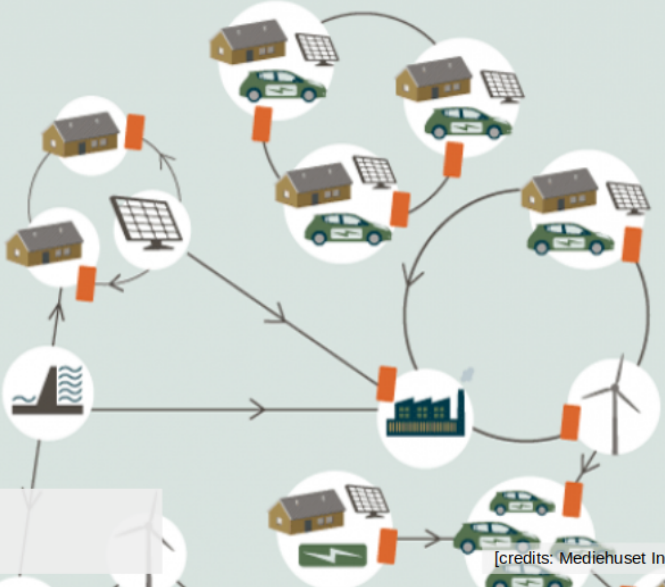
Module 9 – Renewable Energy Forecasting: First Steps

9.2 Going further in a regression framework

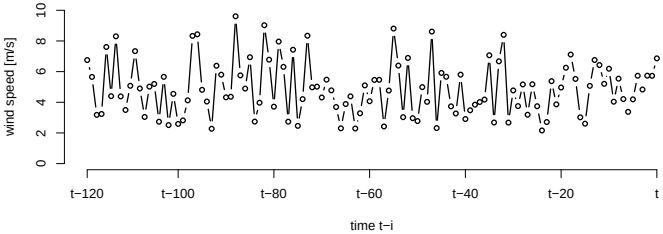


Pierre Pinson
Technical University of Denmark

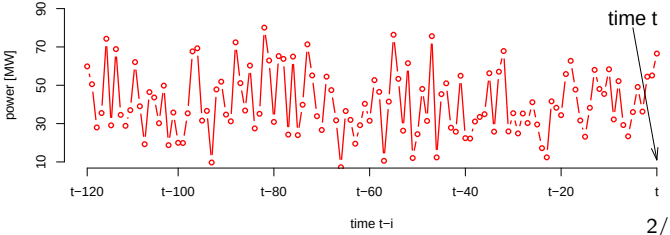
[credits: Mediehuset Ingeniøren]

What is (linear) regression?

- In the simplest case, data is available for:
  - y_i $(i = t - n, \ldots, t)$, the **response** variable, i.e., the variable we will want to predict, eventually
  - x_i $(i = t - n, \ldots, t)$, an **explanatory** variable, i.e., a variable that can help us predict y
- At this stage, imagine that x_i and y_i are your most recent wind speed and corresponding power observations up to current time t



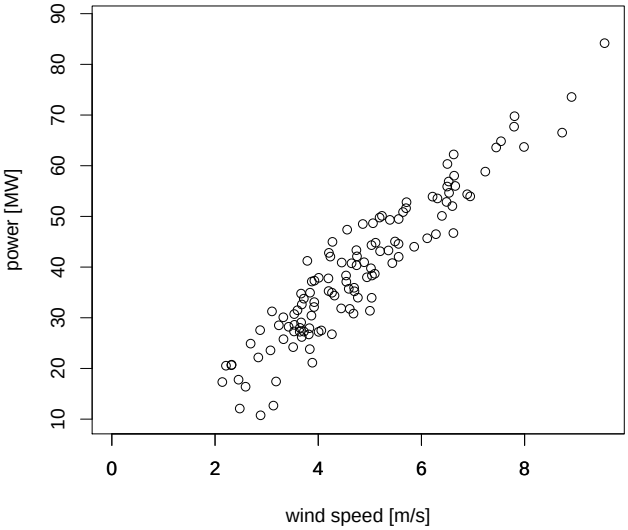
- Example set with the last $n = 120$ observations of
  - wind speed x_i, and
  - corresponding power generation y_i



What is (linear) regression? (continued)

- The aim is to uncovering some relationship between these **explanatory** and **response** variables
- We first do that visually...

