Lecture 9: Linear Models

Dr. Roman V Belavkin

BIS3226

Contents

1 Introduction to Modelling

What is a Model?

- The word '*model*' comes from a Latin word meaning '*small*', and usually we mean some small representation of the real object (e.g. a model of a house, a car or an aircraft)

$$\text{Reality} > \text{Model}$$

- Some models can be larger in size (e.g. a model of an atom). We shall mean 'smaller' in terms of complexity or uncertainty

$$\text{Uncertainty(Reality)} > \text{Uncertainty(Model)}$$

- The remaining part is called the *error* of a model. Thus

$$\text{Reality} = \text{Model} + \text{Error}$$

- Which model is *optimal*?
- We prefer models that have smaller error:

$$\text{Error} \longrightarrow 0 \quad \text{means} \quad \text{Model} \longrightarrow \text{Reality}$$

Table 1: GBP/EUR rates 4–8 Jan, 2010

Date	Today	Tomorrow
2010/01/04	0.89513	0.89966
2010/01/05	0.89966	0.89934
2010/01/06	0.89934	0.89963
2010/01/07	0.89963	0.89771
2010/01/08	0.89771	?

Functions as Mathematical Models

- In mathematics, we represent concepts and objects in reality by elements of some sets (or variables), such as $x \in X$, $y \in Y$.
- Then we try to model (approximate) the dependency between the variables by mathematical functions $f : X \to Y$:

$$y = f(x) + \text{Error}$$

Example 1. Consider a moving object (e.g. a car). We can use variable $x \in [0, +\infty)$ to represent its speed, and $y \in [0, +\infty)$ to represent the distance it travels. In 2 hours, the car travels

$$y \approx f(x) = x \cdot 2$$

In fact, time is another variable $z \in [0, +\infty)$, and so $y \approx f(x, z) = x \cdot z$.

Data-Driven Models

- Data is a 'footprint' of reality.
- Does the credit score depend on a person's income (recall $P(x, y) \neq P(x)P(y)$).
- Can we find a function $f(\cdot)$ such that

$$\text{Tomorrow} \approx f(\text{Today})$$

- Data-driven modelling is a search for such functions that represent the dependencies between different variables.

2 Linear Functions of One Variable

Linear Functions

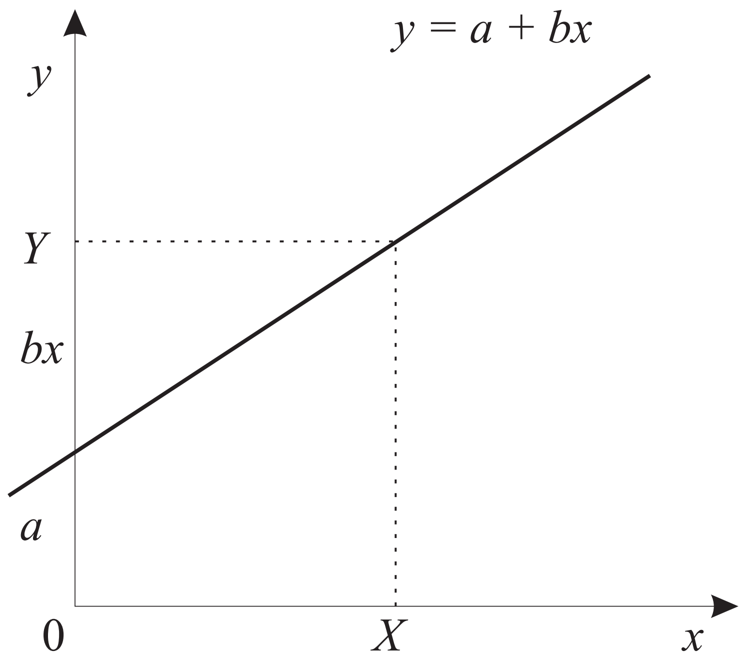
- Linear functions have the properties:

$$f(a+b) = f(a) + f(b)\,, \quad f(\lambda \cdot x) = \lambda \cdot f(x)$$

- The simplest is the linear function of one variable $x \in \mathbb{R}$:

$$f(x) = a + b\,x$$

  where a is the **intercept**, and b is called the **slope**.



Example 2.

