Forecasting & Predictive Analytics

Guillaume Chevillon
chevillon@essec.edu

ESSEC Business School

October-December 2017
1st set of slides
Forecasing Principles

Aims of the Course

- Provide a comprehensive understanding of forecasting & predictive analytics principles and methods
- Provide tools to assess uncertainty and evaluate forecasts in R
- 3 textbooks:

"Forecasting: Principles and Practice" , Rob J. Hyndman and George Athanasopoulos
"Economic Forecasting" by Graham Elliott & Allan Timmermann
"Practical Business Forecasting" by Michael K. Evans

Forecasting Principles

Definition (Forecasting)

Forecasting is about predicting the future as accurately as possible, given all of the information available, including historical data and knowledge of any future events that might impact the forecasts.

It is different from

- "Goals": what you would like to happen,
- "Planning": appropriate actions that are required to make your forecasts match your goals

Definition (Predictive Analytics – Wikipedia)

Predictive analytics is an area of data mining that deals with extracting information from data and using it to predict trends and behavior patterns. Often the unknown event of interest is in the future, but predictive analytics can be applied to any type of unknown whether it be in the past, present or future.

Characteristics

1. What is forecasting ?
   - Predict the Future using (time series related or other) data we have in hand
   - Something interesting because if affects the decisions we make today
2. Characteristics of good Forecasts
   - It should be timely, as accurate as possible and/or reliable
   - The method should be easy to use and understand in most cases
3. Other characteristics
   - A good forecast is more than a single number :
     - Includes a mean value and standard deviation (at least)
     - Includes accuracy range (high and low)
   - Aggregated forecasts are usually more accurate
   - Accuracy erodes as we go further into the future.

What can be forecast(ed)

Forecasting is needed

- In many situations:
  - Forecast future demand to decide whether to build another power generation plant in the next five years
  - Forecast call volumes to schedule staff in a call center each week ...
- For different time horizons, depending on specific applications:
  - Short-term demand forecasts to schedule personnel, production and transportation
  - Medium-term forecasts to determine future resource requirements, purchase raw materials, buy machinery...
  - Long-term forecasts that are used in strategic planning and take into account market opportunities and environnemental factors.

Predictability

- Forecasting is easier in some cases than in others (time of sunrise tomorrow versus tomorrow's lottery numbers)
- The predictability of an event or quantity depends on:
  - How well we understand the factors that contribute to it
  - How much data are avalaible (how can we forecast the sales of a new product?)
  - Whether the forecasts can affect the thing we are trying to forecast (think about the "efficient market hypothesis")
- Unfortunately, some things cannot be forecasted accurately, that is not better than tossing a coin or making naïve forecasts (using the most recent observation as a forecast).

Issues

1. Identifying the forecasting problem: what for? who? when? how?
2. Gathering information: what are the avalaible data? are they reliable?
3. Preliminary exploratory analysis: graphs to see trends, seasonality, breaks, outliers, volatility...
4. Selecting appropriate methods (often 2 or 3) and compare them
5. Evaluating and refining forecasting methods over time.
6. Forecasting (Confidence) Intervals

Examples:

Bank of England, Inflation Report, Nov 2016

Chart 5.1 GDP projection based on market interest rate expectations, other policy measures as announced

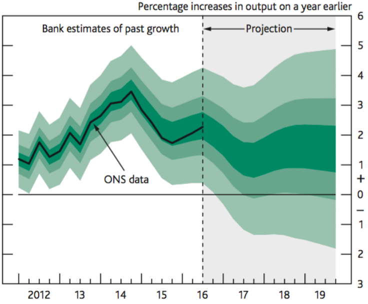
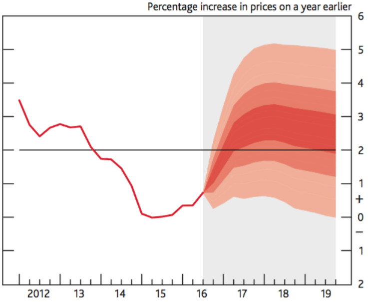


Chart 5.2 CPI inflation projection based on market interest rate expectations, other policy measures as announced



Course Overview

- Forecasting (Predictive Analytics) Principles and Techniques
- Assessment
  - Final Quiz (1/3 of overall assessment)
  - Continuous Assessment (2/3)

Topics

- Forecasting Principles
- Univariate Forecasting Modeling
- Evaluation and Combination
- Vector & Structural Models
- Data Rich environments
- Instability & Robustness
- Nonlinear Forecasting
- Recommendation systems

Outline

Some of Today's words

- Visualizing time series;
- Time series components;
- Loss functions
- Forecasting vs. explanation;
- Performance evaluation and backtesting;

1- Identifying the forecasting problem

In the early stages of a forecasting project, it is absolutely necessary to determine what to forecast:

1. every product line or groups of products?
2. every sales outlet, or for outlets grouped by region, or the total sales?
3. weekly data, monthly data or annual data?
   - at which time horizon?
   - at which frequency?
4. with which explanatory data?
   - are they forecasted?

