How Should Forecasts Be Engineered?
The Indispensable Markets Hypothesis

Abstract

We consider evidence in support of what we term the Weak Indispensable Markets Hypothesis (IMH) provided by the M6 Financial Forecasting Competition. The Weak IMH asserts in the presence of a well-established market, those who eschew prices as inputs for proximate predictive modeling tasks will under-perform out of sample. We also consider the Strong IMH which asserts that forecasting should be considered a market-inspired engineering endeavor *in addition* to a modeling task under the usual rubric of statistics or machine learning methods (put simply: if a market doesn't exist to help you, make one!). The competition established that neither principle's application is obvious to participants or organizers; it hinted at a hidden quality crisis in data science generally; and it suggests that broadening the usual concept of analytic pipelines to insert collective intelligence might be part of the remedy.

Keywords: Predictive modeling, Probability, Efficient Markets Hypothesis, Forecasting, Collective Intelligence

1. There's more to forecasting than models

This paper reports on a simple experimental finding in dire need of explanation. In a worldwide contest involving hundreds of data scientists building the best possible probabilistic prediction models and portfolios, and chasing $300,000 in prize money, not a single team convincingly outperformed a diversified portfolio whose construction used no forecasting at all. In addition, the small percentage of competition entries that outperformed this benchmark on a combination of prediction and investment tasks were almost certainly lucky, as revealed by unrealistic, unrepeatable information ratios no practitioner would take seriously.

Here we document what competitors missed and argue that although they could also have done something equally simple, they might not have found this *conceptually* natural or *instinctive*, and that failure to anticipate the possibility (by participants and organizers alike) points to a possible blind spot in the broader analytic community. They missed the opportunity to harness the forecast power of a "nearby" options market and, by doing so, accidentally allowed for a comparison between prediction using models and prediction using a market that turned out to be unfortunate for the former.

Although it is mostly outside of the present scope, one might be tempted to extrapolate further and question the tradition of equating forecasting with model building. This posture assumes that applying mathematical technique is the primary way to advance prediction quality if not the only one. By implication, all forecasting methods must be understood by a single person and can be written down in excruciating detail in journals such as the present one, so that the precise path from historical data to future predictions can be understood and replicated.

That view appears reasonable, and yet it is similar in many ways to assuming that the only way to organize an economy is through central planning. By that analogy, this ubiquitous position is deeply suspect because it does not address the central problem of how to use information that is possessed partly by many but wholly by none (Hayek (1945)). There is at least one other way to predict, albeit more anarchic, and that is by constructing markets or devices motivated loosely by the same.

Here we mainly deal with the even simpler possibility of using a market that *already exists.* To be crystal clear, unlike other papers written about the M6 Financial Forecasting Competition, the point of this work is not to compare two or more forecasting methods, but to report an empirical comparison between a forecasting community's efforts and the predictive power of a nearby market. This experiment used the M6 Financial Forecasting Competition as a convenient device, but it might be said that it hijacked it for an experimental purpose rather different from the organizers' intent. Our goal was to test whether a "lazy" use of the information contained in the options markets would suffice to perform well.

The laziness characterization is ultimately for the reader to judge (referring to Section 6.1) and is somewhat irrelevant. The real point is that its strong performance suggests that few utilized options prices *at all.* This discussion is predicated largely on acceptance of that stated experimental finding: that a simple use of implied volatility described in Section 6.1 performed better out of sample than the vast majority of analytical approaches unearthed by the contest.

As an aside, it is also *plausible* that the simple entry described might have been revealed as superior to *all entries* eventually had the contest continued to run beyond a year, which would have been an even more stunning result - but this should not distract from the core experimental finding and the reasonable conclusion: *all* entries could have benefited from the use of options market data.

Given that premise, the goal of this paper is foremost to seek a remedy for the poor state of affairs that this appears to reveal in the data science community. At first blush one might think that the Efficient Markets Hypothesis (EMH) would provide a sufficient reminder of the utility of markets for forecasting. But not so. The EMH was prominently mentioned in the framing of the M6 Competition. The EMH was in fact the raison d'etre for the contest. But perhaps that is not so useful a reminder of the utility of markets if nobody is inclined to believe it!

We remark further in Section 2 on our supposition that the EMH is simply too unpopular, on account of its self-evident inaccuracy, to serve as a useful crutch or reminder. We shall introduce in Section 3 what we term the *Indispensable Markets Hypothesis* (IMH). There is no need for flaws in the EMH, some obvious, to detract from the simpler message of this paper or the reminder it should serve in the future. The weaker principle we put a name to, in contrast to the EMH, might survive both empirical research and elementary economic logic while still serving what would appear to be a crucial purpose.[1]

2. The Efficient Markets Hypothesis as failed mascot for the power of markets

We give some consideration as to whether, in order to encourage greater respect for markets and in particular the use of market data as modeling input, any new pedagogical or conceptual mnemonic is required. The most popular expression in the vicinity is the Efficient Markets Hypothesis that, with an irony that is worth repeating, was the very topic under study in the M6 Financial Forecasting Competition.

Unfortunately, as it stands, the Efficient Markets Hypothesis is at best an uncomfortable approximation. In its various guises it asserts that it is difficult or impossible for market participants to extract rent or find predictability in pricing - yet this leaves open the problem of explaining how this pristine state can ever arise, or constitute a true equilibrium given the significant costs

[1]The only real objection to the Indispensable Markets Hypothesis is triviality. However, triviality has been rejected once and for all by the M6 Financial Forecasting Competition.

that can, and rather obviously are, incurred by those participants who drive prices towards near martingales. The point was made by Farmer & Skouras (2012):

> On the one hand, for markets to be efficient, there must be financial agents who detect inefficiencies and exploit them by trading on them, altering prices to make them more efficient. On the other hand, if the market is perfectly efficient, there should be no profit-making opportunities to motivate such agents. Thus the market can never be perfectly efficient: There must be persistent residual inefficiencies to keep financial traders motivated to do their job of making the market more efficient.

Cost is just one of the self-evident flaws that turn many people off the EMH, including those deeply sympathetic to the notion that markets can be excellent aggregators of information *despite* the persistence of residual rent-extraction possibilities.

There is an economic just-so story in which we imagine shovel-carrying market participants filling in holes in the market distribution left by others, potentially at low cost. However elementary calculus can suggest a more nuanced and, from the point of view of the Efficient Markets Hypothesis (EMH), occasionally disappointing picture. We give two simple examples intended to undermine the simplicity of the rationale for the Efficient Markets Hypothesis attributed to Fama (1970), or more precisely to the assumption that rent extraction should be largely impossible.

Consider first a knowledgeable race-goer with logarithmic utility. Suppose they invest the entirety of their wealth in proportion to the true winning probabilities known, we might assume, only to them. Perhaps surprisingly, this action is economically rational even though your decision is made *regardless* of the market probabilities (that is, the odds). In that sense, they are not trying to “fill in any holes”. They act as if they don’t care at all about, and don’t even need to see, the market prices on offer.[2]

This indifference might be dismissed as a mathematical quirk arising from logarithmic utility, but it is not fully resolved even in the case of an investor with no risk aversion at all. We can consider the mechanistic example of a parimutuel market (in keeping with our horse racing theme). There, by construction, state prices are proportional to aggregate investment by all participants

[2] The result requires zero track commission, and follows immediately from the use of Lagrange multipliers, as per Cotton (2013).

and, thereby, it is easy to see that even the hypothetical linear utility investor (again, with perfect knowledge of the probabilities of each horse winning) will *still* not bet a sufficient amount to bring the market and the true probabilities into alignment. Rather, they will bet roughly half that amount - the point where the marginal return falls to zero.