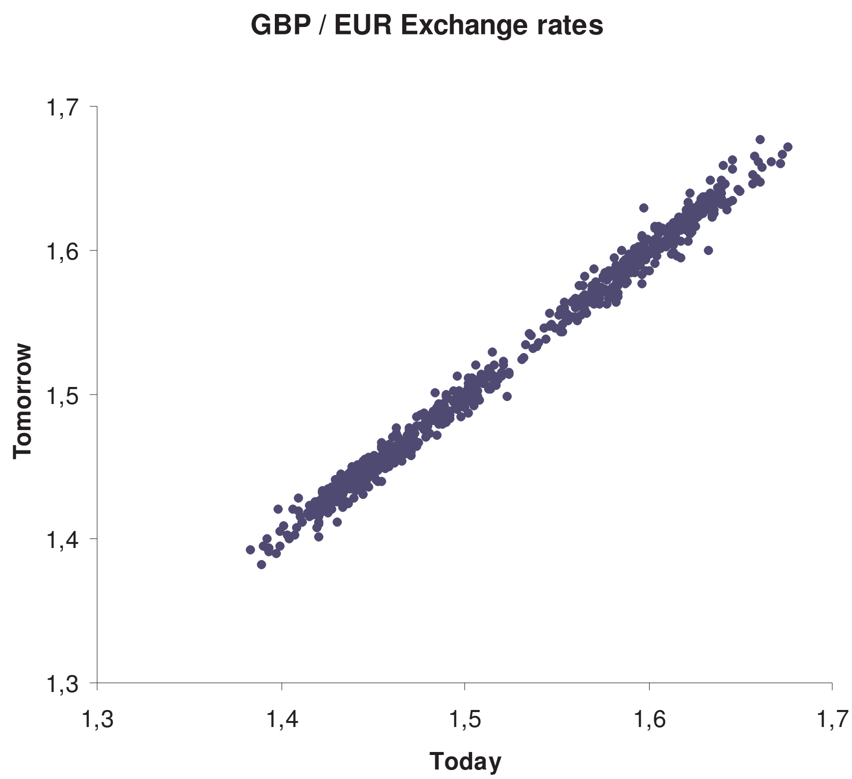
$$y = 2 + 3\,x \qquad (a = 2\,,\ b = 3)$$

What value is y when $x = 4$?

This function defines a line on the xy-plane.

Linear Forecast

- A forecast can be made using a function relating past to the future.
- In the coursework, we can denote the exchange rate of today by x and the rate of tomorrow by y.



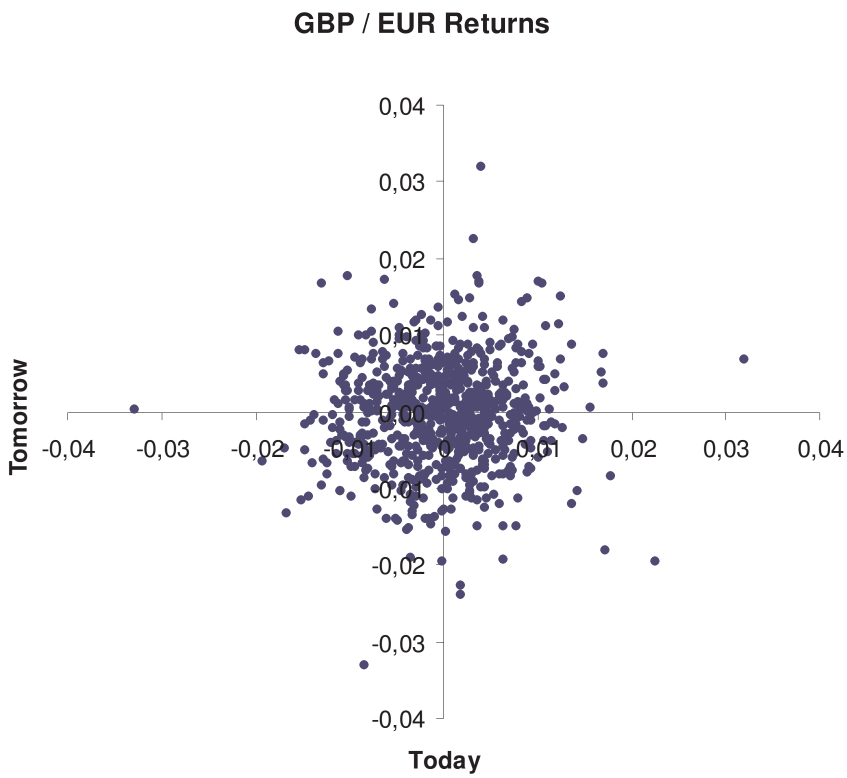
- Does the relation $y = f(x)$ look linear?
- Can you suggest values for the intercept (a) and the slope (b)?

$$y \approx f(x) = a + b\,x$$

Linear Forecast (cont.)

- It would be more useful to predict the *differences* between two consequitive values, or the *returns* Δy

$$\Delta y_i = y_i - y_{i-1}$$



- Does this relation look linear?
- Do you think from the chart that the return of tomorrow depend on the return of today?

