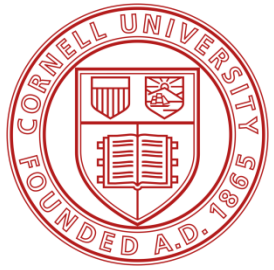


Advances in Financial Machine Learning: Lecture 6/10

Prof. Marcos López de Prado

Advances in Financial Machine Learning

ORIE 5256

What are we going to learn today?

- Backtest Statistics
  - General Characteristics
  - Performance
  - Time-Weighted Rate of Return
  - Drawdown and Time Under Water
  - Implementation Shortfall
  - Efficiency
- Type I and Type II Errors under Multiple Testing
- Understanding Strategy Risk
  - Symmetric Payouts
  - Asymmetric Payouts
  - The Probability of Strategy Failure

Backtest Statistics

General Statistics

- **Time range**: Time range specifies the start and end dates.
- **Average AUM**: This is the average dollar value of the assets under management.
- **Capacity**: A strategy's capacity can be measured as the highest AUM that delivers a target risk-adjusted performance.
- **Leverage**: Leverage measures the amount of borrowing needed to achieve the reported performance.
- **Maximum dollar position size**: Maximum dollar position size informs us whether the strategy at times took dollar positions that greatly exceeded the average AUM.
- **Ratio of longs**: The ratio of longs show what proportion of the bets involved long positions.
- **Frequency of bets**: The frequency of bets is the number of bets per year in the backtest.
- **Average holding period**: The average holding period is the average number of days a bet is held.
- **Annualized turnover**: Annualized turnover measures the ratio of the average dollar amount traded per year to the average annual AUM.
- **Correlation to underlying**: This is the correlation between strategy returns and the returns of the underlying investment universe.

Performance

- **PnL**: The total amount of dollars (or the equivalent in the currency of denomination) generated over the entirety of the backtest, including liquidation costs from the terminal position.
- **PnL from long positions**: The portion of the PnL dollars that was generated exclusively by long positions.
- **Annualized rate of return**: The time-weighted average annual rate of total return, including dividends, coupons, costs, etc.
- **Hit ratio**: The fraction of bets that resulted in a positive PnL.
- **Average return from hits**: The average return from bets that generated a profit.
- **Average return from misses**: The average return from bets that generated a loss.

Time-Weighted Rate of Return (1/2)

- The TWRR for portfolio *i* between subperiods $[t-1, t]$ is denoted $r_{i,t}$, with equations

$$r_{i,t} = \frac{\pi_{i,t}}{K_{i,t}}; \pi_{i,t} = \sum_{j=1}^{J} \left[\left(\Delta P_{j,t} + A_{j,t} \right) \theta_{i,j,t-1} + \Delta \theta_{i,j,t} \left(P_{j,t} - \bar{P}_{j,t-1} \right) \right]$$

$$K_{i,t} = \sum_{j=1}^{J} \tilde{P}_{j,t-1} \theta_{i,j,t-1} + \max \left\{ 0, \sum_{j=1}^{J} \bar{\tilde{P}}_{j,t} \Delta \theta_{i,j,t} \right\}$$

where

- $\pi_{i,t}$ is the mark-to-market (MtM) profit or loss for portfolio *i* at time *t.*
- $K_{i,t}$ is the market value of the assets under management by portfolio *i* through subperiod *t.* The purpose of including the $\max\{.\}$ term is to fund additional purchases (ramp-up).
- $A_{j,t}$ is the interest accrued or dividend paid by one unit of instrument *j* at time *t.*
- $P_{j,t}$ is the clean price of security *j* at time *t.*
- $\theta_{i,j,t}$ are the holdings of portfolio *i* on security *j* at time *t.*

Time-Weighted Rate of Return (2/2)

… where (continued)

- $\tilde{P}_{j,t}$ is the dirty price of security *j* at time *t.*
- $\bar{P}_{j,t}$ is the average transacted clean price of portfolio *i* on security *j* over subperiod *t.*
- $\bar{\tilde{P}}_{j,t}$ is the average transacted dirty price of portfolio *i* on security *j* over subperiod *t.*

- Inflows are assumed to occur at the beginning of the day, and outflows are assumed to occur at the end of the day. These sub-period returns are then linked geometrically as

$$\varphi_{i,T} = \prod_{t=1}^{T}(1 + r_{i,t})$$

- The variable $\varphi_{i,T}$ can be understood as the performance of one dollar invested in portfolio *i* over its entire life, $t = 1, \ldots, T$. Finally, the annualized rate of return of portfolio *i* is