It seems that strict EMH devotees can easily run headlong into Zeno-like paradoxes that call for a large number of informed participants to eliminate rent-extraction possibilities. More's the reason why the substantial computational and human costs of making markets more efficient can't tritely be ignored.

That high cost of achieving efficiency (in the traditional sense) is reflected in data costs, technology expenditure, and compensation, and the latter is considered in several critiques of the financial sector, including Philippon (2015) and previously Henwood (1997). None of this is to suggest that markets should casually be described as inefficient. Instead, it should be possible to refer to an entirely defensible position that acknowledges ongoing rent extraction, rather than the much-criticized EMH. Many practitioners regard it as a contradiction of their daily work, for one thing.

That is why we prefer a weaker notion couched as a statement about probabilistic modeling performed *in the proximity of a market*, rather than being a direct claim about any characteristic of that market per se or a claim about what that market does or does not do (such as aggregating all information). As far as prediction is concerned, markets can be necessary without being sufficient.

Nor do we seek to make a strong comment on rationality that might be in opposition to behavioural observations in Barberis & Thaler (2003) and much work since. Our modest goal is to introduce an abbreviation for something that practitioners, behaviouralists, and EMH skeptics of various flavors *can* believe.

3. The (Weak) Indispensable Markets Hypothesis

The Indispensable Markets Hypothesis, as we term it, states that if a liquid collection of markets exist which *imply* a forecast, even partially, then that market information is to be considered *indispensable* to "nearby" modeling activity. For example in the presence of a collection of options contracts on a stock that incompletely imply a risk-neutral distribution of future returns, any completely independent effort to model future returns by whatever probabilistic means will be inadequate.

The IMH does *not* state that the risk-neutral market distribution will converge to the true distribution - nor that one exists. It states only that price data is vital and there is no complete alternative such as, to pick one example, a combination of machine learning and conformal prediction (say) applied using historical data only (Vovk, Gammerman & Shafer (2005)).

And whereas the EMH states that option markets will incorporate *all* information from the past or present, the IMH only suggests that the options markets will incorporate *some* information that will almost escape the reach of modeling efforts performed by any single individual or group over any reasonable time frame.

It may be possible under the Indispensable Markets Hypothesis for a market participant to occasionally spot inefficiencies and exploit them. It may be possible that some information from the past or present sometimes fails to be incorporated into options markets (say). But it is nonetheless unwise for any individual or group of individuals to completely eschew market information. It is *impossible* for them to completely replace the predictive power of said market on an ongoing basis.

4. The Strong Indispensable Markets Hypothesis

This work will provide evidence for the Weak IMH only, using our small contrivance and the M6 Competition. However, it is motivated by broader concerns and, given what we believe to be a positive affirmation of the weak form IMH, it is worth considering a more controversial version of the same.

The strong form of the Indispensable Markets Hypothesis, as we refer to it, states that where a well-established liquid market does *not* exist in proximity to a forecasting activity, it is still beneficial to *create* a small market (or game) which can serve a similar role and be indispensable in the same sense. This bolder statement has obvious ramifications for forecasting work in generality and, for that matter, the organizational behaviour of firms engaged in modeling work.

The present work advances only indirectly this stronger form of the IMH by providing evidence for the weak. The longer argument is left to Cotton (2022a) and might lean on several additional tenets: that the cost of market construction be dramatically lowered; that those task-specific markets be largely the province of autonomous, self-navigating algorithms rather than humans; that efficiencies of scale are achieved through sharing of real-time feature spaces; and so forth.

For brevity, our use of IMH in this paper refers to the weak form by default. It seems pragmatic to first convince people of the importance of using existing markets, and overcoming the bias against "prices as data" before trying to convince them to create new market-like constructs that will initially be less powerful.

5. The M6 Financial Forecasting Competition as laboratory

When the M6 Financial Forecasting Competition was announced in late 2022, organizers described it as "similar to the previous five" insofar as competitors bringing ideas from statistics, machine learning and other fields would "empirically identify the most appropriate way of forecasting".[3] This time the target would be stock and ETF prices and, the organizers hoped, the competition would test whether prices reflect all information - the version of the Efficient Markets Hypothesis used in the announcement (University of Nicosia (2024)).

But which prices? And which information? The reference was clearly to stock prices that contain no information about stock volatility - the crucial ingredient for the distributional prediction task demanded of participants.

Thus despite the reference to EMH it seemed on the surface that the M6 Competition would be very much like the others as far as volatility prediction was concerned: a straight-up battle between modelers and modeling approaches intended to advance the state of the art. And yet this particular competition, unlike any before to our knowledge, was in some sense tricking participants into performing a distributional prediction activity that was, to a large extent, already being performed by a market - with just enough shrouding to avoid notice.

Entrants in the M6 Competition supplied quintile probabilities for monthly asset returns, a challenge not identical to volatility prediction or option pricing, but *very* closely related (especially if we take a skeptical view on the ability of participants to locate alpha).

The organizers did nothing to reveal the ruse: they focused on the underlying equity markets, not the options market. "The M6 competition will determine if above average financial returns are achieved by one or a combination of the following factors:", read the announcement, including "the ability to properly model market or individual stocks/ETF uncertainty".

[3] Objectives are still listed on the competition website University of Nicosia (2024)

That statement belied the tacit and widespread assumption in predictive analytics that we rail against here. It ignores the possibility of constructing *other types* of forecasting machines or arrangements that extend beyond pure modeling. It might have been possible to produce an entry for the M6 Competition using pipes and pulleys.[4].

The M6 Competition had two sides. In addition to providing quintile probabilities for monthly stock returns (i.e. the probability that any particular asset's monthly return fell in the second quintile with respect to 99 other assets) participants were also required to submit a portfolio and rebalance it monthly - to be judged by daily information ratio. It was up to participants to decide whether to relate one task to the other. They also had to decide whether to take a view on mean returns.

Participants who took positions in assets, or skewed their rank probabilities with direction in mind, believed there were trying to beat the equity markets' estimates of asset value. Those who declined to assert directional views - probably wisely - believed they were pitting their wits against fellow statisticians making volatility and probabilistic predictions. In trying to "empirically identify the most appropriate way of forecasting", as instructed, most seemingly assumed they should be traversing the usual search space used in competitions of this sort: loosely speaking the set of all Python, R or Julia scripts to be brought to bear on a set of regressors and lagged asset values.

Few if any participants realized they were also facing off against the formidable distributional prediction power of the options market, and nor was this commented on by organizers or contemplated in the framing of objectives. The goal of assessing EMH as it applies to the *underlying* asset prices was useful misdirection, from this perspective, as were references to Warren Buffett - known for his directional acumen not his volatility estimation ability.

So, although this competition might have been superficially similar to other forecasting, data science or machine learning contests it clearly was not. It offered a unique scientific opening as with a rare astronomical event: a fluke alignment allowing the best efforts of statisticians, machine learning experts, and others in the forecasting community to unwittingly face off against the predictive power of a sizeable market assuming only that the thin veil, the distinction between volatility estimation and return quintile probabilities, was not pierced.

The M6 Competition provided a single data point to calibrate the relative strength of the

[4]Or more plausibly by entering this competition using another competition, as discussed in Cotton (2019)

forecasting community (and one data point is a lot better than none).

Needless to say the options market serves as a prediction device just as surely as any that might be constructed from code or tea leaves by participants. It aggregates opinions from traders and the quantitative people who support them, all of whom have more financial incentive to provide forecasts than was provided by the relatively generous prize pool for the M6 competition. And while there are well-understood financial reasons why implied volatilities might diverge from real world counterparts, the M6 Competition called upon participants to discern (to first order, and assuming that alpha is hard to come by and correlation secondary) only the *relative* volatility of assets.

