Predictive models and methods

Logistics

Giovanni Righini

Università degli Studi di Milano

Predictive models and methods

The content of this part is well covered by the textbook:

- G. Ghiani, G. Laporte, R. Musmanno, *Introduction to Logistics Systems Management*, Wiley, 2003

Predictive models and methods are used to extract information to make forecasts, in order to support decision processes based on data.

Forecasts may have different time horizons:

- short term (e.g..: number of calls to a call center tomorrow)
- medium term (e.g.: sales in the yearly business plan of a company)
- long term (e.g.: demand of hydrogen-powered cars in the next twenty years)

Classification

Forecasting methods can be classified as qualitative and quantitative.

Qualitative methods:

- Experts opinions
- Market polls
- Delphi method

Quantitative methods:

- Explicative methods: we assume there is a cause-effect relationship that we want to describe;
- Extrapolative methods: we want to extract regularities from the observed data.

Regression analysis

The goal is to identify a functional relationship between an effect and its (assumed) causes.

One observes a quantity y (dependent variable) and assumes it is a function of other quantities x (independent variables).

$$y = f(x)$$

If the independent variable is only one, the method is called simple regression. Otherwise it is called multiple regression.

From previous observations some pairs of values (x_i, y_i) are known; the aim is to find the function $f()$ that best represents them.

Regression analysis

Instead of searching for a complicated function that represents the observations exactly, it is preferred to search for a simple function that represents them approximately.

Hence we allow for a difference between the values computed as $f(x_i)$ and the observed values y_i.

If $f()$ is linear, the method is called linear regression.

$$y = A + Bx + \epsilon$$

The difference ϵ between computed values and observed values is called residual and it is a random variable that must satisfy two requisites:

- normal distribution with null average;
- independence between any two ϵ_i and ϵ_j for each $i \neq j$.

The least squares method

As a measure of the approximation we take

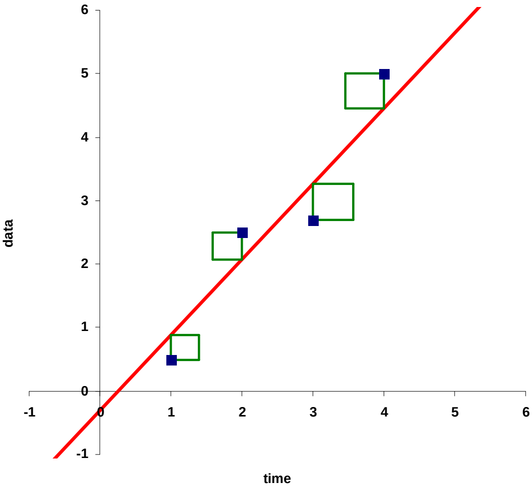
$$Q = \sum_{i=1}^{N} (f(x_i) - y_i)^2$$

and this is the objective function to be minimized.

The unknowns, or decision variables, are the parameters of the line, i.e. A and B.

To find their optimal values, it is sufficient to compute the partial derivatives of Q with respect to A and B and to impose they are equal to 0.

The least squares method



The least squares method

Indicating the average values with

$$\overline{x} = \frac{\sum_{i=1}^{N} x_i}{N} \text{ and } \overline{y} = \frac{\sum_{i=1}^{N} y_i}{N}$$

we have

$$B = \frac{S_{xy}}{S_{xx}} \text{ and } A = \overline{y} - B\overline{x}$$

where

- $S_{xx} = \sum_{i=1}^{N} (x_i - \overline{x})^2$
- $S_{xy} = \sum_{i=1}^{N} (x_i - \overline{x})(y_i - \overline{y})$
- $S_{yy} = \sum_{i=1}^{N} (y_i - \overline{y})^2$

Regression line through the origin

If we want to impose that the prediction line $y = A + Bx$ pass through the origin, then we set $A = 0$ and we estimate only

$$B = \frac{\sum_{i=1}^{N} x_i y_i}{\sum_{i=1}^{N} x_i^2}.$$

Model evaluation

A posteriori, it is very important to evaluate the reliability of the model used, before relying on the forecasts it provides.