3 Choosing a Model

Fitting a Data-Driven Model to Data

- Together, a and b are the *parameters* of a linear model, because their values define the line.
- Different values of $a \in \mathbb{R}$ and $b \in \mathbb{R}$ correspond to different lines (models).
- To select the best model means we need to find the best (optimal) values of a and b which make the line as close as possible to data.
- More generally, the optimal model should minimise the average error:

$$y = f(x) + \text{Error}$$

- To measure the error, we define the *cost function*:

$$c(y, f(x)) = \text{Error}$$

- The choice of $c(y, f(x))$ determines, which model is the best.

Cost Functions

Example 3 (Boolean).

$$c(y, f(x)) = \begin{cases} 0 & \text{if } y = f(x) \\ 1 & \text{otherwise} \end{cases}$$

Example 4 (Absolute deviation).

$$c(y, f(x)) = |y - f(x)|$$

Example 5 (Squared deviation).

$$c(y, f(x)) = |y - f(x)|^2$$

Remark 1. *It is known that for Boolean cost, the optimal model is the maximum likelihood (probability)* $\max P(y \mid x)$*; for squared deviation the optimal model is the conditional expectation* $y \approx E\{y \mid x\}$*.*

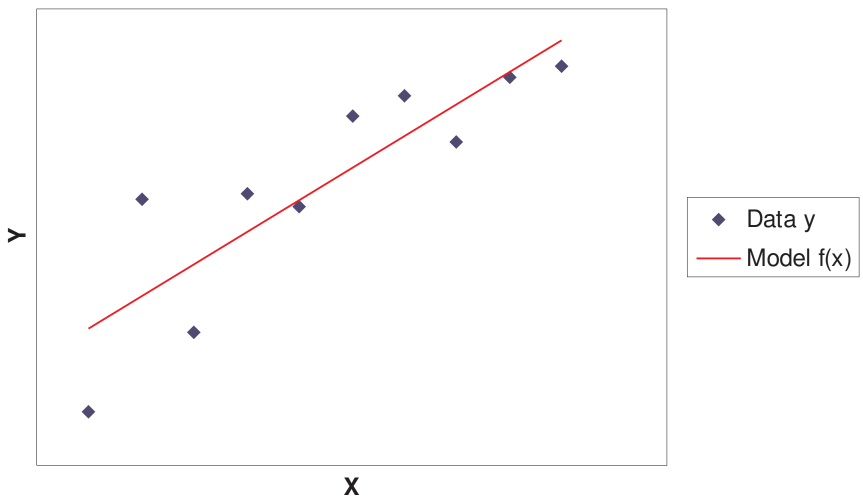
4 Mean-Square Linear Regression

Mean-Square Linear Regression

- We use squared deviation $c(y, f(x)) = |y - f(x)|^2$ as the cost function.
- The conditional expectation is approximated by a linear function

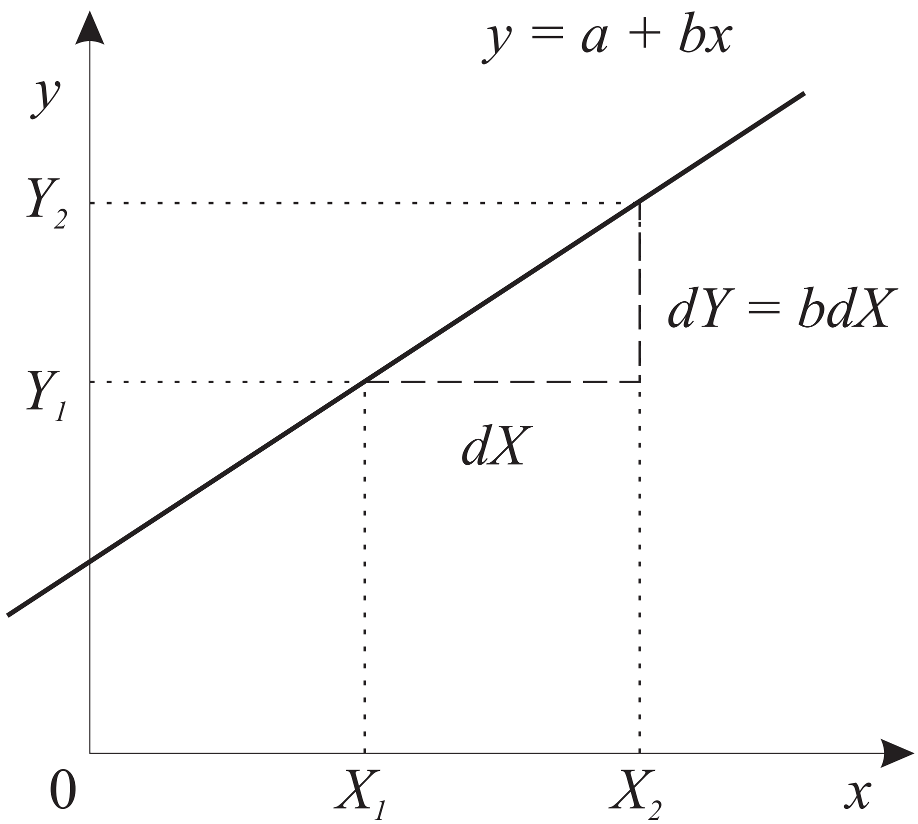
$$y \approx E\{y \mid x\} = a + b\,x$$

- Thus, we only need to find the intercept (a) and the slope (b) of a line that minimises the squared deviation on average.