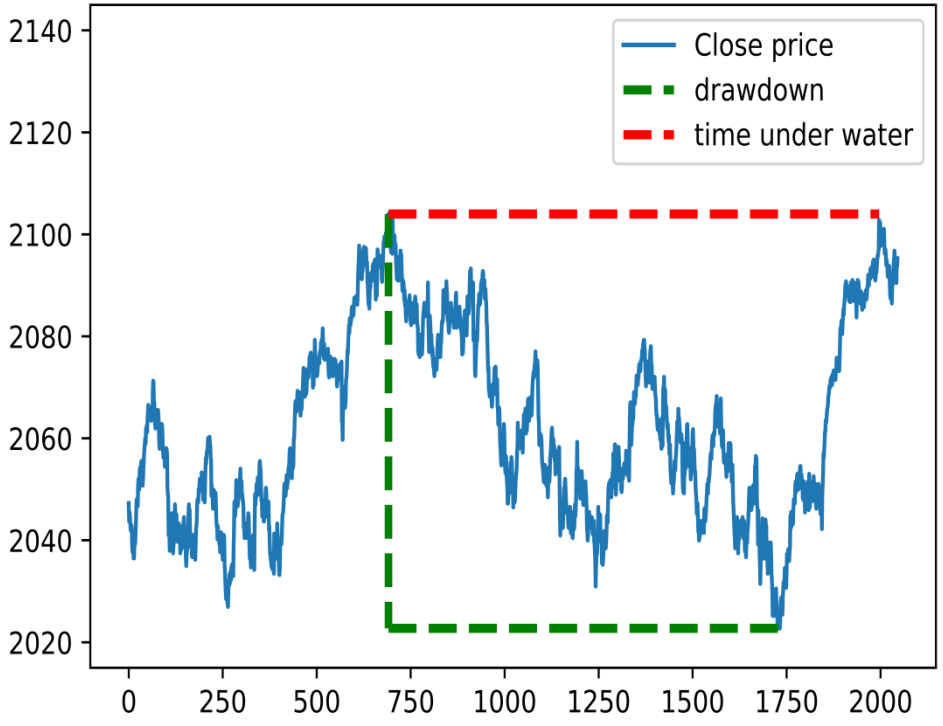
$$R_i = \left(\varphi_{i,T}\right)^{-y_i} - 1$$

where y_i is the number of years elapsed between $r_{i,1}$ and $r_{i,T}$.

Drawdown and Time Under Water

- Intuitively, a **drawdown** (DD) is the maximum loss suffered by an investment between two consecutive high-watermarks (HWMs).
- The **time under water** (TuW) is the time elapsed between an HWM and the moment the PnL exceeds the previous maximum PnL.

```
def computeDD_TuW(series,dollars=False):
    # compute series of drawdowns and the time under water associated with them
    df0=series.to_frame('pnl')
    df0['hwm']=series.expanding().max()
    df1=df0.groupby('hwm').min().reset_index()
    df1.columns=['hwm','min']
    df1.index=df0['hwm'].drop_duplicates(keep='first').index  # time of hwm
    df1=df1[df1['hwm']>df1['min']] # hwm followed by a drawdown
    if dollars:dd=df1['hwm']-df1['min']
    else:dd=1-df1['min']/df1['hwm']
    tuw=((df1.index[1:]-df1.index[:-1])/np.timedelta64(1,'Y')).values # in years
    tuw=pd.Series(tuw,index=df1.index[:-1])
    return dd,tuw
```



Implementation Shortfall

- **Broker fees per turnover**: These are the fees paid to the broker for turning the portfolio over, including exchange fees.
- **Average slippage per turnover**: These are execution costs, excluding broker fees, involved in one portfolio turnover.
- **Dollar performance per turnover**: This is the ratio between dollar performance (including brokerage fees and slippage costs) and total portfolio turnovers.
- **Return on execution costs**: This is the ratio between dollar performance (including brokerage fees and slippage costs) and total execution costs.

Efficiency

- **Annualized Sharpe ratio**: This is the SR value, annualized by a factor $\sqrt{a}$, where a is the average number of returns observed per year.
- **Information ratio**: This is the SR equivalent of a portfolio that measures its performance relative to a benchmark.
- **Probabilistic Sharpe ratio**: PSR corrects SR for inflationary effects caused by non-Normal returns or track record length.
- **Deflated Sharpe ratio**: DSR corrects SR for inflationary effects caused by non-Normal returns, track record length, and selection bias under multiple testing.

Sharpe [1966]

- Consider an investment strategy with excess returns (or risk premia) $\{r_t\}$, $t = 1, \ldots, T$, which follow an IID Normal distribution,

$$r_t \sim \mathcal{N}[\mu, \sigma^2]$$

where $\mathcal{N}[\mu, \sigma^2]$ represents a Normal distribution with mean μ and variance σ^2.

