The M6 Financial Duathlon Competition

A great majority of the eighteen major forecasting competitions held to date have been numerical, asking for and evaluating the accuracy of point or probabilistic forecasts at a single origin, while using concealed data for evaluation purposes. The sole purpose of these competitions has been to learn how to improve forecasting accuracy and decrease forecast uncertainty (see Table 1). The M6 forecasting competition innovates on these objectives in the following three directions (see Makridakis, Fry, Petropoulos and Spiliotis, 2021)

- M6 is live, so that no data are concealed. Having a live competition allows participants to:
  - Search and use any available information for use in their forecasts and investment decisions.
  - include judgmental inputs about the economy at both the macro and the micro (industry and firm specific) levels. This allows participants to utilize both judgmental and numerical (model-based) inputs in order to improve their performance.
- Instead of using a single one, M6 uses multiple rolling origins to evaluate performance. This allows for participants to learn and to adjust their methods and/or models in real‐time. More importantly, considering multiple evaluation rounds allows separating skills from luck and investigating the consistency of the participants' performance over time.
- M6 consists of two challenges (forecasting performance and investment performance). In this sense, the competition is a duathlon, and winners will be announced in both sub-competitions, as well as in the overall combined competition. Additionally, prizes will be awarded quarterly, as well as at the end of the competition, with the "Global" winner being the participant who provides superior performance in both sub-competitions over the entire horizon of the competition. A key metric will thus be the ability of participants to effectively exploit forecasting methods to mitigate uncertainty and to translate their forecasts and findings into meaningful, profitable decisions.
- In summary, the M6 expands on the learning objectives of prior competitions by focusing on the interplay between forecasting and investing, as well as on the importance of forecast accuracy/uncertainty when used to support investment decision making.

Purpose

Forecasting competitions can only ever approximate reality, and their value and usefulness depends on how realistic such approximations can be made to be. From a motivational perspective, participants do not have "skin in the game" as they do not invest their own money, and we rely on modest prize money and professional and academic curiosity to encourage participants. To make the task manageable for a wide range of competitors, we limit the investment universe to 100 assets, but include both individual equity securities and asset classes via a range of ETFs. The M6 competition does allow for forecasting and investment decisions to be made in a truly real-time context, mimicking the true environment that real-world forecasters and investors operate in, but is of necessity limited in terms of time horizon and duration. M6 will run for 1 year, with 12 rolling assessment periods across which forecasts and investments will be made. While this time period limits the ability of M6 awards to properly account for fat tails associated with market returns, the structure of the competition still allows addressing several critical types of questions, including among others the following:

- What are the key differentiating factors associated with the good/poor forecasting and investment performance?
- What is the interplay between objectively measured forecasting performance based on strictly proper scoring rules and investment performance measured using criteria associated with portfolio optimization like the information ratio? A key aspect of M6 is that it captures both return forecasts and portfolio decisions.
- Are excess financial returns achieved by one or a combination of the following factors?
  - The ability to accurately forecast overall market returns, or those of individual assets.
  - The ability to properly model market or individual asset uncertainty.
  - The ability to formally optimize a portfolio when making investment decisions.
  - The ability to use judgement, model-based methods, or some combination thereof when constructing forecasts and making investing decisions.
  - The ability to adjust (or keep fixed) an investment strategy, over time.
  - Other factors, including judgmental and model based prediction and investment decision biases and inefficiencies that can be exploited in forecasting and in investment allocation.

No major competitions have been previously conducted in the area of financial forecasting. The M6 competition aims to provide empirical evidence about how investors can improve the accuracy of their forecasts, mitigate the uncertainty involved in these forecasts, and exploit their findings to build robust, profitable portfolios.

Duration and schedule of the competition

Quarter	Month 1	Month 2	Month 3
0	Sunday, March 6, 2022	Sunday, April 3, 2022	Sunday, May 1, 2022
1	Sunday, May 29, 2022	Sunday, June 26, 2022	Sunday, July 24, 2022
2	Sunday, August 21, 2022	Sunday, September 18, 2022	Sunday, October 16, 2022
3	Sunday, November 13, 2022	Sunday, December 11, 2022	Sunday, January 8, 2023

There will be two separate challenges with quarterly prizes awarded for each of the four quarters of the competition. In addition, there will be global prizes combining the performance of all the quarters (see Table 2). The quarterly prizes for the first, second, and third winners will be awarded to:

- The best performance in terms of forecasting (evaluated by the ranked probability score; see also section "Measuring the performance of the forecasts").
- The best performance in terms of investment decisions (evaluated by the information ratio; see also section "Measuring the performance of the investment decisions").
- The best overall performance of the above two challenges, winning the duathlon prize.