6. An options market informed entry in the M6 Competition

The opportunity to calibrate the predictive power of data scientists against a market did not go completely unnoticed. An entry under the team name *microprediction* was provided that was, morally speaking, *implied* by the options markets (again, to within some choices for correlation estimation that matter less than relative volatilities). In this way a second battle was surreptitiously created - not between teams besting each other with better statistical and machine learning methods, but between the contest population as a whole and a different, hidden but lurking quant community - silent participants in the M6 Competition, as it were, operating from afar as they traded options or provided models for the same.

The result was not hard to predict, in our view, but the extent of the victory for that second community was surprising. And therefore we assert that the key lesson of M6 may lie not with a collection of winning forecasting techniques (interesting as they might be) but the resounding victory for the traders and quants who *didn't* enter the contest and *didn't even know about it.* Their work was translated into rank probabilities by an M6 submission that was (in our view) reasonably representative of their labors. Although one would expect this submission to be score worse than a reasonable fraction of competition entrants just by chance, it was bettered by remarkably few.

Code for the entry that provided a connection between the predictive power of the options markets and the M6 Competition is provided at Cotton (2022b) and was also described previously in Cotton (2023). In short, the entry used the implied volatilities from the options markets to augment an econometric covariance matrix estimate, and then used a very defensive portfolio construction that was not overly sensitive to the off-diagonal entries in the same. The rank probabilities produced in this fashion (as with any other) were primarily a function of the relative scale of implied volatilities for the 100 assets in the portfolio.

We now give more detail.

6.1. Rank probability methodology

Quintile probabilities were computed using entirely standard multivariate normal simulation.[5] The most important input, as we have remarked, is the relative volatilities to assign to assets and for that the *microprediction* entry adopted 30-day implied volatilities for stocks and ETFs with minor shinkage modification, as follows.

Before use in simulation the implied volatilities were shrunk by $\lambda = 0.25$ towards their grand mean, resulting in an estimate $\hat{\sigma}$:

$$\hat{\sigma_i} = (1-\lambda)\sigma_i + \lambda \frac{1}{n} \sum_{k=1}^{n} \sigma_k$$

where σ_k is the k'th implied 30-day volatility.[6] Equity volatilities were inflated 5% relative to ETFs in an ad-hoc attempt to reward the diversification implicit in the latter in the investment side of the contest, and an even more ad-hoc attempt to account for the fatter tails of the former. Correlations were estimated using the precise package Cotton (2022b) and, specifically, an exponentially weighted partial-moments method with a memory of 100 time steps and a decay of 0.01 was employed.[7]

6.2. Portfolio construction methodology

We take the view that as far as differentiating teams was concerned, the statistical power of the forecasting side of the M6 Competition was much greater than on the investing side. Our

[5] See `https://github.com/microprediction/precise/blob/main/precise/skatertools/m6/quintileprobabilities.py` for rank probability code.

[6] Implied volatilities were entered by hand from Market Chameleon and on occasion longer dated volatilities also entered the consideration. The motivation for the shrinkage was increasing the probability of beating the uniform benchmark - something that turned out not to be a serious concern. In retrospect, it might have been better not to perform this step.

[7] See `https://github.com/microprediction/m6entry/blob/main/m6entry/step1_probs.py` for code used and precise details. The partial moments code was in turn adapted from the NNS package Viole (2020).

thesis should therefore rest primarily on the results of quintile probability estimation rather than investment returns, and in this respect the precise details of portfolio construction used in the *microprediction* entry are likely of secondary importance - but described in detail in the appendix for completeness.[8]

Briefly: a covariance matrix was inferred from the options-implied volatilities and correlations described above. Then, a diverse long-only portfolio was constructed using a novel robust portfolio construction method similar in form (and nesting) the better-known Hierarchical Risk Parity (HRP - De Prado (2016)). This top-down allocation scheme made defensive, cautious, partial use of the covariance matrix to avoid the well-documented problems associated with the construction of matrix inversion and minimum variance portfolios (as considered in Jain & Jain (2019), DeMiguel, Garlappi, Nogales & Uppal (2009), Molyboga (2020) and Lan (2015) for example).

Pseudo-code for some ad-hoc covariance adjustments is provided in Algorithm 1. Aside from the orthodox assumption that volatility of assets would be roughly rewarded with slightly higher mean return, and the further assumption that this higher return would be somewhat ameliorated by a "missing volatility premium" (Falkenstein (2012)) no opinion on the relative performance of stocks or ETFs was taken. The final line of pseudo-code in 1 refers back to the mentioned portfolio construction.

7. Results

Table 1 reports the percentile finish for the microprediction entry in the prediction half of the contest - where specification of rank probabilities was demanded. As can been seen, in all five phases the microprediction entry finished in the top quartile of participants.

This would be highly unlikely under any reasonable null hypothesis where all or most participants took advantage of option market data.[9]

Performance in the overall contest combining both prediction and investment was strong too. The microprediction benchmark was so close to the year-long podium that an organizer prematurely

[8]See also `https://github.com/microprediction/m6entry/blob/main/m6entry/step2_portfolio.py` The code provides some ability to favor or shy away from individual stocks or ETFs that was *not* used in the *microprediction* entry. Similarly, the capability to slightly modify stocks with upcoming earnings was underutilized.

[9]We do not reject the possibility, but we are yet to unearth *any* evidence, in forums, social media feedback or elsewhere, that *anyone* else used implied volatilities.

Algorithm 2 Pseudo-code for modifications to an estimated covariance matrix used by the microprediction entry. Line 17 calls out to a new type of top-down portfolio construction known as Schur Complimentary Allocation. That methodology nests both minimum variance portfolio construction and also Hierarchical Risk Parity, and is explained in the appendix.

1: $\mu \leftarrow (2 + \text{diag}(\text{cov}))/3$
2: risk_exponent $\leftarrow$ 0.5 ▷ Higher vol is relatively riskier
3: earnings_penalty $\leftarrow$ 1.1 ▷ Higher for riskier earnings
4: LS $\leftarrow$ False
5: **if** LS **then**
6:     exclusion_penalty $\leftarrow$ 10.0
7: **else**
8:     exclusion_penalty $\leftarrow$ 1.02
9: **end if**
10: risk_penalties $\leftarrow$ exp(normalize(log($\mu^{\text{risk\_exponent}}$)))
11: exclusion_penalties $\leftarrow$ [exclusion_penalty if *tck* not in INCLUDE else 1.0 for *tck* in tickers]
12: earnings_penalties $\leftarrow$ [earnings_penalty if *tck* in EARNINGS_SOON else 1.0 for *tck* in tickers]
13: penalties $\leftarrow$ exclusion_penalties $\times$ earnings_penalties $\times$ risk_penalties
14: $S \leftarrow$ scatter(penalties)
15: penalized_cov $\leftarrow$ cov $\times$ S
16: penalized_cov $\leftarrow$ nearest_pos_def(penalized_cov)
17: long_w $\leftarrow$ port(cov = penalized_cov)

Stage	Percentile
Pilot	85
Q1	78
Q2	77
Q3	86
Q4	89
Overall	91

Table 1: The microprediction entry's percentile rank amongst all participants for the five parts of the year-long rank probability estimation competition cannot be attributed to chance.

announced it as a cash winner on social media - not realizing there was another update to come. After that final update, the entry was still so close that only a difference of 0.0002 in the Brier score would have been required to finish in the top five of close to 200 teams.