- The SR (non-annualized) of such strategy is defined as

$$SR = \frac{\mu}{\sigma}$$

- Because parameters μ and σ are not known, SR is estimated as

$$\widehat{SR} = \frac{\mathrm{E}[\{r_t\}]}{\sqrt{\mathrm{V}[\{r_t\}]}}$$

Lo [2002]

- Under the assumption that returns follow an IID Normal distribution, Lo [2002] derived the asymptotic distribution of $\widehat{SR}$ as

$$\left(\widehat{SR} - SR\right) \overset{a}{\rightarrow} \mathcal{N}\left[0, \frac{1 + \frac{1}{2}SR^2}{T}\right]$$

- Under the assumption that returns follow an IID non-Normal distribution, Mertens [2002] derived the asymptotic distribution of $\widehat{SR}$ as

$$\left(\widehat{SR} - SR\right) \overset{a}{\rightarrow} \mathcal{N}\left[0, \frac{1 + \frac{1}{2}SR^2 - \gamma_3 SR + \frac{\gamma_4 - 3}{4}SR^2}{T}\right]$$

where γ_3 is the skewness of $\{r_t\}$, and γ_4 is the kurtosis of $\{r_t\}$ ($\gamma_3 = 0$ and $\gamma_4 = 3$ when returns follow a Normal distribution).

Bailey and López de Prado [2012] (1/2)

- Christie [2005] and Opdyke [2007] discovered that, in fact, the Mertens [2002] equation is also valid under the more general assumption that returns are stationary and ergodic (not necessarily IID).
- Bailey and López de Prado [2012] utilized those results to derive the Probabilistic Sharpe Ratio (PSR).
- PSR estimates the probability that an observed $\widehat{SR}$ exceeds SR^* as

$$\widehat{PSR}[SR^*] = Z\left[\frac{\left(\widehat{SR} - SR^*\right)\sqrt{T-1}}{\sqrt{1 - \hat{\gamma}_3\widehat{SR} + \frac{\hat{\gamma}_4 - 1}{4}\widehat{SR}^2}}\right]$$

where $Z[.]$ is the CDF of the standard Normal distribution, T is the number of observed returns, $\hat{\gamma}_3$ is the skewness of the returns, and $\hat{\gamma}_4$ is the kurtosis of the returns. Note that $\widehat{SR}$ is the non-annualized estimate of SR, computed on the same frequency as the T observations.

Bailey and López de Prado [2012] (2/2)

- For a given SR^*, $\widehat{PSR}$ increases with
  - greater mean returns ($\mathrm{E}[\{r_t\}]$)
  - lower variance of returns ($\mathrm{V}[\{r_t\}]$)
  - longer track records (T)
  - positively skewed returns ($\hat{\gamma}_3$)
  - thinner tails ($\hat{\gamma}_4$)
- This result also allows us to answer the question: *"How long should a track record be in order to have statistical confidence* $(1-\alpha)$ *that its estimated Sharpe ratio (*$\widehat{SR}$*) is above a given threshold (*SR^**)"* (minimum track record length)

$$MinTRL = 1 + \left[1 - \hat{\gamma}_3 \widehat{SR} + \frac{\hat{\gamma}_4 - 1}{4} \widehat{SR}^2\right] \left(\frac{Z_\alpha}{\widehat{SR} - SR^*}\right)^2$$

where Z_α is the value of the Standard Normal CDF that leaves a probability α in the right tail.

Bailey and López de Prado [2014] (1/2)

- The Deflated Sharpe Ratio computes the probability that the Sharpe Ratio (SR) is statistically significant, after controlling for the inflationary effect of multiple trials, data dredging, non-normal returns and shorter sample lengths.

$$\widehat{DSR} \equiv \widehat{PSR}\left(\widehat{SR}_0\right) = Z\left[\frac{\left(\widehat{SR} - \widehat{SR}_0\right)\sqrt{T-1}}{\sqrt{1 - \hat{\gamma}_3\widehat{SR} + \frac{\hat{\gamma}_4 - 1}{4}\widehat{SR}^2}}\right]$$

where $\widehat{SR}_0$ is the estimate provided by the False Strategy theorem,

$$\widehat{SR}_0 = \sqrt{\mathrm{V}\left[\left\{\widehat{SR}_k\right\}\right]}\left((1-\gamma)Z^{-1}\left[1 - \frac{1}{K}\right] + \gamma Z^{-1}\left[1 - \frac{1}{Ke}\right]\right)$$

- DSR packs more information than SR, and it is expressed in probabilistic terms.

Bailey and López de Prado [2014] (2/2)

- The standard SR is computed as a function of two estimates:
  - Mean of returns
  - Standard deviation of returns
- DSR deflates SR by taking into consideration five additional variables (it packs more information):
  - The non-Normality of the returns $(\hat{\gamma}_3, \hat{\gamma}_4)$
  - The length of the returns series (T)
  - The amount of data dredging $\left(\mathrm{V}\left[\{\widehat{SR}_k\}\right]\right)$
  - The number of independent trials involved in the selection of the investment strategy (K)

The key to preventing selection bias is to record all trials, and determine correctly the number of effectively independent trials (*K*).