At the end of the competition, after 12 four-week periods of actual submissions, there will be global prizes for the top-performing participants awarded to the winners (see Table 2): The global prizes for the first, second, third, fourth, and fifth winners will be awarded to:

- The best performance in terms of forecasting.
- The best performance in terms of investment decisions.
- The best overall performance of the above two challenges, winning the duathlon prize.

In the extreme case of a tie, then the participants with the same score will share the respective prizes. For instance, if two participants tie on the 2nd place, then they will share the prizes for the 2nd and 3rd places.

Input data

The investment universe will consist of two classes of assets:

- 50 stocks from the Standard and Poor's (S&P) 500 index, and
- 50 international exchange-traded funds (ETFs).

The 50 stocks and 50 ETFs will be selected such that they are broadly representative of the market. We will announce the names of the 100 assets closer to the commencement of the M6 competition.

Submission format

Apart from the trial run for February 6, 2022, the actual competition has 12 submission points. The submission deadline for each point will be on 18:00 GMT the Sunday before the start of the corresponding investment period. The objective is to submit (i) forecasts and (ii) investment decisions at each successive monthly submission point that specify your forecasts and your investment strategy over the next four week period. In other words, the forecast horizon is four weeks (usually 20 trading days). There will be no overlapping evaluation periods, i.e., the second submission point will take place four weeks after the first submission point.

Example: The deadline for the first submission point is on 18:00 GMT 06-March-2022 (Sunday). The participants are to submit forecasts and investment decisions that will reflect on the closing value of the last trading day of the next four weeks, i.e., 01-April-2022 (Friday).

At each submission point, a participant may submit a single file consisting of seven columns of 100 values each (one per asset):

- The first column must indicate the asset to which the forecasts and the investment decisions of the respective row refer to. The acronym of each asset will serve as an identifier.
- The second to sixth columns must be values summing to unity that refer to the probabilities of the ranks of the forecasted percentage return for each asset (stocks or ETFs); rank 1 is the lowest forecasted percentage return and rank 5 is the highest forecasted percentage return (see also the section on "Measuring the performance of the forecasts"). If the sum of the submitted values (up to five decimal points) for each asset does not sum to 1, then the submission will be invalid. If any of the submitted values in columns 2 to 6 is negative, then the submission will be invalid. Participants will be notified about invalid submissions in order to resubmit.
- The seventh column must contain numerical values corresponding to the weights for investing on each asset. These values (up to five decimal points) must be positive for long positions, negative for short positions, or zero for no position. For instance, if three assets are assigned weights 0.5, 0.3 and -0.2, respectively, and all other assets a weights of 0, this means that the

participant wishes to invest in only three assets with positions long, long and short and with a budget allocation of 50%, 30% and 20% respectively. If the sum of the absolute weights exceeds 1 (or 100%), then the submission will be considered invalid. If the sum of the absolute weights is less than 1 (less than 100%), then the remainder is assumed to be assigned to an asset with zero return and zero risk (i.e., no investment). However, if the sum of the absolute weights is below 0.25 (25%), then a warning message will be given and the submission will be considered invalid, unless modified (i.e., some investment must be made and some risk must be taken).

Example: The following is an example for the first 8 rows of a submission file. In this case, the participant decides to invest in three assets (3rd, 6th and 7th) with weights 50, 30 and 20% (or 0.5, 0.3, and 0.2) and positions long, long and short respectively. Additionally, the participant forecasts that there is a probability of 0.1, 0.2, 0.5, and 0.2 that the first asset (MMM) with rank 2nd, 3rd, 4th, and 5th, respectively, with regards to the expected percentage return. Equally, the participant's forecast is that the second asset (ATVI) will be ranked 3rd.

