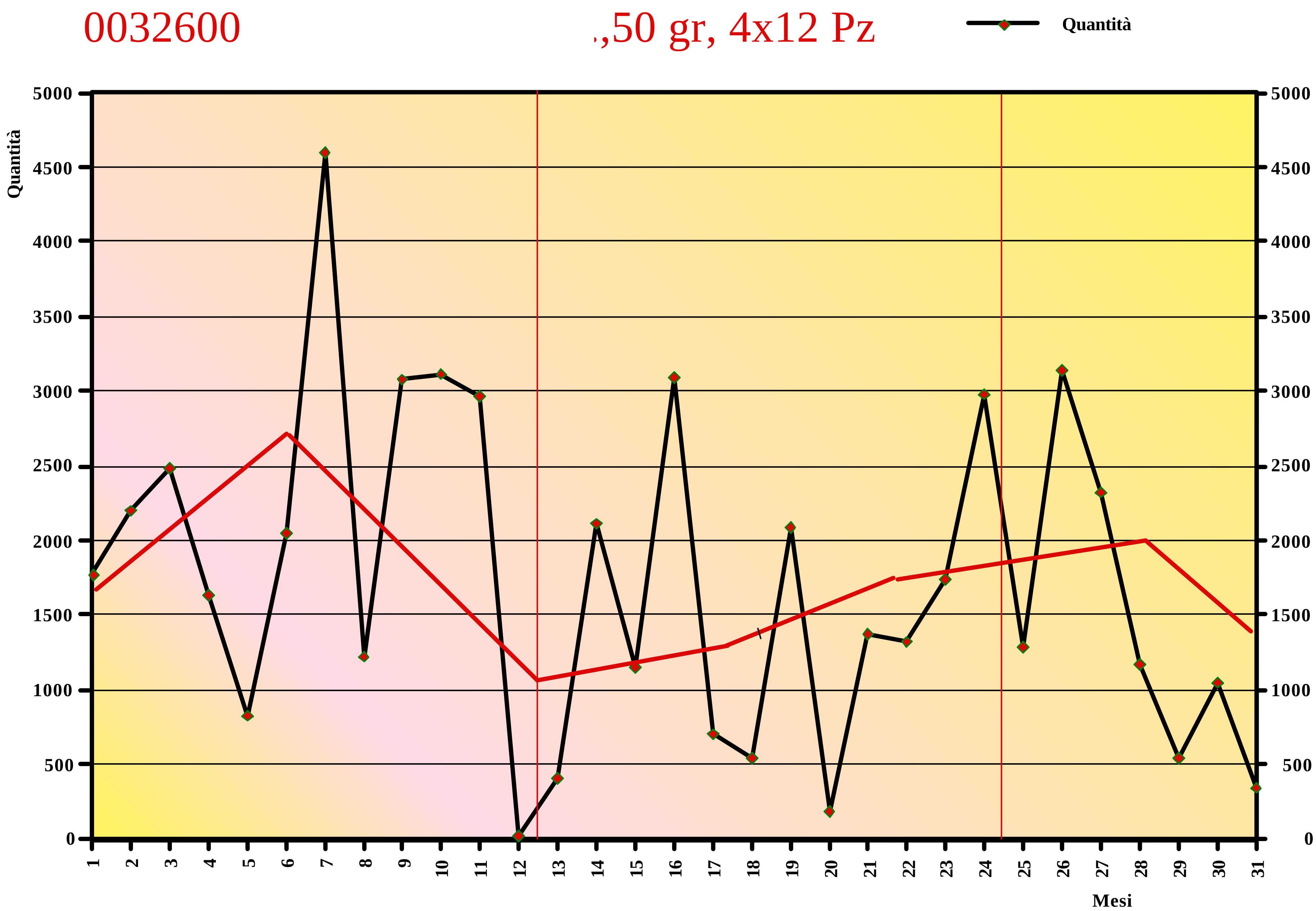
Statistical Analysis

- Periodgram (Fourier Analysis)
  (No clear seasonality)
- Randomness Test (Runs above and below median, Runs up and down, Box-Pierce test)
  (All serie pass every test)

Series Decomposition

- Trend, Seasonality, Cyclicity, Noise
  - Additive Model $Y_t = T_t + C_t + S_t + R_t$
  - Multiplicative Model $Y_t = T_t * C_t * S_t * R_t$

- Decomposition
  - Smooth the data using a moving average of length equal to the length of seasonality. This estimates the **trend-cycle**.
  - Divide the data by the moving average (if using the multiplicative method) or subtract the moving average from the data (if using the additive method). This estimates the **seasonality**.
  - Average the results for each season separately and rescale so that an average month equals 100 (multiplicative) or 0 (additive). This gives the **seasonal indices**.
  - Adjust the data for the estimated trend-cycle and seasonality, yielding the **irregular** or residual component.
  - Divide the original data values by the appropriate seasonal index if using the multiplicative method, or subtract the index if using the additive method. This gives the **seasonally adjusted data**.