Multiple Testing

The Neyman-Pearson Framework

- Under the standard Neyman-Pearson [1933] hypothesis testing framework:
  - We state a null hypothesis H_0, and an alternative hypothesis H_1
  - We derive the distribution of a test statistic under H_0 and under H_1
  - We reject H_0 with confidence $(1-\alpha)$ in favor of H_1 when we observe an event that, should H_0 be true, could only occur with probability α
- This framework is the statistical analogue to a "*proof by contradiction*" argument.
- Four probabilities associated with a predicted positive $(x > \tau_\alpha)$:
  - $\mathbf{P}\left[\boldsymbol{x} > \boldsymbol{\tau_\alpha} \middle| \boldsymbol{H_0}\right] = \boldsymbol{\alpha}$ is the Type I error probability, significance or false positive rate
  - $\mathbf{P}\left[\boldsymbol{x} > \boldsymbol{\tau_\alpha} \middle| \boldsymbol{H_1}\right] = \mathbf{1} - \boldsymbol{\beta}$ is the power, recall or true positive rate; $\mathrm{P}\left[x \leq \tau_\alpha \middle| H_1\right] = \beta$ is the Type II error probability or false positive rate
  - $\mathbf{P}\left[\boldsymbol{H_0} \middle| \boldsymbol{x} > \boldsymbol{\tau_\alpha}\right]$ is the false discovery rate (FDR)
  - $\mathbf{P}\left[\boldsymbol{H_1} \middle| \boldsymbol{x} > \boldsymbol{\tau_\alpha}\right]$ is the test's precision
- The p-value (or α) does <u>not</u> give the probability that the null hypothesis is true.

Familywise Error Rate (FWER)

- When Neyman and Pearson [1933] proposed this framework, they did not consider the possibility of conducting multiple tests and select the best outcome.
- When a test is repeated multiple times, the combined α increases.
- Consider that we repeat for a second time a test with false positive probability α.
  - At each trial, the probability of *not* making a Type I error is $(1-\alpha)$
  - If the two trials are independent, the probability of not making a Type I error on the first *and* second tests is $(1-\alpha)^2$
  - The probability of making *at least one* Type I error is the complementary, $1-(1-\alpha)^2$
- After a "family" of K independent tests, we reject H_0 with confidence $(1-\alpha)^K$.
- FWER the probability that *at least one* of the positives is false, $\alpha_K = 1-(1-\alpha)^K$.
- **The Šidàk Correction: For a given *K* and α_K, then $\alpha = 1-(1-\alpha_K)^{\frac{1}{K}}$.**

FWER vs. FDR

- Two definitions of Type I error in the multiple testing literature:
  - Familywise Error Rate (FWER): The probability that *at least one* false positive takes place.
  - False Discovery Rate (FDR): Expected value of the ratio of false positives to predicted positives.
- In most scientific and industrial applications, FWER is considered overly punitive.
  - For example, it would be impractical to design a car model where we control for the probability that a single unit will be defective.
- However, **in the context of finance, we advise against the use of FDR**.
  - The reason is, an investor does not typically allocate funds to all strategies with predicted positives within a family of trials, where a proportion of them are likely to be false.
  - Instead, investors are only introduced to the single best strategy out of a family of thousands or even millions of alternatives.
- **Investors have no ability to invest in the discarded predicted positives.**
  - Following the car analogue, in finance there is actually a single car unit produced per model, which everyone will use. If the only produced unit is defective, everyone will crash.

Type I Errors under Multiple Testing

Test Statistic

- Consider an investment strategy with returns time series of size T.
- We estimate the Sharpe ratio, $\widehat{SR} = \frac{\mathrm{E}\left[\{r_t\}_{t=1,\dots,T}\right]}{\sqrt{\mathrm{V}\left[\{r_t\}_{t=1,\dots,T}\right]}}$, where $\{r_t\}$ are excess returns.
- We define two hypothesis:
  - The null hypothesis, $H_0: SR = 0$
  - The alternative hypothesis, $H_1: SR > 0$
- Following Bailey and López de Prado [2012], if the true Sharpe ratio equals SR^*, the statistic $\hat{z}[SR^*]$ is asymptotically distributed as a Standard Normal,

$$\hat{z}[SR^*] = \frac{\left(\widehat{SR} - SR^*\right)\sqrt{T-1}}{\sqrt{1 - \hat{\gamma}_3 \widehat{SR} + \frac{\hat{\gamma}_4 - 1}{4}\widehat{SR}^2}} \xrightarrow{a} Z$$

where $\hat{\gamma}_3$ and $\hat{\gamma}_4$ respectively are the skewness and kurtosis of $\{r_t\}$.