ID	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5	Decision
MMM	0	0.1	0.2	0.5	0.2	0
ATVI	0	0	1	0	0	0
GOOGL	0.1	0.1	0.1	0.1	0.6	0.5
APH	0.5	0.4	0.05	0.05	0	0
BMY	0.2	0.2	0.2	0.2	0.2	0
CB	0	0	0.1	0.4	0.5	0.3
EXR	0.7	0.3	0	0	0	-0.2
MSI	0	0	1	0	0	0

The stocks included in this example were randomly selected to illustrate their use.

Important notes:

- It is possible that a participant decides not to submit forecasts and investment decisions at a particular submission point. In such a case, we will assume that their previous submission carries over. In other words, their forecasts and investment decisions do not change (percentages for investment decisions will be kept the same as at the start of the last submitted period).
- If a participant does not submit forecasts and investment decisions in the first month of a particular quarter, and if there are no submissions to be carried over, then they are not eligible for the prize for that particular quarter.
- If a participant is not eligible for a prize for a single quarter, then they are automatically not eligible for the global prizes (awards based on the performance across all 12 submission points). In other words, in order for a participant to be eligible for a global prize, they have to submit forecasts and investment decisions from the very first month of the competition.
- Incomplete files of less than 100 rows will be invalid submissions.

Measuring the performance of the forecasts

The forecasting performance for a particular submission point will be measured by the Ranked Probability Score (RPS).

The realised percentage total returns of all assets (stocks and ETFs) over the period are divided into quintiles, ranking from 1 (worst performing) to 5 (best performing). Given 100 assets, 20 of these will receive a rank of 5, 20 a rank of 4, and so forth. In cases involving a tie on the margins of the classes, the tied assets will all be assigned the respective average rank. For instance, if four assets are tied at the 18th place, then they will all get a rank of (5+5+5+4)/4=4.75 , with the three "5's" in this exp

ression being the rank of the 3 assets in the first quintile, and the "4" being the rank of the asset in the second quintile. The actual return ranking of each asset is described by a vector $q_{i,T}$ of order 5.

- In the case of no ties on the margins of the classes, the elements in this vector, $q_{i,T,k}$ with k ∈ 1,…,5 are set equal to one if the asset is ranked in quintile k and zero otherwise. For instance, if asset *i* is ranked in quintile 3 at time *T*, then $q_{i,T}$ has values 0, 0, 1, 0, and 0.
- In the case of ties on the margins of the classes, then the values assigned to the elements of the vector $q_{i,T}$ are calculated such that the tied classes are assigned non-zero weights, with the respective weighted average being equal to the actual rank. For instance, following the above example of a 4.75 rank, the values of $q_{i,T}$ would be 0, 0, 0, 0.25, and 0.75, such that 0×1+0×2+0×3+0.25×4+0.75×5 = 4.75.

Similarly, we construct a vector denoting the probabilities of each rank for a particular asset $f_{i,T}$, as submitted by a participant. The RPS for asset *i* in period *T* is then:

$$\boldsymbol{RPS_{i,T} = \frac{1}{5}\sum_{j=1}^{5}\left(\sum_{k=1}^{j} q_{i,T,k} - \sum_{k=1}^{j} f_{i,T,k}\right)^2}$$

$$\boldsymbol{RPS_T = \frac{1}{100}\sum_{i=1}^{100} RPS_{i,T}}$$

The overall RPS for multiple submission points t_1 to t_2 is:

$$\boldsymbol{RPS_{T_1\text{-}T_2} = \frac{1}{100(T_2 - T_1 + 1)}\sum_{T=T_1}^{T_2}\sum_{i=1}^{100} RPS_{i,T}}$$

The RPS is zero for a perfect score, and positive otherwise.