0032600
.,50 gr, 4x12 Pz
Quantità
Quantità
Mesi
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31

Trimester Analysis

- By aggregating data by trimesters, a more stable behavior emerges
- Eight typical patterns

QuickTime™ and a Animation decompressor are needed to see this picture.

QuickTime™ and a Animation decompressor are needed to see this picture.

QuickTime™ and a Animation decompressor are needed to see this picture.

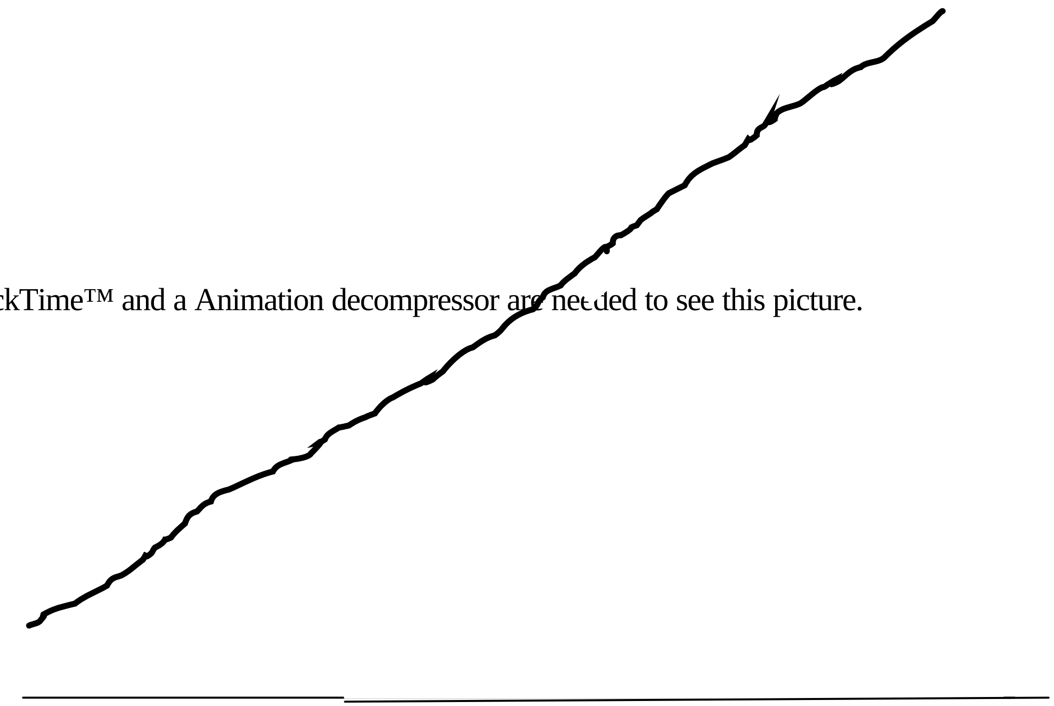
Trimester Analysis

- All relevant variables values are compared on the trimester level, year by year
- A unique sequence is formed for better approximability

Cumulative Analysis by Year

QuickTime™ and a Animation decompressor are needed to see this picture.