Šidàk Correction

- Familywise Type I errors occur with probability

$$\mathrm{P}\left[\max_k\{\hat{z}[0]_k\}_{k=1,\ldots,K} > z_\alpha \middle| H_0\right] = 1-(1-\alpha)^K = \alpha_K$$

- For a FWER α_K, Šidàk's correction gives us a single-trial significance level

$$\alpha = 1-(1-\alpha_K)^{\frac{1}{K}}$$

- Then, the null hypothesis is rejected with confidence $(1-\alpha_K)$ if

$$\max_k\{\hat{z}[0]_k\}_{k=1,\ldots,K} > z_\alpha$$

where

$$z_\alpha = Z^{-1}[1-\alpha] = Z^{-1}\left[(1-\alpha_K)^{1/K}\right]$$

is the critical value of the Standard Normal distribution that leaves a probability α to the right, and $Z[.]$ is the CDF of the standard Normal distribution.

FWER Estimation

- We can derive the Type I error under multiple testing (α_K) as follows:

1. Apply the clustering procedure described in López de Prado and Lewis [2018a] on the trials correlation matrix, to estimate clusters' returns series and $\mathrm{E}[K]$.
2. On the selected cluster's returns, estimate

$$\hat{z}[0] = \max_k \{\hat{z}[0]_k\}_{k=1,\ldots,K}$$

3. Compute the Type I error for a single test,

$$\alpha = 1 - Z\left[\hat{z}[0]\right]$$

4. Correct for multiple testing, $\alpha_K = 1 - (1-\alpha)^K$, resulting in

$$\alpha_K = 1 - Z\left[\hat{z}[0]\right]^{\mathrm{E}[K]}$$

Type II Errors under Multiple Testing

True Positive Probability

- Suppose that the alternative hypothesis ($H_1: SR > 0$) for the best strategy is true, and $SR = SR^*$. Then, the power of the test associated with a FWER α_K is:

$$\mathrm{P}\left[\max_k\{\hat{z}[0]_k\}_{k=1,\ldots,K} > z_\alpha \middle| SR = SR^*\right] =$$

$$1 - Z\left[z_\alpha - \frac{SR^*\sqrt{T-1}}{\sqrt{1 - \hat{\gamma}_3\widehat{SR} + \frac{\hat{\gamma}_4 - 1}{4}\widehat{SR}^2}}\right] = 1 - \beta$$

where $z_\alpha = Z^{-1}\left[(1-\alpha_K)^{1/K}\right]$. Accordingly, the power of the test increases with SR^*, the sample length, and the skewness, however it decreases with the kurtosis. The power of the test also increases with *K*, because multiple testing reduces the test's confidence.

$\boldsymbol{\beta}$ Estimation

- We can derive the Type II error under multiple testing (β_K) as follows:

1. Given a FWER α_K, which is either set exogenously or it is estimated as explained in the previous section, compute the single-test critical value, z_α.
2. The probability of missing a strategy with Sharpe ratio SR^* is

$$\beta = Z[z_\alpha - \theta]$$

where

$$\theta = \frac{SR^*\sqrt{T-1}}{\sqrt{1 - \hat{\gamma}_3\widehat{SR} + \frac{\hat{\gamma}_4 - 1}{4}\widehat{SR}^2}}$$