It is further tempting to speculate that, had the contest continued to run for a second year or a third, the microprediction entry would have steadily improved and more contestants fallen back. A trajectory of one top competitor is shown in Figure 2 and, clearly, there appears to be some significant volatility in overall ranking, even well into Round 5. That is not the case for the microprediction entry, which rose steadily through the ranks after a sluggish start.

It is possible that *any* robust portfolio construction method would have been sufficient to beat most of the entrants, and in fact the relative success of the equal weighted portfolio was remarked on by the organizers in Makridakis, Spiliotis, Hollyman, Petropoulos, Swanson & Gaba (2023).

The difference between the portfolio construction algorithm used in the microprediction entry and HRP is not likely to be of sufficient magnitude to materially alter the main finding here, but details of the new construction are provided for full transparency.

The performance of this entry might not provide complete support for the Efficient Markets Hypothesis as it applies to the options market, since only the *relative* volatilities matter to first order in the determination of quintile probabilities. It would be easy to construct EMH-defying models for option markets that are consistent with our experimental result by appealing to the difference between realized and implied volatility, for example (cf. Driessen, Maenhout & Vilkov (2009), Broadie, Chernov & Johannes (2009), Bollerslev & Todorov (2011), Taleb (2007)).

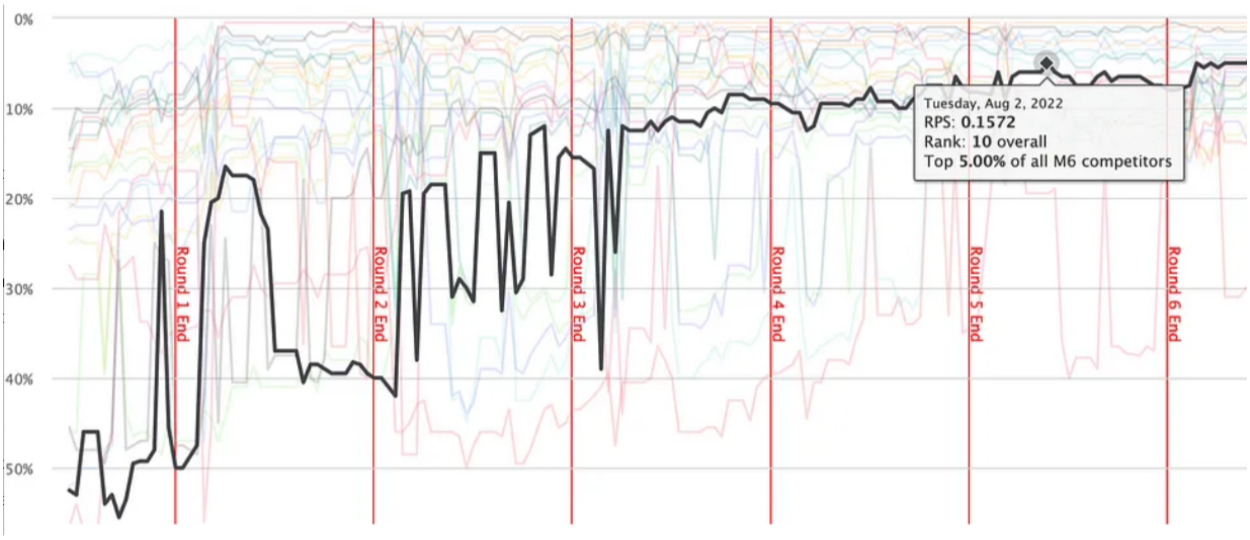

Figure 1: Steady progression of the options-market benchmark in the overall contest as shown on a screenshot of a tool built by JB Kirkland for the benefit of participants. The scale on the x-axis is roughly one year. The y-axis shows the percentile amongst participants.

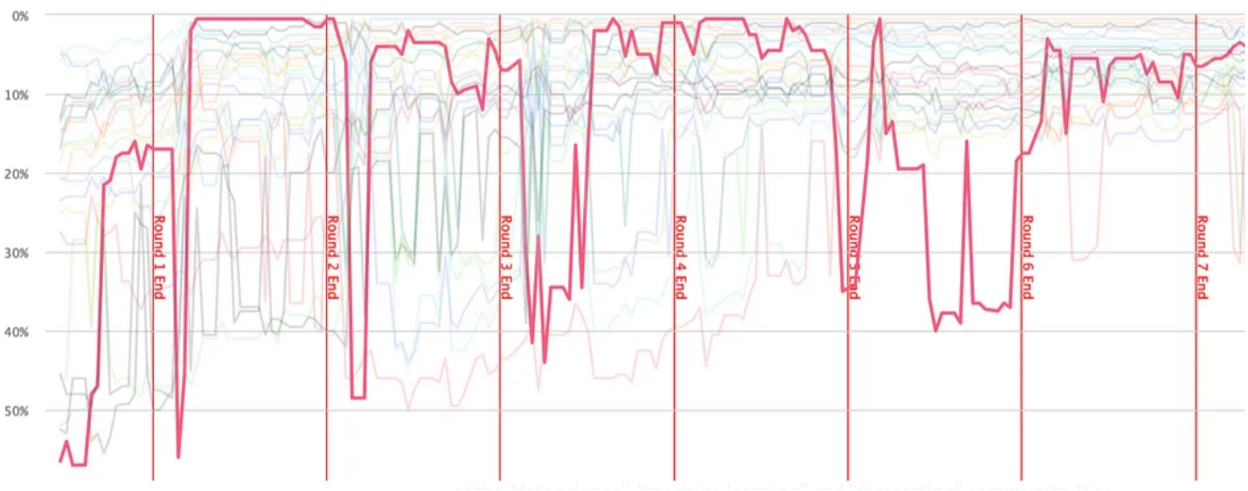

Figure 2: Less than steady progression of a competing top-ten entry in the overall contest. Once again the scale on the x-axis is approximately one year.

Position	Team	OR	RPS	Rank	IR	Rank
1	StanekF_STU (CERGE-EI)	6	0.15735	6	13.39013	6
2	MP - Miguel Pérez Michaus	13	0.15661	1	3.65944	25
3	Peters_STU	16.5	0.15801	11	4.77344	22
4	Innovation Team (Mathco)	22	0.15922	18	3.00801	26
5	Rogal Dorn	22.5	0.15968	27	6.58754	18
6	QuantMk6.ai	25	0.15979	30	4.97293	20
7	microprediction	27.5	0.15850	13	0.76081	42

Table 2: Summary of team positions and performance metrics collated near the end of the contest. Here OR stands for overall rank: the average of the rankings on either side of the contest. RPS is rank probability score. IR is information ratio, and largely implausible for all but the microprediction entry.

8. Related observations

In relating our findings and main recommendation to those of Makridakis et al. (2023), we note that hypothesis 7 made by organizers prior to the contest stated that "teams will be measurably overconfident in the accuracy of their forecasts". Their subsequent analysis showed this to be true in several ways including the comparison of frequency against entrant probabilities reproduced in Figure 3.

Shrinkage may have ameliorated that situation and was advocated before the contest (also in Cotton (2021b)). Failure to adequately shrink to was anticipated by organizers. This referred to shrinkage towards the uniform distribution for rank probabilities (for example). However in a similar vein, participants conscious of what we have termed the Indispensable Markets Hypothesis might have been inclined to shrink towards a different anchor: the relative volatilities supplied by the options market.

The M6 Competition served as a litmus test for whether people in quantitative fields would race to use market prices as data. That possibility would appear to be soundly rejected.

