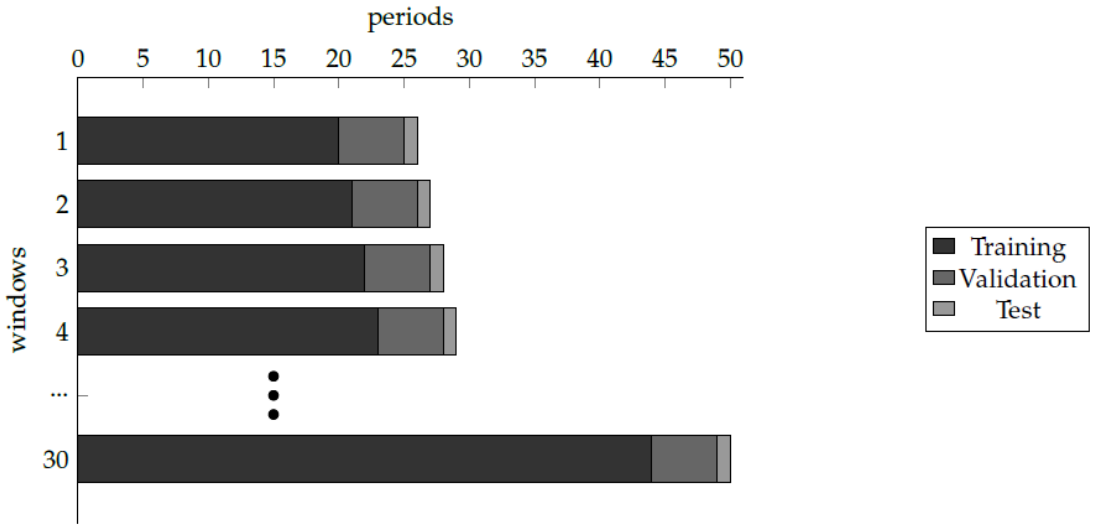
Modeling Exercise

Your task is to create a simple walk forward backtest using real stock market data. The general idea of walk forward backtest is to model and predict the target variable using the features in a rolling manner. Use preferably R or Python.

Data:

- Each row represents a stock at a specific point in time
- DATE is the date (there are multiple stocks at each time point)
- RET1M (target) is a future one month return for that stock
- FACTORS (features) describe the attributes of the stock (eg. P/E ratio) for that date
- INDUSTRY is the industry in which the stock belongs to

1. Read and explore the data
2. Pre-process the raw data to some standardized format
    a. Normalize/rank it across time dimension and/or groups (eg. sectors)
    b. Deal with missing data and outliers if you need
3. In a rolling fashion fit/train a model of your choice to data.
4. Predict one step ahead with the model and then increment or slide the training data by one step. You can choose the step size freely. It can be anything from 1 month to 12 months.
5. Report the accuracy/performance of your predictions in the out-of-sample periods. See the figure below how to obtain 'out-of-sample' predictions through time.



Remember, there are no right or wrong answers to this exercise. The financial returns are notoriously difficult to predict, and you might not get any good results at all. The purpose of this task is more to see how comfortable you are in wrangling and modelling the data. Articulate the choices you make throughout the code and submit a simple notebook/report via email to us.

Eero.Saarialho@ilmarinen.fi
Nikolaus.Juntunen@ilmarinen.fi
Heikki.Kinnunen@ilmarinen.fi