A Line Through Two Points

- Given two points (X_1, Y_1) and (X_2, Y_2), find a and b.
- Two points uniquely define the line.



$$b = \frac{dY}{dX} = \frac{Y_2 - Y_1}{X_2 - X_1}$$

$$a = Y_1 - b\,X_1$$

We can also write the equation for the line without a:

$$f(x) = Y_1 + b\,(x - X_1)$$

A Line Through Several Points

- For several points, the slope (b) is computed using the measures of dispersion (covariance and variance)
- For the intercept (a), we use the fact that the line must go through the centre of gravity $(E\{x\}, E\{y\})$.

x	y
X_1	Y_1
X_2	Y_2
$\vdots$	$\vdots$
X_n	Y_n

$$b = \frac{Cov(x,y)}{Var(x)}$$

$$a = E\{y\} - b\,E\{x\}$$

And the equation for the line is

$$y \approx E\{y\} + b\,[x - E\{x\}]$$

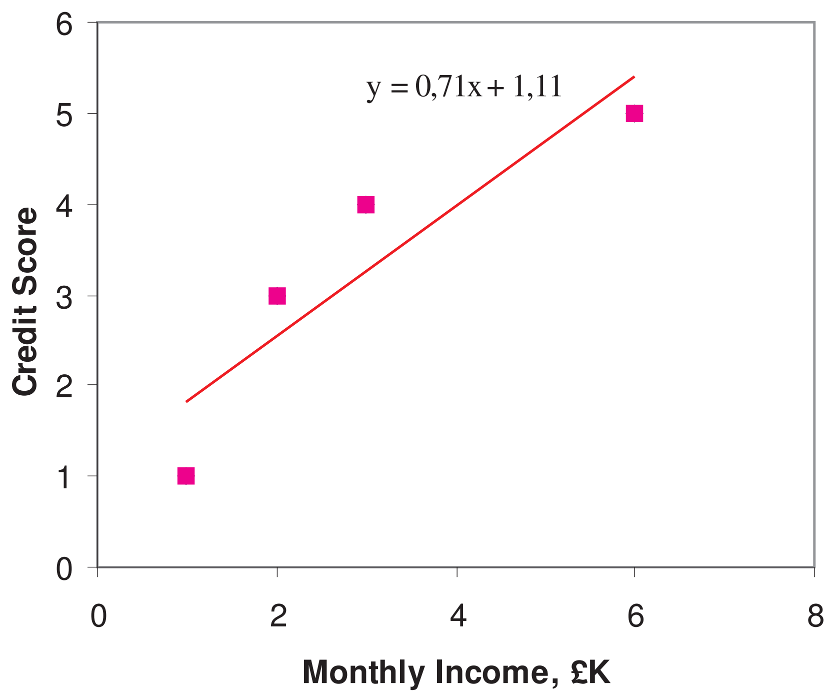
A Simple Model for Credit Score

Monthly Income (£ K)	Credit Score
2	3
1	1
6	5
3	4

- Denote by x the income and by y the credit score.
- Construct a linear model $y \approx a + b\,x$
- We need to find slope (b) and intercept (a) from the data

Solution

$$\begin{aligned}
E\{x\} &= (2+1+6+3)/4 = 3 \\
E\{y\} &= (3+1+5+4)/4 = 3,25 \\
Cov(x,y) &= [(2-3)(3-3,25) + \cdots + (3-3)(4-3,25)]/4 = 2,5 \\
Var(x) &= [(2-3)^2 + \cdots + (3-3)^2]/4 = 3,5
\end{aligned}$$



$$
\begin{aligned}
{\color{blue}b} &= \frac{Cov(x,y)}{Var(x)} = 0,71 \\
{\color{red}a} &= E\{y\} - {\color{blue}b}\,E\{x\} = 1,11
\end{aligned}
$$

Summary

- Models are simplified representations of reality
- The unexplained part of reality results in an error of the model
- Linear functions are the simplest mathematical models, which can be used for data-driven modelling
- Linear mean-square regression is a standard method of computing such models
- Linear functions model linear dependencies in data