Statistics for Decision Making: Broad Introduction

Basics for Forecasting or Prediction

Le Wang

2018-08-27

Forecasting or Prediction

Mindful of the goals, we can figure out what type of forecats/predictions that we are interested in

1. **Machine Learning** Forecasting an outcome of interest (irrespective of the underlying true process)
2. **Econometrics** Causal Inference (e.g., Predict the consequences of a strategic move)

Forecasting or Prediction

Mindful of the goals, we can figure out what type of forecats/predictions that we are interested in

1. Forecasting an outcome of interest (irrespective of the underlying true process)

Example: a company's hiring decisions depend on its predictions about how much demand there would be and how much the company needs to produce in order to meet such demand.

Forecasting or Prediction

Mindful of the goals, we can figure out what type of forecats/predictions that we are interested in

2. Causal Inference (e.g., Predict the consequences of a strategic move)

Example: before taking this course or program, many of you may need to form your own expectations of the future returns to taking it whether and how much this particular program or course could help to increase your job prospect or the chances of getting into a Ph.D program, or deepen your understanding of econometrics for future research.

Forecasting or Prediction

Types of Forecasts:

Qualitative vs. **Quantitative**

The former tells us only the *direction* of the future, while the later is concerned with the actual amount of changes.

Forecasting or Prediction

How do we form such quantitative forecasts?

1. **Scientific**: Rigorous statistical and mathematical methods to discover the patterns in historical data.
2. **Art**: Requires a lot of personal judgement as well.

e.g., whether the assumptions are plausible and realistic in practice and how the results should be interpreted.

Types of Quantitative Forecasts

1. By Data types
   1.1 Cross-sectional Data
   1.2 Time Series Data
   1.3 Panel Data
   1.4 Hierarchical Time Series
2. By Variables types
   2.1 Continuous Variables
   2.2 Qualitative Discrete Variables
   2.3 Quantitative Discrete Variables
   2.4 Mixed Type Variables
3. By Information
   3.1 Point Forecast
   3.2 Interval Forecast
   3.3 Density Forecast

Types of Quantitative Forecasts (by data types)

1. **Cross-sectional data** the information is observed only at time period, t, but for many different economic agents (e.g., individual consumers, countries, and firms). We predict what may happen to some units that we do have information for.

Example: Housing Prices When you are purchasing (or selling) a house, the (final) price is generally not observable to us because it has not even happened yet. Thus, we generally collect information on other houses sold during the same period (week, month or year) such as number of rooms and baths and their final prices. Using the information, we can predict what the price will be for your house.

Types of Quantitative Forecasts (by data types)

2. **Time Series data** the information is observed for only one agent, but for many time periods.

Example Daily IBM stock prices or monthly rainfall for Norman, Oklahoma.

Types of Quantitative Forecasts (by data types)

3. **Panel Data** a combination of both cross-section and time-series. For this type of data, the information is observed for more than one economic agents and for many time periods.

Example: Government expenditures for the 50 states during the period 1900-2014.

Types of Quantitative Forecasts (by data types)

4. **Hierarchical time series**

Example: We collect information on national economic account, which can be divided into three categories: income/outlay, production, capital transaction. Within the category of income/outlay, we can further divide it into government, persons, and companies.

This kind of hierarchical structure is specifically taken into account in modelling.

Types of Quantitative Forecasts (by variable types)

Textbook

1. Continuous variables: assumes an infinite number of values within a given range (uncountably infinite number of values).
2. Discrete varaibles: assumes only certain clearly separated values (finite or countably infinite number of distinct values).

Not very accurate, nort are they mathematically precise. We will define them more formally later but look at some examples.

Types of Quantitative Forecasts (by variable types)

1. **Continuous Variables:** Amount of income tax paid, weight of a student, age of a student, or annual rainfall in Norman.
2. **Qualitative Discrete Variables** Gender (male vs female), Marital status (married with a spouse present, married without a spouse present, divorce, widow, single and never married), party affiliation (democrat, republican, independent).
3. **Quantitative Discrete Variables** Number of children in a household, number of tornadoes in Norman!!!!!!
4. **Mixed Type** Wages with a mass of zeros but infinite possible values greater than zero.

In practice, we will treat qualitative variables just as quantitative variables by assigning numeric values to each category.