- Same example, with the last $n = 120$ observations of
  - wind speed x_i, and
  - corresponding power generation y_i
- In this *scatterplot*, there seems to be a (linear) relationship between wind speed and power



What is (linear) regression? (continued)

- Such a linear relationship between x and y can be written as

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = t - n, \ldots, t$$

where

  - β_0 and β_1 are the model parameters (called *intercept* and *slope*)
  - ε_i is a noise term, which you may see as our forecast error we want to minimize

The linear regression model can be reformulated in a more compact form as

$$y_i = \boldsymbol{\beta}^\top \mathbf{x}_i + \varepsilon_i, \quad i = t - n, \ldots, t$$

with

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \qquad \mathbf{x}_i = \begin{bmatrix} 1 \\ x_i \end{bmatrix}$$

- It is often easier to deal with such compact formulations...

Least Squares (LS) estimation

- Now we need to find the best value of $\boldsymbol{\beta}$ that describes this cloud of point
- Under a number of assumptions, which we overlook here, the (best) model parameters $\hat{\boldsymbol{\beta}}$ can be readily obtained with **Least-Squares (LS) estimation**

The **Least-Squares (LS) estimate** $\hat{\boldsymbol{\beta}}$ of the linear regression model parameters is given by

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \sum_i \varepsilon_i{}^2 = \arg\min_{\boldsymbol{\beta}} \sum_i \left(y_i - \boldsymbol{\beta}^\top \mathbf{x}_i\right)^2 = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$$

with

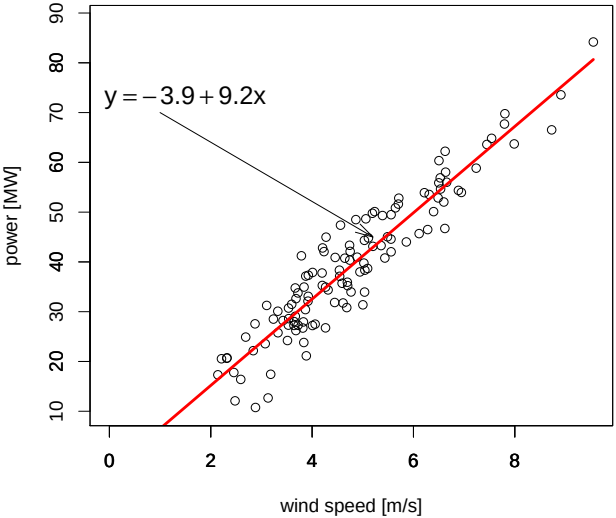
$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix}, \qquad \mathbf{X} = \begin{bmatrix} 1 & x_{t-n} \\ 1 & x_{t-n+1} \\ \vdots & \vdots \\ 1 & x_t \end{bmatrix}, \qquad \mathbf{y} = \begin{bmatrix} y_{t-n} \\ y_{t-n+1} \\ \vdots \\ y_t \end{bmatrix}$$

- Even better: some functions in R/Matlab can do it for you!

The resulting (linear) regression

- For the same example set with the last $n = 120$ observations of
  - wind speed x_i, and
  - corresponding power generation y_i
- The LS-estimate of the model parameters is:

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} -3.9 \\ 9.2 \end{bmatrix},$$



- This type of model and estimation can then be *incorporated within in a forecasting approach*

Forecasting in a (linear) regression framework

- At a given time t, you (as a forecaster) identified a good model:

$$y_i = \beta_0 + \sum_{j=1}^{m} \beta_j x_{j,i} + \varepsilon_i, \quad i = t-n, \ldots, t$$

(or, equivalently: $\quad y_i = \boldsymbol{\beta}^\top \mathbf{x}_i + \varepsilon_i$)

where

  - y_i is still your response variable (say, wind power generation) observed at time i
  - $x_{j,i}$ is the observation at time i for the jth explanatory variable ($j = 1, \ldots, m$)
  - β_j is the model parameter for the jth explanatory variable
  - ε_i is a noise term, which you may see as our forecast error we want to minimize

- Based on the last n observations, you obtain an **LS-estimate** $\hat{\boldsymbol{\beta}}_t$, valid at time t
- And you can issue forecast using these estimates $\hat{\boldsymbol{\beta}}_t$, for any new values of the explanatory variables, i.e.