Multiple Testing Adjustment for $\boldsymbol{\beta}$

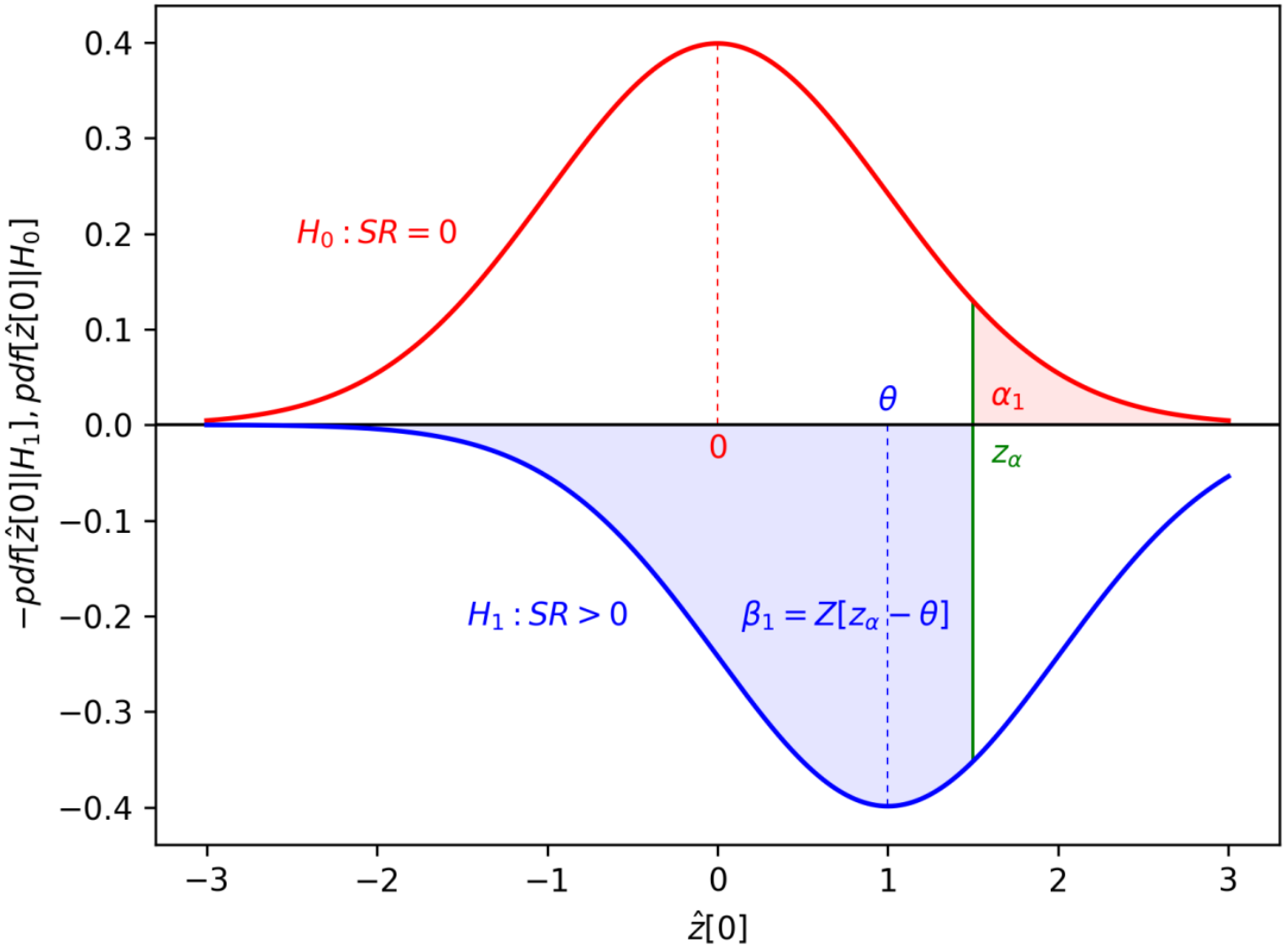
- After one trial, the probability of making a type II error is β.
- After K independent trials, the probability of making a type II error on all of them is

$$\beta_K = \beta^K$$

- Note the difference with FWER:
  - In the false positive case, we are interested in the probability of making *at least one* error. This is because a single false alarm is a failure.
  - However, in the false negative case, we are interested in the probability that *all* positives are missed.