There are likely some other matters at play. We speculate that some and possibly many participants eschewed explicit simulation when determining rank probabilities. It seems plausible that many simply used "mental models", poorly motivated rules of thumb, or heuristics for probability generation - as famously advocated in Taleb (2007). A survey of participants would be interesting, in this regard.

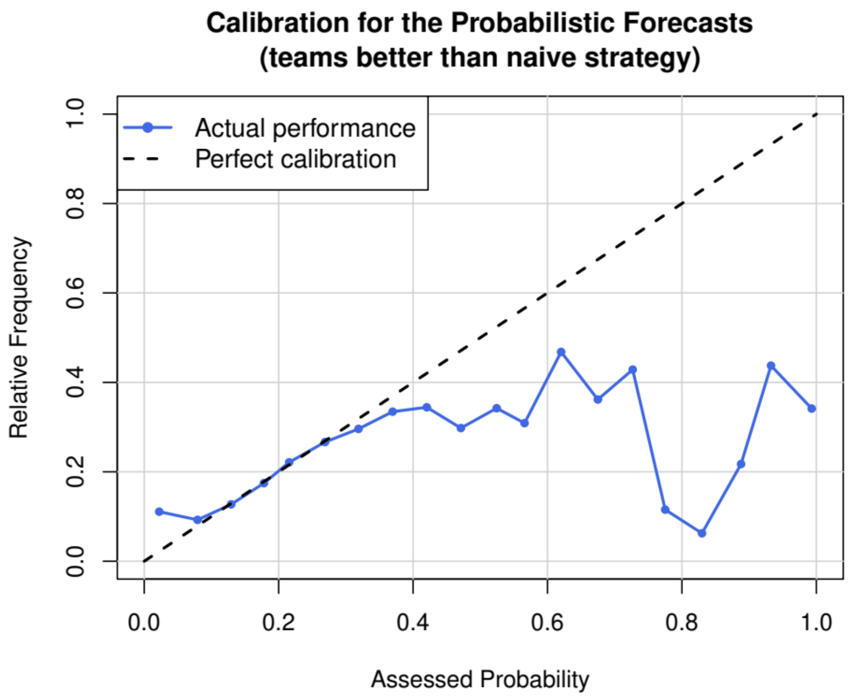

Figure 3: Comparison of frequency versus assessed probability for rank quintile probabilities in the M6 Competition reproduced from Makridakis et al. (2023).

Turning to another hypothesis made by organizers, the 8'th pre-competition conjecture stated that "averaging forecast rankings (investment weights) across all teams for each asset will yield rankings (weights) that outperform those of the majority of the teams." Subsequently in Makridakis et al. (2023) it is noted that "averaging the investment weights of the top 5 or 10% of the teams according to the IR results to significantly better performance than the best performing team".

This finding would ordinarily be interpreted only as a statement about modeling. However if we compare it to the parimutuel example mentioned in Section 1 we notice a possible comparison to markets. In the log-utility agent model, as described, the market state prices are wealth-weighted combinations of the subjective opinions of investors.[10]

So in this sense, the ex-post averaging of participants gives a hint as to the result of an experiment that was *not* run. For, it might have been possible to run a contest like the M6 Competition

[10]This mathematical comment ostensibly applies to a horse race, but is rather easily translated to options markets when we replace horses with Arrow-Debreu securities that can combine in linear combinations to replicate, exactly or approximately, options payouts.

side by side with a more market-styled tournament (i.e. with some clearing mechanism and incentive compatible reward) and then compare the result against a pre-specified meta-model applied to the former.

9. Remark on luck versus skill in investment

It is difficult to argue that the success of the microprediction entry was purely a matter of chance. It is easier to argue that luck played a role in the very few cases where other entrants finished ahead of it in the overall contest.

Toward the end of the contest, a snapshot of the leaderboard was saved for the purpose of a blog article. This update is shown in Table 2 and, at that time, the microprediction entry was in 7'th place overall. However this table reveals a striking fact: every single participant beating the microprediction entry had an information ratio in excess of 3, versus only 0.76 for the microprediction entry.[11]

A fund would have a hard time convincing institutional investors that ratios above 3 are likely to be repeated, certainly if such a hypothetical manager were to be subject to the stringent investment constraints implied by the contest structure: namely a restriction of the trading universe to 100 predetermined assets and a requirement that trading occur once a month!

So, setting aside the fact that the median entrant was roundly defeated by the microprediction entry one is also inclined to reject the idea that *any* team scored a believably repeatable victory over this option market benchmark in the overall competition - which is not to say that they might not have done so given a much longer time period or fewer constraints.

10. Remark on luck versus skill in prediction

One might also consider the role of luck versus skill in the prediction side of the contest in various ways, and we suggest a simple thought experiment as follows.

1. Simulate a true covariance matrix, then
2. Assume player 1 estimates rank probabilities based on empirical covariance from n_1 samples from the true matrix, and

[11] As a minor point, the information ratios reported were relative to a zero return and thus departed from the custom.

3. Assume player 2 estimates rank probabilities based on empirical covariance from n_2 samples from the same true matrix, and
4. Simulate the actual outcome of such a contest multiple times, compute Brier scores, and determine how often one player beats the other.

This need not be interpreted literally. The difference between n_1 and n_2 can represent superior technique (or wiser use of exogenous data).[12] We might cautiously inform our views about luck versus skill in rank probability estimation with experiments of this kind although one quickly discovers that this only makes the task of explaining the microprediction entry's success in all five stages of the contest more difficult, not less.

One struggles to defend the notion that the entry of *microprediction* prevailed due to only slightly better relative volatility knowledge. Elementary numerical experiments such as the one we suggest establish that a large skill gap is required. Avoiding a full Bayesian analysis, let us merely consider the case where the first player beats the second player quarterly on 75% of occasions (whereupon the observed evidence in Table 1 might almost be believed). In order that covariance knowledge alone provide the differentiating factor we require a skill (or knowledge) gap is *equivalent* to one player seeing 5000 samples of the truth and the other player a measly, rank-deficiency inducing 25 samples.

As an aside, many people would expect a much smaller skill difference to be sufficient and that incorrect intuition is likely to colour debate about the M6 Competition. Table 3 reports a Linked-in poll where participants were provided one such hypothetical experiment and asked to estimate the better player's chance of winning (when $n_1 > n_2$ but only by ten percent). Many respondents believed, incorrectly, that this marginally superior knowledge would almost certainly be rewarded.

11. Tentative conclusions on modeling quality absent an ingrained appreciation for the Indispensable Markets Hypothesis

Whatever the reasons for the success of the options proxy, this experimental result gives us pause. Is it a hint of the quality of data gathering writ large? Is probabilistic modeling of this

[12] Steps 2 and 3 above are not intended as advice against shrinkage. The only intent is to establish a skill difference. See Ledoit & Wolf (2004)

Table 3: Poll Results: Probability that the better informed person wins.

Option	Percentage
About 51%	30%
About 60%	24%
About 95%	30%
Dude, you have a bug	16%

Table 4: Results from a poll of size 168 when participants were asked to estimate the probability of one player beating another in a stylized M6-like contest where one player's knowledge of covariance is equivalent to $n_1 = 1100$ samples from the true distribution and the other player's ability is equivalent to only $n_2 = 1000$ samples. Approximately 1/3 of the participants chose the correct answer of 51%, but just as many grossly overestimated it.

nature typically performed at a relatively low standard (relative to markets) and this paucity only revealed by occasional accidental alignment with a market? If a hidden quality crisis in data science made plain only when the work is proximate in this sense, we might need more markets!