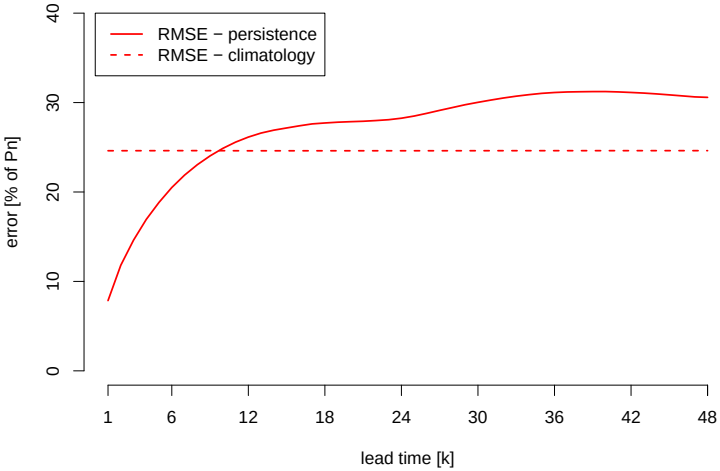
$$\hat{y}_{t+k|t} = \boldsymbol{\beta}_t^\top \mathbf{x}_{t+k}$$

- **Potential problem here: we do not know future values of the x variable** (e.g., wind speed)!

Example application: combining persistence and climatology

- *Persistence* and *climatology* were shown to be good benchmarks (difficult to outperform), though
  - persistence is good for **short lead times**
  - climatology is good for **longer lead times**
- *Why no combining them, as function of the lead time k?*

- Reminder of the quality of the *persistence* and *climatology* forecasts,
  - in terms of RMSE
  - as a function of the lead time *k*



Proposal of a combination model and estimation

- Our proposal model has the following form, for a given lead time k,

$$y_i = \beta_{k,\text{pers}} \hat{y}^{(\text{p})}_{i|i-k} + \beta_{k,\text{clim}} \hat{y}^{(\text{c})}_{i|i-k} + \varepsilon_i, \quad i = t-n, \ldots, t$$

  where

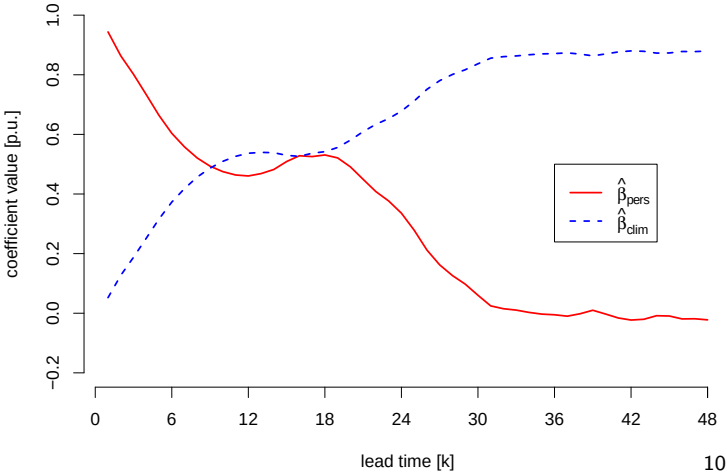
  - $\hat{y}^{(\text{p})}_{i|i-k}$ and $\hat{y}^{(\text{c})}_{i|i-k}$ are the persistence and climatology forecasts, issued at time $i-k$ for time i
  - $\beta_{k,\text{pers}}$ and $\beta_{k,\text{clim}}$ are *intercept* and the weights to be given to persistence and climatology forecasts, respectively
  - ε_i is a noise term, which you may see as our forecast error we want to minimize

- It therefore combines persistence and climatology forecasts
- The weight given to each of these forecasts can change with the lead time k

Estimation of the model coefficients

- For the example of the 28 April 2002 (as in first slides),
  - the necessary vectors and matrices are formed, with $n = 200$ last values
  - LS estimates $\hat{\beta}_{k,\text{pers}}$ and $\hat{\beta}_{k,\text{clim}}$ are computed for every lead time k ($k = 1, \ldots, 48$)

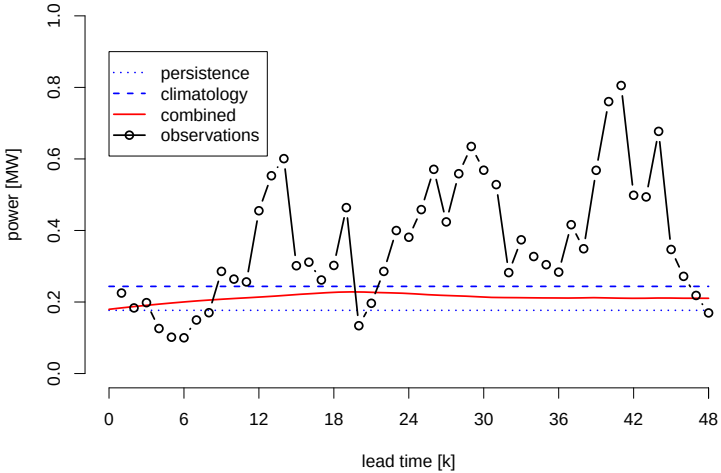
- *Right:* Evolution of the estimated model parameters $\hat{\beta}_{k,\text{pers}}$ and $\hat{\beta}_{k,\text{clim}}$ as a function of the lead time k
- persistence is given less weight for further lead times
- climatology is given more weight instead