The Trade-Off Between Type I and Type II Errors

The α_1 vs β_1 trade-off

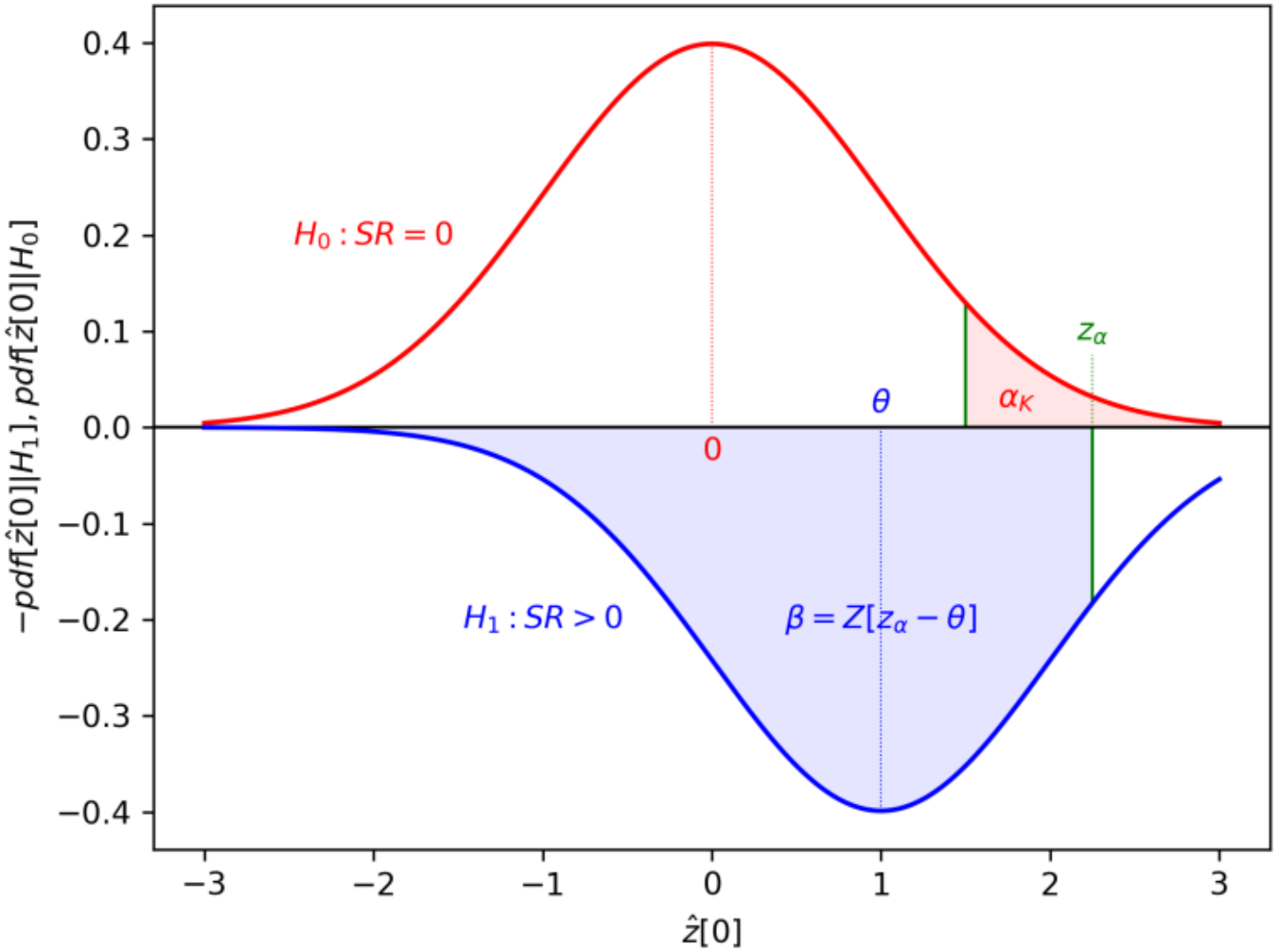


The red distribution models the probability of $\widehat{SR}$ estimates under the assumption that H_0 is true. The blue distribution (plotted upside-down, to facilitate display) models the probability of $\widehat{SR}$ estimates under the assumption that H_1 is true, and in particular under the scenario where $SR^* = 1$.

The sample length, skewness and kurtosis influence the variance of these two distributions.

Given an actual estimate $\widehat{SR}$, those variables determine the probabilities α_1 and β_1, where **decreasing one implies increasing the other**.

The α_K vs β trade-off



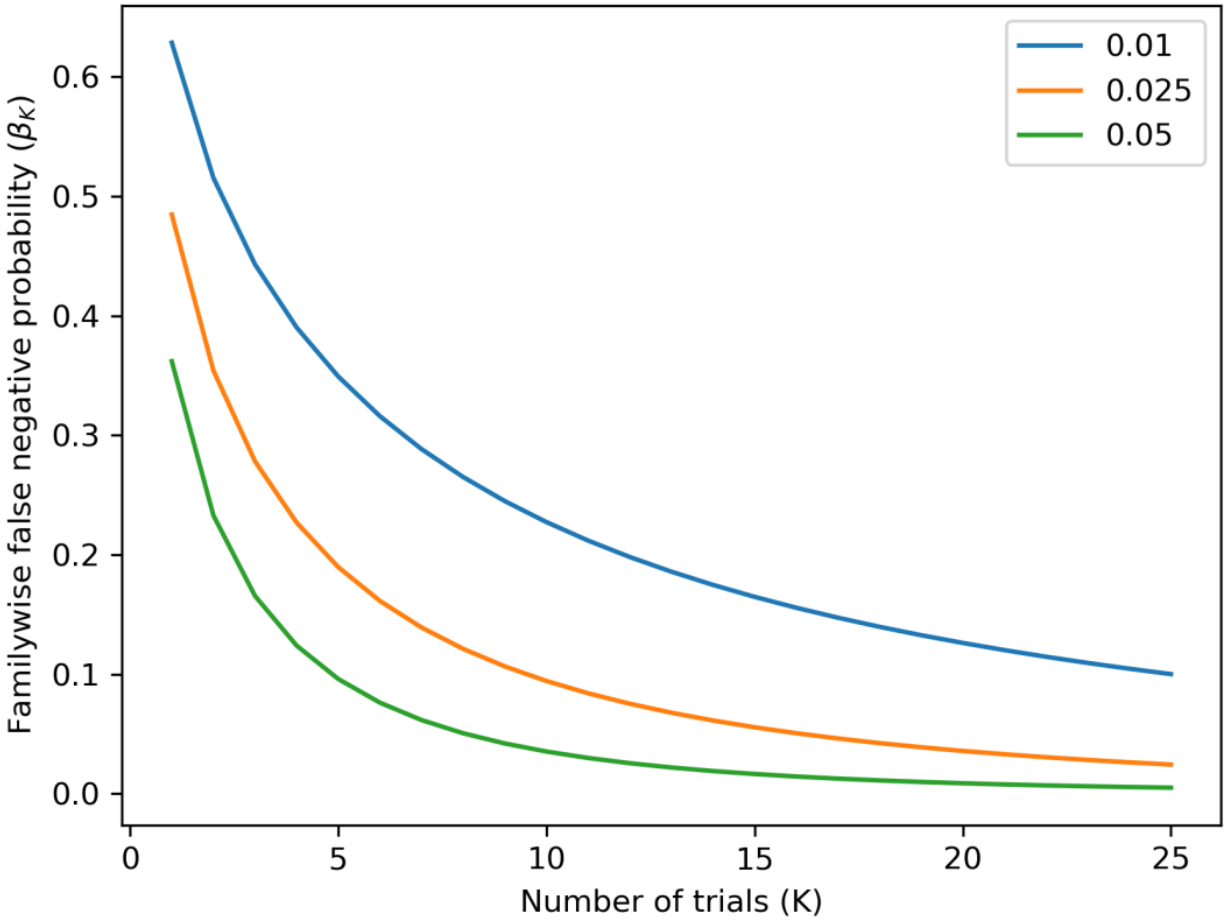
$$\beta = Z\left[Z^{-1}\left[(1-\alpha_K)^{1/K}\right] - \theta\right]$$