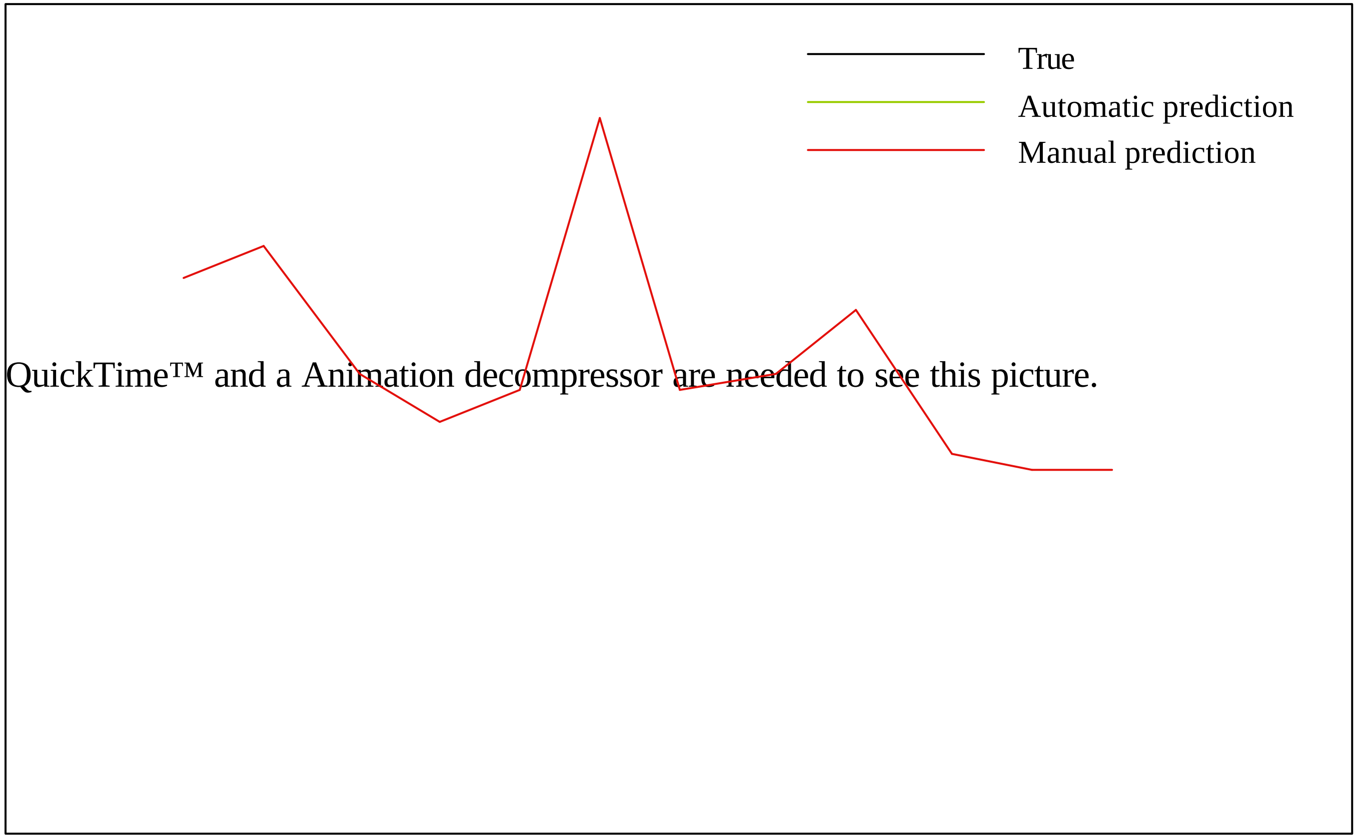
Global Time Series



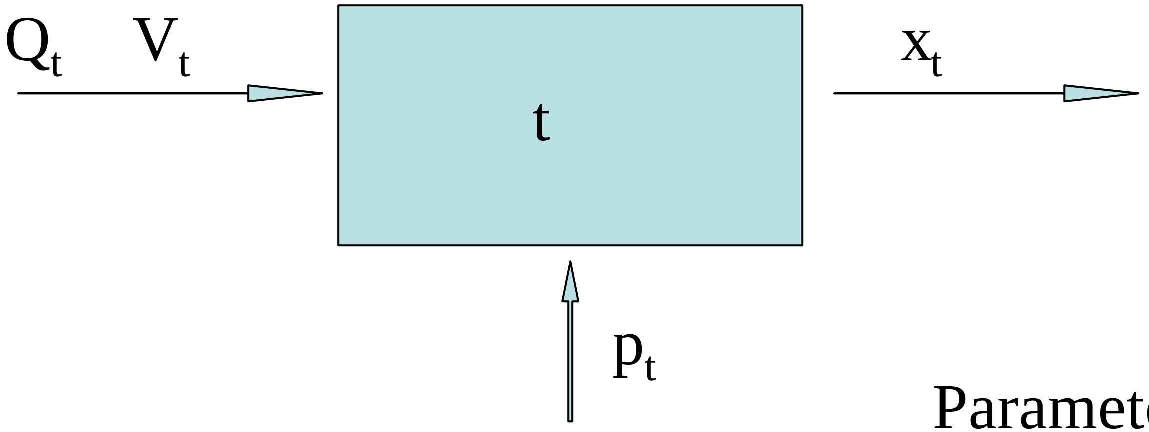
Comparison on Cumulative Graphs

QuickTime™ and a Animation decompressor are needed to see this picture.