Example: We wish to compute the overall RPS of a participant for a particular asset, i , at one submission point, T. The submitted probabilities for the ranks are 0, 0.2, 0.3, 0.4, and 0.1. The actual rank was 4. As such $q_{i,t}$ = [0, 0, 0, 1, 0] and $f_{i,\,t}$ [0, 0.2, 0.3, 0.4, 0.1], $RPS_{i,T}$ is calculated as:

$$RPS_{i,t} = \frac{(0-0)^2 + (0-0.2)^2 + (0-0.5)^2 + (1-0.9)^2 + (1-1)^2}{5} = 0.06$$

Measuring the performance of the investment decisions

The performance of the investment decisions is measured by means of a variant of the Information Ratio, IR, defined as the ratio of the portfolio return, ret, to the standard deviation of the portfolio return, sdp. Namely, risk adjusted returns are defined as:

$$IR = \frac{ret}{sdp}$$

where ret denotes continuously compounded portfolio returns, and sdp denotes the standard deviation of these returns, measured at a daily frequency. Note that all reported IR values are annualized. Additionally, IR is a variant of the typical Information Ratio, but with the benchmark return set equal to 0; and is also a variant of the Sharpe Ratio, but with the risk free rate set to 0. All return calculations begin with the daily portfolio holding period return, calculated as:

$$\boldsymbol{RET_i = \sum_{i=1}^{N} w_i \left(\frac{S_{i,t}}{S_{i,t-1}} - 1\right)}$$

where N denotes the number of assets, wi is a portfolio weight, and Si,t denotes the price of asset i at the end of trading day t, with t — 1 referring to the previous trading day. In all return calculations, prices are adjusted closing prices. Continuously compounded portfolio returns are then calculated as:

$$ret_i = ln(1 + RET_i)$$

In the above expressions, RETt , is measured for a single day, t, and is the percentage return (gain/loss) associated with each asset selected for investment, averaged by the corresponding investment decision weight for each asset. Returns for a holding period longer than one day are calculated as the sum of daily returns. In particular, the return for the holding period from t 1 to t 2 is calculated as:

$$\boldsymbol{ret_{t_1:t_2} = \sum_{t=t_1}^{t=t_2} ret_t}$$

The standard deviation, $sdp_{t1:t2}$, is calculated using the same $t_2 - t_1 + 1$ values of ret_t as those used in the calculation of $ret_{t1:t2}$. In particular:

$$\boldsymbol{\mathrm{varp}_{t_1:t_2} = \frac{1}{T-1} \sum_{t=t_1}^{t_2} \left(\mathrm{ret}_t - T^{-1}\mathrm{ret}_{t_1:t_2}\right)^2}$$

And

$$\mathrm{sdp}_{t_1,t_2} = \sqrt{\mathrm{varp}_{t_1:t_2}}$$

With $T = t_2 - t_1 + 1$

Higher IR values suggest better investment performance.

Example: We wish to construct the Information Ratio of one monthly investment decision of a participant, carried over the 4 week assessment period. We calculate daily compound returns, yielding 20 r_t observations. Summing these observations yields $ret_{1:20}$= 0.01. Also, we calculate that $sdp_{1:20}$=0.01. Then, we have:

$$IR_{t_1:t_2} = \frac{\frac{21}{20} \times 12 \times 0.01}{\sqrt{252} \times 0.01} = 0.79$$

Note that in this example, as in all of our investment performance assessments, daily returns on investment decisions are utilized. This allows for more degrees of freedom when calculating the standard deviation.

Measuring the combined performance of the forecasts and the investment decisions

The combined performance is measured by means of the arithmetic mean of the ranks of the ranked probability score, RPS and performance of the investment decision, IR, which assumes equal importance between the two tasks. As such, the overall rank for submission t, OR, is calculated as:

$$\boldsymbol{OR} = \frac{\textbf{rank}(\boldsymbol{RPS}) + \textbf{rank}(\boldsymbol{IR})}{\mathbf{2}}$$

where rank(•) returns the rank of a participant relative to all other participants for that measure (RPS or IR). To calculate the overall forecasting rank, RPS, across all 12 submission points, we take the arithmetic mean of the RPS as calculated in each month.

Reference

Spyros Makridakis, Chris Fry, Fotios Petropoulos and Evangelos Spiliotis, 2021,
The future of forecasting competitions: Design attributes and principles.
https://arxiv.org/abs/2102.04879

These Guidelines were prepared by the M6 Academic Committee:
Spyros Makridakis (University of Nicosia); Anil Gaba (INSEAD); Ross Hollyman (University of Bath); Fotios Petropoulos (University of Bath); Evangelos Spiliotis (NTUA); and Norman Swanson (Rutgers University).

mofc@unic.ac.cy mofc.unic.ac.cy