While the M6 Financial Forecasting Competition was ostensibly a test of the Efficient Markets Hypothesis applied to the equity markets, it was nearer to being a test of the EMH applied to the *options* market. And more precisely, it was an excellent test of the Indispensable Markets Hypothesis - a variant we introduce that is weaker but also more broadly applicable.

In a potential future experiment, a fair test between human orchestrated prediction and a market could be arranged that is similar to the M6 Competition but with one key difference. In this augmented design, participants would describe their methdologies to a committee of experts *before* the start of the competition. That committee would decide how to weight the entrants' contributions, again *before the competition*, and then a head-to-head out-of-sample comparison could proceed over the course of a year or longer between the market and this mildly bureaucratic meta-model.

In another hypothetical experiment, half of the participants could be secretly provided options market data or encouraged to use it. (Unfortunately, the cat might be out of the bag as far as the relevance of the options market is concerned, and we might need to wait for the next opportunity when some other data science competition, by chance, wanders into the gravitational pull of a highly relevant market.)

We have also sought with the definition of the Strong Indispensable Markets Hypothesis to provoke discussion on a broader notion of forecasting, one that characterizes that task by the objective and not the means by which it is done. From this vantage point the act of forecasting might encompass modeling work but also other kinds of engineering intended to avoid the inherent limitations of reliance on a single mind - such as creating a market or some other apparatus encouraging collective intelligence.

It is perhaps a shame that engineering activities are regarded as more cumbersome than pure modeling, but with appropriate tooling that could change. Moreover, the creation of market-like mechanisms can have an ongoing medium-term return on investment, whereas models require constant maintenance to remain close to the state of the art.

Perhaps someday this will all seem obvious, because the construction of forecasting pipelines that are designed to allow anyone to improve them without asking permission (as with markets) will be commonplace. At that time, there will be abundant evidence for their efficacy or otherwise, and attention might turn to the question of which collective intelligence contraptions work better than others.

However, advancing that possibility presents a Catch-22 because some, if not most, in the forecasting community must first be convinced by evidence. The old tradition of prediction markets has not proven sufficient (Wolfers & Zitzewitz (2006), Horn, Ivens & Brem (2014)) it would seem despite early advocacy from luminaries in economics Arrow, Forsythe, Gorham, Hahn, Hanson, Ledyard, Levmore, Litan, Milgrom, Nelson et al. (2008). Evidence of significant statistical power is harder to arrange for prediction markets as they are usually conceived, since these often are viewed as a tool for singular events like elections rather than ongoing operational needs.

Ability to experiment is hampered by regulation and censoring of evidence for the *indispensable* nature of markets is inevitable. A paper with a negative result citing the inability of one author to *completely replace* the predictive power of a market is unlikely to be deemed interesting - whereas tests of the Efficient Markets Hypothesis, such as various "proofs" of rent extraction (a much easier task) are commonplace.

Here we present a negative result for a large number of forecasters simultaneously. In the future also, competitions might present our best opportunity to further evaluate the Indispensable Markets Hypothesis - but only if the task at hand is sufficiently similar to, and informed by, an existing market and only if a large number of entrants eschew market prices as data. These are rare and necessarily flawed comparisons, accidentally arranged as with the M6 Competition, yet as scientists we should take what evidence is presented.

The M6 Competition prediction task is not a *perfect* match to the information presented in the options market - as noted - but on the other hand it seems unwise to take that imperfection as an excuse to wait for a better alignment (akin to waiting for Halley's comet to return given the infrequency of these competitions and the likelihood they come close to a market). Furthermore if we choose to wait for an even more precise match between a forecasting task and a market, in order to perform a test that is free of second order differences, then we may be disappointed. Participants might find the connection too obvious.

That is why the comparison between the options market proxy and efforts by a forecasting community should not devolve into the minutiae of decisions made in mapping the former to the latter. Clearly, in the construction we have described, a method that essentially replaced forecasting completely with implied volatilities, outperformed the vast majority of entrants whom we *reasonably presume* did not take advantage of the same possibility.

The message on the investing side might be equally strong, but where is the bar? For example, is beating 97% of entrants sufficient to make our point? Or should failure of the options market proxy to finish in first place amongst two hundred justify rejection of the weak IMH and also neglect of the strong? That would result in a continuation of what we regard as group think in forecasting: namely that models are the only thing you can construct using ten fingers and a keyboard that serve a predictive purpose.

The great temptation is to cherry pick one or two submissions *after* the M6 tournament is now complete and compare them to the single market benchmark we provided - one that was, we have argued, reasonably representative of the *only* market that could serve this purpose. That retrospective error would be especially problematic given the relatively short investment horizon, monthly holding periods, and lack of factor neutralization in the M6 Competition, some of which are acknowledged in post-competition discussion by organizers (Makridakis et al. (2023)).

To conclude, we assert that the (weak) Indispensable Market Hypothesis was strongly supported by the M6 competition. We think a reasonable judgement, notwithstanding the issues mentioned, is that standalone modeling is not a full substitute for the predictive power of a market in this instance. That would appear to be true even in the unusual circumstance where the same task is undertaken by a diverse population of quantitative people from around the world (in the presence of a healthy prize-money budget).

As a final remark, the prediction side of the M6 Competition was very well designed to test the Weak IMH because contestants could not "sit out" (they were forced to provide quintile probabilities for every asset in every period). As a group, those contestants were bested by a simple proxy for the options market and worse, that proxy could *probably* have been improved in obvious ways (say by using implied distributions implied by the collection of all strikes, not just near the money volatilities).[13]

12. Conclusion on data search

The apparent victory for the market implicit in the success of the *microprediction* entry does not establish the Efficient Markets Hypothesis applied to the options market, only the weaker Indispensable Markets Hypothesis we have defined. But more pertinently, if participants were truly sympathetic to *either*, it should have been automatic for them to look for proximate market predictions. What can excuse the failure to locate even the one, single most important piece of exogenous data - the options market - never mind *all* or most relevant exogenous data?

The answer is that nothing can excuse it. Even if it were to be revealed that several other teams did use options market data (not the case to our knowledge) what possible relevance would that have to companies building data science teams? Are they to turn a blind eye to the fact that 95% (say) of data scientists might not locate the first important piece of exogenous information when performing a modeling task? What are their odds of choosing the data scientists who will find *most* relevant exogenous data for any given task?

Our conclusion must be that the Efficient Markets Hypothesis is so tainted that it fails to serve as a critical reminder of the power of markets. An interesting question for further discussion or study is *why* application the Indispensable Markets Hypothesis, in contrast, isn't obvious or even second nature in the broad community (as compared with some much narrower quant communities). Perhaps, as we have suggested, it is common to mentally compartmentalize market data points (like implied volatilities) from other data (such as lagged values). If that is the case, it is all the more

[13]However we shy away from an ex post assessment given the obvious danger in the same. Using implied distributions was considered beforehand, but would have required more time and better data.

important that students are taught to value the Indispensable Markets Hypothesis, by whatever name.

Solving the question of why a blindspot might exist, or trying to educate data scientists, should not hold up development of a cure by other means. A medium-term solution to data search mediocrity (which quite possibly is widespread if this competition result is extrapolated) might involve replacing or augmenting conventional project-based data science with pipelines that are infused with market-inspired mechanisms. Other contraptions might be far more granular in their use of the IMH than existing markets or prediction markets. See Cotton (2022a) for extended discussion. A simple operating example is provided at the time of writing by the "monte carlo market" described in Cotton (2021a). This apparatus effects probabilistic prediction of sector returns every week, and anyone is free to find a new source of relevant exogenous data.

