Quantitative strategies on High Frequency Data

Assessment of the sections (labs)

prof. Piotr Wójcik

academic year 2025/2026

General information

In teams of at most 2 persons students will build and backtest different trading strategies for **2 groups of assets**. Please inform the lecturer about the team members by email pwojcik@wne.uw.edu.pl (mailto:pwojcik@wne.uw.edu.pl) **the latest by midnight 2025-12-05**.

The data is **exactly the same for all teams** and consists of intraday data of in the `parquet` format in 1 minute frequency for group 1 and 5 minute frequency for group 2. The data covers real market quotations of selected futures contracts for the **period of 2023-01 – 2025-12** and is **divided in 12 quarterly files**.

For the purpose of **strategy selection and parameters search** students are initially given the **data just for 7 in-sample quarters** (2023Q1, 2023Q3, 2023Q4, 2024Q2, 2024Q4, 2025Q1 and 2025Q2). The remaining data for **5 out-of-sample quarters** (2023Q2, 2024Q1, 2024Q3, 2025Q3 and 2025Q4) will be delivered by the lecturer **after the submission of presentations** (see below).

Groups of assets

The groups of assets include:

1. Group 1 – two assets (1 min frequency, **traded during NYSE sessions** - on working days between 9:30 and 16:00 CET):
   - `SP` – futures contract for S&P 500 index (transaction cost = 12$, point value = 50$).
   - `NQ` – futures contract for NASDAQ index (transaction cost = 12$, point value = 20$).
2. Group 2 – four assets (5 min frequency, **traded almost 24 hours a day** with 1 hour break betweem 17:00 and 18:00 CET - quotations start on Sundays at 18:00 and last until 17:00 on Friday):
   - `CAD` – futures contract for Canadian dollar (transaction cost = 10$, point value = 100000$).
   - `AUD` – futures contract for Australian dollar (transaction cost = 10$, point value = 100000$).
   - `XAU` – futures contract for gold (transaction cost = 15$, point value = 100$).
   - `XAG` – futures contract for silver (transaction cost = 10$, point value = 5000$).

> CAUTION: There are separate data files for **group 1** and **group 2** for each quarter.

Any combinations within groups allowed

Within **each of the above groups** of assets you can:

- trade just a single asset, or
- put (selected) assets together in pair(s) as spreads, or
- trade each of selected assets separately and treat them as a portfolio (applying the same or different strategy for each asset).

If trading more than one asset (`spread`), remember to include **positive transaction costs for each of them**.

Trading sizes

Assume trading just with **one unit** of any security/spread, so the only positions available are:

- flat / neutral (`0`),
- short (`-1`),
- long (`+1`).

Different approaches, entry/exit techniques

For each of the (groups of) assets please consider and compare **at least 2 different types of entry techniques (approaches)**, each with **several combinations of parameters** (memories of moving statistics, multipliers, etc.).

As different approaches one may treat (each for the **trend following** or **mean reverting** strategy) for example an entry/exit technique based on:

- a single moving average/moving median/moving quantile,
- two or more intersecting moving averages/moving medians/moving quantiles,

- a single moving average/moving median/moving quantile and a selected volatility measure (breakout models),
- any other that comes to your mind.

Additional filtering

Additional filtering may be added (e.g. in pair trading strategies):

- based on correlation between two (or more) assets,
- based on regression between two (or more) assets,
- based on testing for cointegration between two (or more) assets,
- based on testing for Granger causality between two (or more) assets,
- any other that comes to your mind.

Common assumptions

Common assumptions for **group 1**:

- **do not use in calculations** the data from the first and last 10 minutes of the session (`9:31-9:40` and `15:51-16:00`) – put missing values there,
- do not hold positions overnight (**exit all positions** 20 minutes before the session end, i.e. at `15:40`),
- **do not trade** within the first 25 minutes of stocks quotations (`9:31-9:55`), but **DO use** the data for `9:41-9:55` in calculations of signal, volatility, etc.

Common assumptions for **group 2**:

- do not hold positions during the breaks (**exit all positions** 10 minutes before the break starts, i.e. at `16:50`),
- **do not trade** within the first 10 minutes after the break (until `18:10`).

One may make **additional assumptions**, however they should be **clearly explained and justified**, e.g. **stop-loss condition**, etc.

Selection of best strategy

> CAUTION !!!!! As mentioned before, the data are divided in **two parts** – **in-sample quarters** and **out-of-sample quarters**. At first teams are provided just with the in-sample data to do a research and **select the best strategy for each group of assets separately**.

Exactly the **same strategy** (the same entry/exit technique and parameters) has to be applied for a particular group of assets **in each quarter**.

For example if after research you find that for a particular asset the best strategy is a trend following strategy based on the cross-over of two exponential moving averages – EMA60 and EMA10 – you should apply this particular strategy with **the same parameters** and all other assumptions to **every quarter of your data** (first in-sample, then out-of-sample once available) and report the results.

The best/optimal strategy may be different for different assets, but again – it has to be **consistently applied on all quarters of data**.

> Selecting different best strategies (or just different parameters) for the same asset in different quarters of the data **is not allowed**.