- Slope of the line: the model is considered non-significant if a given confidence interval for B contains the value 0.
- Linear correlation coefficient (Pearson index): $r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$.

  It always holds $-1 \leq r \leq 1$.
  If $r > 0$ the line increaes, if $r < 0$ it decreases.
  If $|r| \approx 1$, the linear correlation is strong; if $|r| \approx 0$, it is weak.
- Estimator of the variance: $s^2 = \frac{\sum_{i=1}^{N}(f(x_i)-y_i)^2}{N-2} = \frac{1}{N-2}(S_{yy} - BS_{xy})$.

Time series

A time series is a sequence of values y_t taken by a quantity of interest at given points in time t. If these points in time define a discrete set, the time series is a discrete time series. We consider discrete time

series with points in time uniformly spaced (years, weeks, days,...). A

time series can be seen as a particular realization of a stochastic process and a formal treatment of time series requires concepts from statistics.

Classification

Extrapolative methods can be used to forecast a single period or multiple periods in the future.

- Time series decomposition
- Exponential smoothing
  - Brown model
  - Holt model
  - Winters model
- Autoregressive models:
  - Autoregressive models (AR)
  - Moving average models (MA)

Models of time series

We assume that the observed values y_t be the result of a combination of several components of different nature:

- long period trend, m_t
- long term economic cycles, v_t
- seasonal component, s_t (given a period L)
- random residual, r_t.

We consider two ways in which these components can interact: additive and multiplicative models.

- Additive models: $y_t = m_t + v_t + s_t + r_t$
- Multiplicative models: $y_t = m_t * v_t * s_t * r_t$

In the next slides we will consider a multiplicative model but the same concepts apply to additive ones, just replacing products with sums.

Averaging on a period

If we know the period L of the seasonal component, we can remove the component by computing the average on all time windows of length L.

- L odd: $(mv)_t = \frac{\sum_{i=t-\frac{L-1}{2}}^{t+\frac{L-1}{2}} y_i}{L}$
- L even: $(mv)_t = \frac{\frac{1}{2}y_{t-\frac{L}{2}} + \sum_{i=t-\frac{L}{2}+1}^{t+\frac{L}{2}-1} y_i + \frac{1}{2}y_{t+\frac{L}{2}}}{L}$

To separate the trend component m from the economic cycle component v, we assume the former is linear and we compute it via simple linear regression, where time is the independent variable.

Seasonal indices

The seasonal component and the random component are obtained from $(sr)_t = \frac{y_t}{(mv)_t}$.

Seasonal indices $\overline{s}_1, \ldots, \overline{s}_L$ are obtained as

$$\overline{s}_t = \frac{\sum_k (sr)_{t+kL}}{N_t}$$

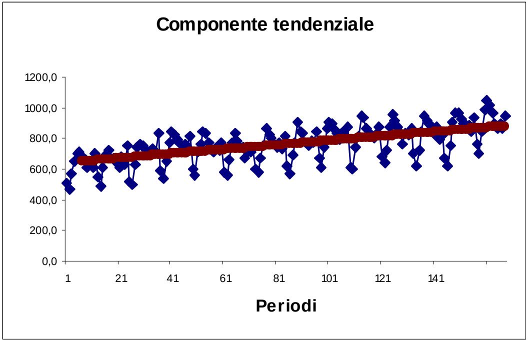
where the extreme values of the range of the sum are suitably chosen to cover all the (sr) values previously computed and N_t indicates the number of terms in the sum.