Whatever remedies are attempted, we are as a result of the M6 Competition more confident in making the general assertion that the set of data used in any given prediction task should be determined collectively by the most inventive authors or algorithms, not left to the apparently limited peripheral vision of a single data scientist or small team. This suggests more attention might be paid to crowd-sourcing mechanism design (as in Goel, Nikzad & Singla (2014), Jhang-Li & Chiang (2023), Li (2017) for example) versus algorithm design alone, and that conclusion does not rest on the finer details of the *microprediction* entry, only on the importance of locating relevant exogenous data.

Appendix: Schur complementary portfolio construction

In the interest of completeness, and because the method is not published, a sketch of the Schur complementary portfolio construction used in the microprediction entry is provided. The method is similar to Hierarchical Risk Parity (HRP) but uses a more principled way to split and sub-allocate capital in a top-down fashion.

In De Prado (2016), where HRP was introduced, the top-down allocation scheme is described in procedural fashion. A terse recursive expression will serve us better, where intra-group allocations $w(A)$ and $w(D)$ are determined by a method $w()$ that calls out to a terminating portfolio algorithm $w_{term}()$ once the dimension is sufficiently small. Inter-group allocation is determined by a group fitness function such as robustly estimated inverse group covariance $1/\nu$. We write

$$w(\Sigma) \propto \begin{cases} \begin{pmatrix} \frac{1}{\nu(A)} w(A) \\ \frac{1}{\nu(D)} w(D) \end{pmatrix} & \text{if } \dim(\Sigma) > m, \\ w_{\text{term}}(\Sigma) & \text{otherwise.} \end{cases} \tag{1}$$

to represent the overall allocation (left hand side) and it is implicit that the right hand side will be normalized $\sum_i w_i = 1$. The notation A and D refers to submatrices of Σ:

$$\Sigma = \begin{pmatrix} A & B \\ C = B^T & D \end{pmatrix} \tag{2}$$

The novelty in Hierarchical Risk Parity is that grouping and inter-group allocation occurs *after a reordering of assets* that places similar securities close to one another. This seriation step, as it is called in statistics, is implicit in Equation 1. [14]

The splitting can occur multiple times, terminating after some rule is triggered (say less than five assets) with an application of ω_{term} that may or may not return a long-only portfolio. For avoidance of doubt, in 1 both $w(A)$ and $w(D)$ are column vectors and $\nu(A)$ and $\nu(D)$ are scalar.

The method used in the microprediction benchmark entry was quite similar in spirit but for newcomers to the methodology, this similarity might not be obvious in the code.[15]. The approach is best understood as a top-down scheme similar to HRP where the sub-covariance matrices are modified as we descend. But it makes use of the off-diagonal matrices B whereas Hierarchical Risk Parity ignores them. We write:

$$w(\Sigma) \propto \begin{pmatrix} \frac{1}{\nu(A')} w\left(A''\right) \\ \frac{1}{\nu(D')} w\left(D''\right) \end{pmatrix} \tag{3}$$

where, in contrast with 1, the $A' = A'(\Sigma)$ and $A'' = A''(\Sigma)$ are modifications of A designed to incorporate some information from the full covariance matrix Σ. Likewise D is also augmented in two different ways.

The next section motivates these augmentations with an example, and the code contains some parameters that allow it to span a continuum between minimum variance portfolio construction

[14] One applies the permutation to the covariance matrix Σ, determines w, and then applies the inverse permutation to solve the original problem.

[15] See `https://github.com/microprediction/m6entry/blob/main/m6entry/step2_portfolio.py` which in term references `https://github.com/microprediction/precise/blob/main/precise/skaters/portfoliostatic/schurport.py` from the precise Python package Cotton (2022b)

and Hierarchical Risk Parity. In the former extreme the sub-covariance matrix A is swapped out for an augmented matrix A'':

$$A'' = \frac{A - BD^{-1}C}{b_A b_A^T} \tag{4}$$

where pointwise division is indicated, and where

$$b_A = \overrightarrow{1} - AC^{-1}\overrightarrow{1}. \tag{5}$$

Similarly matrix A' is the result of "pointwise multiplication of the Schur complement by $b_A b_A^T$ in the precision domain", where the vector b_A was defined in 5. To expand that comment, we use the shorthand $(\cdot)^{*b}$ to denote the operation:

$$Q^{*b} := \left(Q^{-1} \cdot (bb^T)\right)^{-1} \tag{6}$$

then the inter-group allocation matrix sent to $\nu()$ can be written

$$A' = \left(A - BD^{-1}C\right)^{*b_A} \tag{7}$$

or even more briefly

$$A' = (A^c)^{*b_A}$$

where A^c denotes the Schur complement. The matrix D' is computed in analogous fashion. These new matrices will be used to determine the inter-group allocation ratio:

$$1/\nu(A') : 1/\nu(D').$$

where $\nu(\cdot)$ is the inverse fitness measure that, as with Hierarchical Risk Parity, is one of the components that can be chosen by the user. The differences between Schur portfolio construction and its predecessor Hierarchical Risk Parity are summarized in Table 5.

The smallest non-trivial example suffices to illustrate how 4 and 7 can reproduce the minimum variance portfolio whereas Hierarchical Risk Parity (if we limit ourselves to bisection) does not. Take:

$$\Sigma = \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix}$$

Allocation	HRP	Schur
Inter-group	A or $diag(A)$	$(A^c(\gamma)^{-1} \cdot b_A b_A^T)^{-1}$ where we set $A^c(\gamma) = A - \gamma BD^{-1}C$ and $b_A(\lambda) = \overrightarrow{1} - \lambda BD^{-1}\overrightarrow{1}$.
Intra-group	A	$\left(A - \gamma BD^{-1}C\right) / \left(b_A b_A^T\right)$ element-wise division

Table 5: Comparison of covariance matrices used to allocate capital between and within a group of securities in two divide-and-conquer schemes: HRP as proposed in De Prado (2016) and Schur Hierarchical Portfolios as proposed herein.

The desired portfolio is the symmetric one, all else being equal. However bisection must break the symmetry. Without loss of generality we split $\{1, 2\}, \{3\}$ and, consequently, the covariance splits per 2 into

$$A = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, B = \begin{pmatrix} \rho \\ \rho \end{pmatrix}, C = \begin{pmatrix} \rho, \rho \end{pmatrix}, D = \begin{pmatrix} 1 \end{pmatrix}$$

Most hierarchical methods will use a sub-allocation within $\{1, 2\}$ that will allocate evenly amongst the two assets (since A is symmetric). Then every dollar allocated towards this part of the portfolio, as compared with the third asset, incurs variance $v_A = \frac{1+\rho}{2}$. A dollar invested in the third asset incurs unit variance, naturally. Thus a seemingly reasonable hierarchical portfolio allocation assigning capital inversely proportional to variance will lead to an inter-group portfolio allocation:

$$\frac{2}{3+\rho} : \frac{1+\rho}{3+\rho}$$

Then, splitting the allocation to $\{1, 2\}$ in half we have

$$w = \frac{1}{3+\rho} \begin{pmatrix} 1 \\ 1 \\ 1+\rho \end{pmatrix}$$

So, despite the reasonableness of this methodology, it evidently over-allocates to asset 3 when $\rho > 0$ and under-allocates when $\rho < 0$.