Types of Quantitative Forecasts (by information types)

Description	Technical Terms	Notation
Object to Analyze	Time Series	$\{y_t\}, t = 1 \dots T$
Value at Present Time t	Known Value of the Series	y_t
Future at time $t+h$	**Random Variable**	Y_{t+h}
Value at Future Time $t+h$	Unknown value of the Randome Variable	y_{t+h}
Collection of Information	**Information Set (Univariate)** **Information Set (Multivariate)**	$I_t = \{y_1, \dots, y_t\}$ $I_t = \{y_1, \dots, y_t \; x_1, \dots, x_t\}$
Final Objective	**Forecasts: or** 1-step-ahead h-step-ahead forecasts	or $\widehat{y}_{t+1\|t}$ $\widehat{y}_{t+h\|t}$
Uncertainty	**Forecast Errors**	$e_{t+h} = y_{t+h} - \widehat{y}_{t+h\|t}$

Types of Quantitative Forecasts (by information types)

1. If what we are interested in is time series data, a collection (or set) of realizations of the subject of interest, denoted by $\{y_t\}$ for time periods $1 \dots T$.
2. We also have a collection of information available to us at time period t, including all the past information on the outcome itself, y, and other potential predictors, x. The information set is denoted by I_t. For a univariate time series, the information set contains all information on past outcomes, $I_t = \{y_1, \dots, y_t\}$. For multivariate time series, the information contains all information on past outcomes and other predictors, $I_t = \{y_1, \dots, y_t\ x_1, \dots, x_t\}$. Note that each element of $x_1, \dots, x_t$ can be a vector.
3. Based on the information available at time period t, our task is to come up with a forecast for the object of interest in the future at time $t + h$, Y_{t+h}. This is a **random variable**.

Types of Quantitative Forecasts (by information types)

4. Our forecasts are denoted by $y_{t+h|t}$, where h is **forecast horizon** or **the step**. The step, h, is simply how many periods ahead that we would like to forecast.
   It could be one minute, day, or month, depending on the frequency of the time series. The subscript summarizes two things: conditional on information available at time period t, we are trying to forecast h-step(period)-ahead.
5. For the forecasts that we make, there will be some **forecast errors**, $e_{t+h} = y_{t+h} - \widehat{y}_{t+h|t}$, the difference between the true value and our forecasts.

Types of Quantiative Forecasts (by information types)

Our forecasts can also be categorized into three groups:

1. **Point Forecast:** This is just a single value. For example, we may say that our 1-year-ahead forecast for the inflation rate will be 2%.
2. **Interval Forecast:** This is a range of possible values. For example, we may say that our 1-year-ahead forecast for the inflation rate will be between 1% and 3% with a 90% confidence.
3. **Density Forecast:** This gives a complete density function for the future. This would allow us to construct forecasts like $\Pr[Y_{t+2} \geq 100\%] = 0.01$, that is, the probability of large inflation of 100% occurring in two years is very unlikely, about one percent. \end{enumerate}

What will determine the accuracy or performance of our forecasts?

Question What do you think?

Which one can we forecast better, Electricity demand vs Exchange Rates (or Stock Prices)? Why?

What will determine the accuracy or performance of our forecasts?

1. How well we understand the factors that determine the outcome? This is related to the specification of our models; in other words, what goes into our model (i.e., potential determinants) and how it goes into our model (i.e., restrictions on functional forms).

What will determine the accuracy or performance of our forecasts?

2. How much information is available?
   2.1 Not only the number of data points, but also the independent information that each data point carries. This will affect the precision of our estimates and hence forecasts.
   2.2 Sampling. By sampling, we mean we want to try our best to collect data of larger size and better quality. But we also recognize that sometimes the data are a result of lack of effort, but just of reality and individual decisions

Example: 1. Fighter Jet 2. Estimation of returns to education among women 3. Death Penalty Survey in Jail

What will determine the accuracy or performance of our forecasts?

3. Whether the forecasts can affect the thing we are trying to forecast. This is a more subtle point

What will determine the accuracy or performance of our forecasts?

4. Cross-sectional analysis: Partial equilibrium vs General equilibrium effects

Violation of an oft-assumed condition: i.i.d.

What will determine the accuracy or performance of our forecasts?

4. Cross-sectional analysis: Partial equilibrium vs General equilibrium effects

Example:

1. The effects of vacine against a contagious disease. (underestimate)
2. The effects of education on life-time earnings.
3. Large-scale Job training programs funded by tax (which could also decrease the demand)