The resulting forecast

- For the example of the 28 April 2002 (as in first slides),
  - LS estimates $\hat{\beta}_{k,\text{pers}}$ and $\hat{\beta}_{k,\text{clim}}$ are used to combine the available *persistence* and *climatology* forecasts
  - The combination is different for every lead time k ($k = 1, \ldots, 48$)

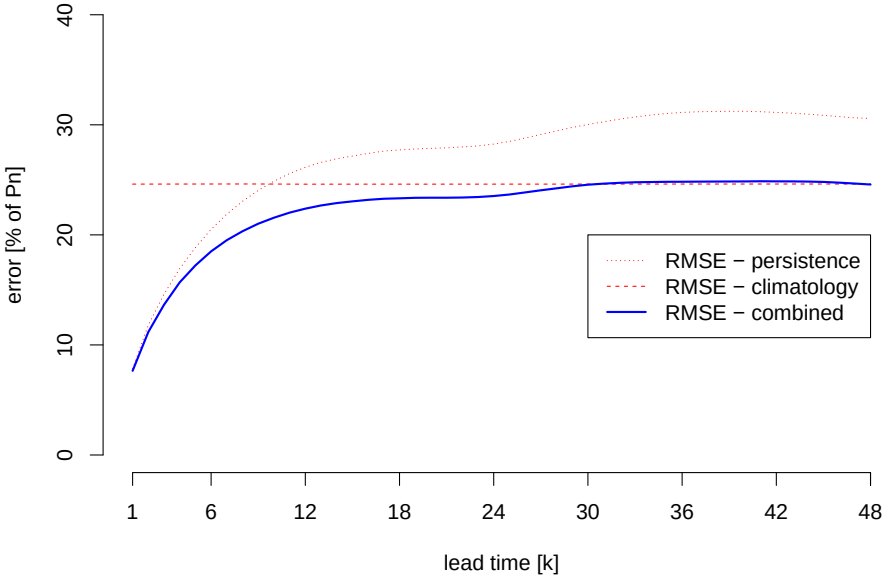
- *Right:*
  Example of a **combined** forecast for Klim (persistence and climatology), issued on 28 April 2002, 00:00UTC



- Let us similarly apply that strategy for a whole sample year (2002), and analyse its performance

Evaluation of the combined forecasts

- **RMSE** only, for *persistence*, *climatology*, and the *combined forecasts*

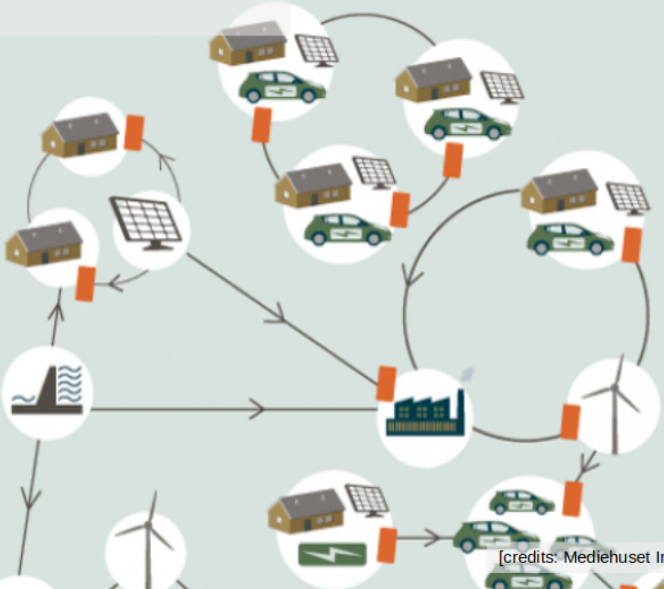


- With this combination strategy, we are getting the best out of the original simple benchmarks!

A few (more) conclusions at this stage

- We have now learned to handle more variables and data
- The forecasting approaches do not look impressive still
- *What could we do?*
  - *extracting more information* within available data
  - go further than using *simple linear relationships only* (to be discussed in the next Module)

Use the self-assessment quizz to check your understanding!



[credits: Mediehuset Ingeniøren]