The indices obtained in this way are then normalized:

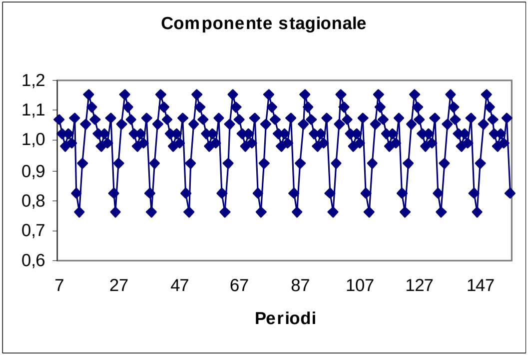
$$s_t = \frac{L\overline{s}_t}{\sum_{t=1}^{L} \overline{s}_t} \quad \forall t = 1, \ldots, L.$$

So we have $s_{t+kL} = s_t$ for each t and for each integer k.

Example: trend component

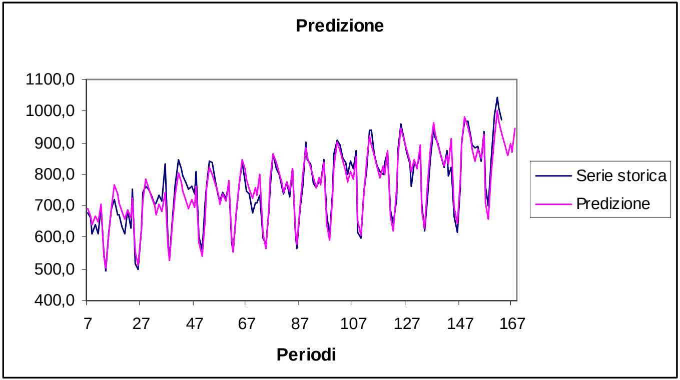


Example: seasonal component



Example: model and forecast

The forecast is done by combining the trend component *m* and the seasonal component *s*.



Exponential smoothing

Exponential smoothing methods are simple, versatile and accurate methods for forecasts based on time series.

There are various models taking into account or not the existence of trend and seasonal components in the time series.

The basic idea is to give more importance to recent observations than to remote ones.

This makes the smoothing methods able to adapt to unknown and sudden variations in the values of the time series owing to events that change the regularity of the observed phenomenon (technical failures, special discounts, bankruptcy of competitors, financial crisis,...).

Brown model

This is the simple exponential smoothing method.

Smoothed average:

- $s_t = \alpha y_t + (1 - \alpha)s_{t-1} \ \ \forall t \geq 2$
- $s_1 = y_1$
- Prediction: $f_{t+1} = s_t$

with $0 \leq \alpha \leq 1$.

For α close to 0 the model is inertial;
for α close to 1 thwe model is reactive.

The optimal value of α is obtained by minimizing the mean square error of the forecasts.

Holt model

This is the exponential smoothing method with trend correction.

Smoothed average:

- $s_t = \alpha y_t + (1-\alpha)(s_{t-1} + m_{t-1}) \ \ \forall t \geq 2$
- $m_t = \beta(s_t - s_{t-1}) + (1-\beta)m_{t-1} \ \ \forall t \geq 2$
- $s_1 = y_1$
- $m_1 = y_2 - y_1$
- Prediction: $f_{t+1} = s_t + m_t$.

with $0 \leq \alpha \leq 1$ and $0 \leq \beta \leq 1$, optimized by minimizing the mean square error.

Winters model

This is the exponential smoothing method with trend and seasonality correction.

Snoothed average:

- $s_t = \alpha \frac{y_t}{q_{t-L}} + (1-\alpha)(s_{t-1} + m_{t-1}) \;\; \forall t \geq 2$
- $m_t = \beta(s_t - s_{t-1}) + (1-\beta)m_{t-1} \;\; \forall t \geq 2$
- $q_t = \gamma \frac{y_t}{s_t} + (1-\gamma)q_{t-L} \;\; \forall t \geq L+1$
- $s_1 = y_1$
- $m_1 = y_2 - y_1$
- $q_t = \frac{y_t}{\sum_{\tau=1}^{L} y_\tau / L} \;\; \forall t = 1, \ldots, L$

with $0 \leq \alpha \leq 1$, $0 \leq \beta \leq 1$ and $0 \leq \gamma \leq 1$, optimized as before.