Comparison on Differential Graphs



Model of the Selling Process



Parameters

Q = Target quantity to buy

Q_t = Bought quantity up to t

V_t = Expenditure up to t

Δ_t = Missing quantity

p_t = Offered price at t

x_t = Quantity to buy at t

$$Q_t = \sum_{i=1}^{t-1} (x_i + r_i) \qquad V_t = \sum_{i=1}^{t-1} v_i$$

$$\Delta_t = Q - Q_t$$

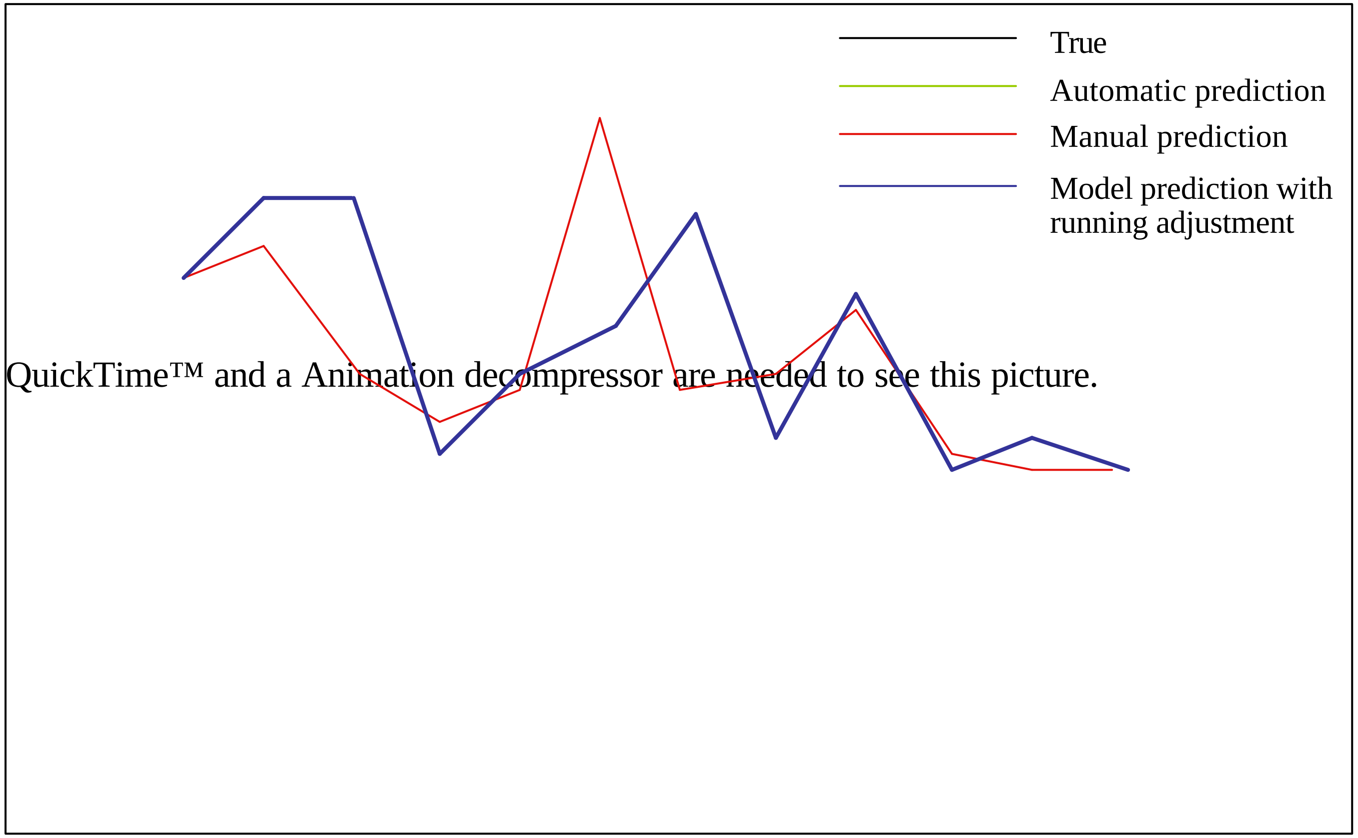
$$x_t = \frac{Q}{12}\left(1 + \frac{A}{p_t}\, e^{-\frac{a}{\Delta_t} - b V_t}\right)$$

Prediction from Model

- Regression on Cumulative function

$$x_t = y(t) + \frac{A}{p_t} e^{-\frac{a}{\Delta_t} - b V_t}$$

Comparison on Differential Graphs



Conclusions

- Prediction below the monthly scale is impossible
- Prediction at the trimester scale is possible
- Real time adjustments can be made if a two months delay is acceptable
- Results can be used as a Min-Max strategy in Game Theory