2- Gathering information

Do we have:

- Cross sectional data: collected at a single point in time
  - that will lead to models of Predictive Analytics
- Time series data: collected at regular intervals over time that will lead to
  - this is more specific to Forecasting problems
- And are these data reliable ?

2- Gathering information

- Can you simplify easily the forecasting problem ?
- Adjusting the historical data can lead to simpler forecasting models, by removing known sources of variation or making the pattern more consistent across the whole data set.
- Mathematical transformations: if the volatility is higher when the level is higher, you could use a log-transformation. The log transformation is useful when one wants to model a quantity that should stay positive
- Calendar Adjustments for data that are affected by calendar effects (subway traffic...)
- Population Adjustments for data that are affected by population changes (GDP, number of hospital beds...)
- Inflation Adjustments for data that are affected by the value of money (house price, oil price, ...) with price indexes

3- Preliminary exploratory analysis

- A pre-requisite: the graph
  - Before modeling, it is essential to graph data, in order to detect: Missing data or outliers
  - Links between series
  - The nonstationarity of time series:
  - Trend (non constant mean over time) Non-constant volatility over time Breaks presence on level or volatility
  - The presence of seasonality

... So many elements that must be taken into account in the modeling.
The type of data determines what forecasting method to use but also the appropriate graphs.

Definitions

Some definitions for preliminary analysis for time series

Definition (Trend)

A trend exists when there is a long-term increase or decrease in the data. It does not have to be linear and can be ”changing direction”.

Definition (Cycle)

A cycle occurs when the data exhibit rises and falls that are not on a fixed period, due to economic conditions (usually at least 2 years).

Definition (Seasonal Pattern)

A seasonal pattern occurs when a time series is affected by seasonal factors such as the time of the year or the day of the week. Seasonality is always of a fixed and known period.

4- Choosing, Fitting & Assessing the model(s)

- Loss functions and expected Loss
- Plug-In estimator vs Loss Based estimator
- Parametric vs non parametric
- Cross-validation (Backtesting)
- Residuals
- Combination & Encompassing

Forecast Construction

- Model must be *congruent*: i.e. valid whichever way it is assessed.
- Various issues (see Clements & Hendry, 2005)
  1. unconditional vs. conditional models;
  2. internal vs. external standards;
  3. checking constancy vs. adventitious significance;
  4. ex ante vs. ex post evaluation;
  5. one-step vs. multi-horizon forecasts;
  6. in-sample fixed coefficients vs. continuous updating;
  7. whether the process is stationary or non-stationary; and
  8. whether the final objective is forecasting or not (e.g. policy).

Forecast evaluation

- Comparison in-sample fit versus out-of-sample forecasting performance
- Mincer-Zarnowitz regressions(forecast rationality/efficiency):

Test $(\beta_1, \beta_2) = (0, 1)$ in

$$y_t = \beta_1 + \beta_2 \widehat{y}_{t|t-h} + \nu_t$$

and in the generalized version, add predetermined regressors (known at $t-h$) and test that their coefficients are zero.

Forecast Comparison

- Diebold-Mariano statistic:

To compare the forecasts from two models A and B. Define

$$d_t = \left(y_t - \widehat{y}^A_{t|t-h}\right) - \left(y_t - \widehat{y}^B_{t|t-h}\right)$$

then, letting $\mathsf{V}[d_t]$ the long run variance of d_t (estimated using Newey-West kernel estimator), estimate the models over subsample $\mathcal{S}_R$ and evaluate over $\mathcal{S}_F$. The statistic for $\mathsf{H}_0 : \mathsf{E}\left[d_t\right] = 0$ is

$$DM = \frac{\overline{d_{t\in\mathcal{S}_F}}}{\sqrt{\mathsf{V}\left[d_t\right]}} \underset{\mathsf{H}_0}{\sim} \mathsf{N}\left(0, 1\right)$$

where large negative values imply that A outperforms B (and conversely). CAREFUL WHEN THE MODELS ARE NESTED

- Forecast encompassing test: can $\left(1-\alpha\right)\widehat{y}^A_{t|t-h} + \alpha\widehat{y}^B_{t|t-h}$ be better than $\widehat{y}^A_{t|t-h}$? i.e. test $\alpha = 0$ in

$$y_t - \widehat{y}^A_{t|t-h} = \alpha\left(\widehat{y}^B_{t|t-h} - \widehat{y}^A_{t|t-h}\right) + \eta_t$$