Performance measures

For the selected best strategy for each group of assets **aggregate the strategy P&Ls to daily** and based on daily results calculate the following measures (separately for each quarter):

- **gross SR** – annualized Sharpe ratio based on gross daily P&L (**without** transaction costs, denoted in monetary terms),
- **net SR** – annualized Sharpe ratio based on net daily P&L (**with** transaction costs included, denoted in monetary terms),
- **gross CR** – annualized Calmar ratio based on gross daily P&L (**without** transaction costs, denoted in monetary terms),
- **net CR** – annualized Calmar ratio based on net daily P&L (**with** transaction costs included, denoted in monetary terms),
- **gross cumP&L** – **cumulative** profit and loss at the end of the investment period (last value of the cumP&L series) **without** transaction costs, denoted in monetary terms,
- **net cumP&L** – **cumulative** profit and loss at the end of the investment period (last value of the cumP&L series) **with** transaction costs included, denoted in monetary terms,
- **av.ntrades** – average daily number of trades.

and **report them in a table at the end of the presentation and report**.

Based on the above mentioned measures the final **summary statistic** will be calculated for each quarter separately. The **formula for the summary statistic** is the following:

$$stat = (netSR - 0.5) * max(0, log(abs(\frac{net.\ PnL}{1000})))$$

This promotes strategies which give positive net Sharpe ratios (above 0.5) and higher net pnl.

Please **add this statistic to the summary table** and in addition use codes that will **save this table as a `csv` file**.

In the end the **sum** of the above mentioned **summary statistic** over **all quarters** (in-sample and out-of-sample) will be used to **rank the teams**, divide them in quartile groups and give points for strategy performance.

Points

In total 100 points can be collected, given for:

- presentation in class prepared in **Quarto** including **working python codes** (`10 pts`),
- final written report prepared in **Quarto** including **working python codes** (`40 pts`),
- strategies performance (`20 pts`) – ranking based on a **summary statistic** described above, max. `10 pts.` per each (group) of assets results:

    - `10` if strategy performance in top quartile group (best),
    - `7.5` if strategy performance in the 2nd quartile group (good),
    - `5` if strategy performance in the 3rd quartile group (below average),
    - `2.5` if strategy performance in the 4th quartile group (unlucky),
- **obligatory** homeworks (`30 pts`)
- **extra points** can be collected for solving non-obligatory homeworks

Presentations

The **presentation** prepared in **Quarto** has to be submitted by email to the lecturer pwojcik@wne.uw.edu.pl (mailto:pwojcik@wne.uw.edu.pl) **until midnight 2026-01-18** (presentation should be submitted both as the source `*.qmd` file **and also** in the version compiled to `html` or `pdf` format). The python codes included in the `qmd` file should **load the data** from source files for each quarter, **apply** the BEST finally selected strategy on the data for ALL quarters, **calculate** P&Ls and **report** the results in the desired form.

> Do NOT include all the testing codes which you applied for strategy selection, parameter search, etc. ONLY a simple code for a FINALLY selected strategy i.e. with the selected set of best performing parameters for each group of assets – check the sample `qmd` files that will prepared by the lecturer by early December.

All teams will give presentations (**10 minutes**) informing about strategies considered and their **in-sample results**. The presentations do not have to inform about all the details of considered strategies.

> Only teams that submit presentations in a desired format with working python codes behind will obtain the **out-of-sample data**. Teams which do not provide `qmd` file with working python codes behind which apply their best strategies **will not be valued**.

All presentations will take place on 2026-01-20 (lecture or labs time, **16:45-18:20** or **18:30-20:00**).

> Groups that do NOT present their results in class **will get 0 points for presentation and for the out-of-sample performance part**.

Out-of-sample data

After all the presentations on 2026-01-20 in the evening the lecturer will provide the **out-of-sample data** to enable verifying the strategy performance and finishing a **final report**. Having prepared the report in Quarto with working python codes behind will make your analysis of the out-of-sample data very quick and on the other hand would allow the lecturer to verify reported results and check if all assumptions are met.

Final report

The **final written report** should be submitted by midnight 2026-01-25 (report should be submitted both as the source `*.qmd` file **and also** in the version compiled to `docx` format). It should include a **detailed explanation** of the finally selected strategy **for each group of assets** (approach, type and elements of the strategy, entry technique, assumptions, parameter values, etc.) and also shortly explain the process of final strategy selection. Measures of strategy performance should also be reported in a table (gross and net SR, gross and net cum P&L, average daily number of trades) together with **at least one figure** showing gross and net cumulative P&L of the strategy (**based on daily aggregated data**).

> Students who do not submit their final report **before the deadline** will not be allowed to **take the final written exam** in winter session.

Important dates again

- 2025-12-05 by 23:59 – submission of information about the **team members**
- 2026-01-18 by 23:59 – submission of **presentations** of **in-sample results** in Quarto **with working python codes** behind,
- 2026-01-20 – in class **presentations** of in-sample results – after that obtaining out-of-sample data.
- 2026-01-25 by 23:59 – **final report** submission

Each submission should be done **via email** to pwojcik@wne.uw.edu.pl (mailto:pwojcik@wne.uw.edu.pl) **before midnight** of the deadline day **if not stated otherwise**.

GOOD LUCK !!!!!!!!!!!!!