This inelegance is fixed in the new approach. We can see explicitly how this is accomplished, and how it relates to the matrix inversion identity motivating 3, 4 and 7. We compute the Schur complement inverse:

$$(A^c)^{-1} = \frac{1}{\phi(\rho)} \begin{pmatrix} 1+\rho & -\rho \\ -\rho & 1+\rho \end{pmatrix}$$

where $\phi(\rho) := 1 + \rho - 2\rho^2$. Completing the rest of the algebra the identity

$$\Sigma^{-1} = \begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} (A^c)^{-1} & 0 \\ 0 & (D^c)^{-1} \end{pmatrix} \begin{pmatrix} 1 & -BD^{-1} \\ -CA^{-1} & 1 \end{pmatrix}$$

manifests in this example as

$$\begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix}^{-1} = \frac{1+\rho}{\phi} \begin{pmatrix} 1 & -\frac{\rho}{1+\rho} & 0 \\ -\frac{\rho}{1+\rho} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & -\rho \\ 0 & 1 & -\rho \\ -\frac{\rho}{1+\rho} & -\frac{\rho}{1+\rho} & 1 \end{pmatrix}$$

The right hand side has broken symmetry but we can check that weights of the minimum variance portfolio are equal:

$$w \propto \Sigma^{-1}\overrightarrow{1} \propto \begin{pmatrix} 1 & -\frac{\rho}{1+\rho} & 0 \\ -\frac{\rho}{1+\rho} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & -\rho \\ 0 & 1 & -\rho \\ -\frac{\rho}{1+\rho} & -\frac{\rho}{1+\rho} & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \frac{1-\rho}{1+\rho} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

as they should be.

Acknowledgements

During the course of the contest I was employed by Intech Investments, who also served as a sponsor for the contest and supported open-source projects useful in this context. I thank JB Kirkland for creating the rank visualizer used in Figures 1 and 2 that also made the contest more fun. Needless to say, hats off to organizers for providing year-long entertainment and discussion.

References

Arrow, K. J., Forsythe, R., Gorham, M., Hahn, R., Hanson, R., Ledyard, J., Levmore, S., Litan, R., Milgrom, P., Nelson, F. et al. (2008). The promise of prediction markets. *Science*, *320*, 877–878.

Barberis, N., & Thaler, R. (2003). A survey of behavioral finance. In *Handbook of the Economics of Finance* (pp. 1053–1128). Elsevier volume 1.

Bollerslev, T., & Todorov, V. (2011). Tails, fears, and risk premia. *The Journal of Finance*, *66*, 893–927.

Broadie, M., Chernov, M., & Johannes, M. (2009). Understanding index option returns. *The Review of Financial Studies*, *22*, 4493–4529.

Cotton, P. (2013). Keeping punters (log) happy: some thoughts on the total cost of providing market efficiency. URL: `https://finmathblog.blogspot.com/2013/11/keeping-punters-log-happy-some.html` accessed: 2024-07-22.

Cotton, P. (2019). Self Organizing Supply Chains for Micro-Prediction: Present and Future uses of the ROAR Protocol.

Cotton, P. (2021a). A new generative model competition to test the old against the new. *Medium*, . URL: `https://microprediction.medium.com/a-new-generative-model-competition-to-test-the-old-against-the-new-189d3d460669`.

Cotton, P. (2021b). Six ways to train for the m6 financial forecasting competition. *Medium*, . URL: `https://medium.com/geekculture/six-ways-to-train-for-the-m6-financial-forecasting-competition-cacaf3af58b5`.

Cotton, P. (2022a). *Microprediction: Building an Open AI Network*. Cambridge, MA: MIT Press.

Cotton, P. (2022b). Precise: A Python Package for Streaming Covariance Prediction. URL: `https://github.com/microprediction/precise`.

Cotton, P. (2023). The options market beat 97% of participants in the m6 financial forecasting contest. URL: `https://medium.com/geekculture/the-options-market-beat-94-of-participants-in-the-m6-financial-forecasting-contest-fa4f47f57d33`.

De Prado, M. L. (2016). Building diversified portfolios that outperform out of sample. *Journal of Portfolio Management*, *42*. doi:`10.3905/jpm.2016.42.4.059`.

DeMiguel, V., Garlappi, L., Nogales, F. J., & Uppal, R. (2009). A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Science*, *55*. doi:`10.1287/mnsc.1080.0986`.

Driessen, J., Maenhout, P., & Vilkov, G. (2009). Option-implied correlations and the price of correlation risk. *The Review of Financial Studies*, *22*, 3133–3164.

Falkenstein, E. G. (2012). *The Missing Risk Premium: Why Low Volatility Investing Works*. CreateSpace Independent Publishing Platform.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, *25*, 383–417.

Farmer, J. D., & Skouras, S. (2012). An ecological perspective on the future of computer trading. *Quantitative Finance*, *12*, 377–397.

Goel, A., Nikzad, A., & Singla, A. (2014). Mechanism design for crowdsourcing markets with heterogeneous tasks. In *Proceedings of the 23rd International Conference on World Wide Web* (pp. 521–522). ACM. URL: `https://las.inf.ethz.ch/files/goel14mechanismdesign.pdf`.

Hayek, F. (1945). The Use of Knowledge in Society FA Hayek. *American Economic Review*, . doi:`10.1257/aer.98.5.i`.

Henwood, D. (1997). *Wall Street: How It Works and for Whom*. Verso.

Horn, S., Ivens, B. S., & Brem, A. (2014). Prediction markets: A literature review. *Journal of Prediction Markets*, *8*, 89–126. URL: `https://www.uni-bamberg.de/fileadmin/bwl-marketing/Bilder/Publikationen/`

Prediction_Markets_a_literature_review_-__JOPM-_8_2_5_Horn_et_al.pdf.

Jain, P., & Jain, S. (2019). Can machine learning-based portfolios outperform traditional risk-based portfolios? The need to account for covariance misspecification. *Risks*, *7*. doi:10.3390/risks7030074.

Jhang-Li, J.-H., & Chiang, I.-R. (2023). An analysis of reward mechanism and knowledge sharing for crowdsourcing-based open innovation contest. *Kybernetes*, *ahead-of-print*. URL: https://doi.org/10.1108/K-03-2023-0350. doi:10.1108/K-03-2023-0350.

Lan, C. (2015). An Out-of-Sample Evaluation of Dynamic Portfolio Strategies. *Review of Finance*, *19*. doi:10.1093/rof/rfu052.

Ledoit, O., & Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, *88*. doi:10.1016/S0047-259X(03)00096-4.

Li, Z.-H. (2017). Crowd intelligence in ai 2.0 era. *ResearchGate*, . URL: https://www.researchgate.net/publication/313270428_Crowd_intelligence_in_AI_20_era.

Makridakis, S., Spiliotis, E., Hollyman, R., Petropoulos, F., Swanson, N., & Gaba, A. (2023). The m6 forecasting competition: Bridging the gap between forecasting and investment decisions. *Journal Name Here*, *Volume Number Here*, Page Numbers Here. Submitted on 20 Oct 2023.

Molyboga, M. (2020). A Modified Hierarchical Risk Parity Framework for Portfolio Management. *Journal of Financial Data Science*, *2*. doi:10.3905/jfds.2020.1.038.

Philippon, T. (2015). Has the us finance industry become less efficient? on the theory and measurement of financial intermediation. *American Economic Review*, *105*, 1408–38.

Taleb, N. N. (2007). *The Black Swan: The Impact of the Highly Improbable*. Random House.

University of Nicosia (2024). The m6 competition. https://mofc.unic.ac.cy/the-m6-competition/.

Viole, F. (2020). Nowcasting with nns. *SSRN*, . 10 Pages Posted: 5 Jun 2020 Last revised: 11 Mar 2021.

Vovk, V., Gammerman, A., & Shafer, G. (2005). *Algorithmic learning in a random world*. Springer Science & Business Media.

Wolfers, J., & Zitzewitz, E. (2006). Prediction markets in theory and practice. *NBER Working Paper Series*, *12083*.