Prediction: $f_{t+1} = (s_t + m_t)q_{t-L+1}$.

Removing trend and seasonality

To remove the trend component or the seasonality component from a time series:

- compute the moving average to remove the seasonality;
- compute iterative differences $B_t(h) = y_t - y_{t-h}$ to remove the trend;
- identify the trend with regression analysis;
- identify the seasonality by decomposing the series.

Therefore it is possible:

- to use Winters model on the original time series;
- to use Holt model after removing seasonality;
- to use Brown model after removing trend and seasonality.

Autoregressive models

Autoregressive models are based on the assumption that the values in a time series are correlated with the past values.

The autocorrelation of a series is the correlation between its values and the previous ones.

We define autocovariance of order p of a time series Y_t the covariance between the values of Y_t and the values of the same series shifted in time by p:

$$\gamma_p = cov(Y_t, Y_{t-p})$$

Autoregressive models

We define autocorrelation or order p

$$corr(Y_t, Y_{t-p}) = \rho_p = \frac{cov(Y_t, Y_{t-p})}{\sqrt{\sigma^2_{Y_t}\sigma^2_{Y_{t-p}}}}$$

Since $\gamma_0 = \sigma^2_{Y_t}$, we have

$$\rho_p = \frac{\gamma_p}{\gamma_0}.$$

Usually γ_p and ρ_p tend to 0 for $p \to \infty$, because they represent the "memory" or the "persistence" of the underlying (unknown) system that generates the time series.

Stationarity

For an autoregresive model to produce realiable predictions it is required that the time series be stationary, i.e. its mean and variance do not depend on time. Therefore the time series must not contain trend components.

To remove the trend component we can replace the time series $Y = \{y_t\}$ with a series given by the differences between consecutive values, i.e. $Y' = \{y_t - y_{t-1}\}$.

In an autoregressive model of order p ($AR(p)$) we assume

$$y_t = \beta_0 + \beta_1 y_{t-1} + \ldots + \beta_p y_{t-p} + \epsilon_t \quad \forall t$$

where ϵ_t is a (hopefully small) random noise with zero average (white noise).

To make a forecast we must select a proper value for p and we must estimate the parameters β.

Model calibration: selecting p

In the choice of p we must consider the trade-off between the complexity of the model (number of parameters) and its accuracy.

- If p is too small, we loose information carried by the most remote observations.
- If p is too large, the model is unnecessarily complex and it can be affected by noise.

There are several criteria to select p:

- significance test on the parameter with the largest index.
- Bayesian Information Criterion (BIC).
- Akaike Information Criterion (AIC).

Model calibration: estimating β

The most suitable values for parameters β can be found by minimizing the mean square error of the predictions.

The forecast for period $t+1$ is

$$\hat{y}_{t+1} = \beta_0 + \sum_{i=1}^{p} \beta_i y_{t-i}$$

and in the same way we can obtain the forecasts for all next periods.

Moving average models (MA)

In a moving average model of order q ($MA(q)$), we assume

$$y_t = \sum_{i=0}^{q} \theta_i \epsilon_{t-i}$$

where ϵ is a (hopefully small) white noise.

This implies zero average for y. Therefore, before interpreting a time series with a *MA* model, we must remove its trend (by differentiation) and also its average value (by subtracting it from the series).

To make a forecast, we must select a proper value for q and we must estimate the parameters θ, as before.

ARMA models

In an $ARMA(p, q)$ model, we sum an $AR(p)$ and a $MA(q)$ model.

SARMA models (where *S* stands for *seasonal*) are used for time series with seasonalities.

A complete and rigorous treatment of *ARMA* models requires the study of stochastic processes, which is a branch of statistics (and does not fit into this course).