The analytic solution we derived for Type II errors makes it obvious that this tradeoff also exists between α_K and β, although in a not so straightforward manner as in the $K = 1$ case.

For a fixed α_K, as *K* increases, α decreases, z_α increases, hence β increases.

But what about β^K?

What is the Combined Effect on β_K?



As K increases, there are two competing forces at play:

- For a fixed α_K, α decreases, z_α increases, hence β increases.
- $\beta_K = \beta^K$ decreases.

Although β increases with K, the overall effect is to decrease β_K.

For a fixed α_K, the equation that determines β_K as a function of K and θ is

$$\beta_K = \left(Z\left[Z^{-1}\left[(1-\alpha_K)^{1/K}\right]-\theta\right]\right)^K$$

Precision and Recall under Multiple Testing

Precision and Recall after One Trial (1/3)

- Consider s investment strategies. Some of these strategies are false discoveries, in the sense that their expected return is not positive. We can decompose these strategies between true (s_T) and false (s_F), where $s = s_T + s_F$. Let θ be the odds ratio of true strategies against false strategies, $\theta = \frac{s_T}{s_F}$. In a field like financial economics, where the signal-to-noise ratio is low, false strategies abound, hence θ is expected to be low. The number of true investment strategies is

$$s_T = s\frac{s_T}{s_T + s_F} = s\frac{\frac{s_T}{s_F}}{\frac{s_T + s_F}{s_F}} = s\frac{\theta}{1+\theta}$$

- Likewise, the number of false investment strategies is

$$s_F = s - s_T = s\left(1 - \frac{\theta}{1+\theta}\right) = s\frac{1}{1+\theta}$$

Precision and Recall after One Trial (2/3)

- Given a false positive rate α (type I error), we will obtain a number of false positives, $FP = \alpha s_F$, and a number of true negatives, $TN = (1-\alpha)s_F$. Let us denote as β the false negative rate (type II error) associated with that α. We will obtain a number of false negatives, $FN = \beta s_T$, and a number of true positives, $TP = (1-\beta)s_T$. Therefore, the precision and recall of our test is

$$precision = \frac{TP}{TP+FP} = \frac{(1-\beta)s_T}{(1-\beta)s_T + \alpha s_F} = \frac{(1-\beta)s\frac{\theta}{1+\theta}}{(1-\beta)s\frac{\theta}{1+\theta} + \alpha s\frac{1}{1+\theta}} = \frac{(1-\beta)\theta}{(1-\beta)\theta + \alpha}$$

$$recall = \frac{TP}{TP+FN} = \frac{(1-\beta)s_T}{(1-\beta)s_T + \beta s_T} = 1-\beta$$

Precision and Recall after One Trial (3/3)

- Before running backtests on a strategy, researchers should gather evidence that a strategy may indeed exist. The reason is, the precision of the test is a function of the odds ratio, θ. If the odds ratio is low, the precision will be low, even if we get a positive with high confidence (low *p*-value). In particular, a strategy is more likely false than true if $(1-\beta)\theta < \alpha$.
- For example, suppose that the probability of a backtested strategy being profitable is 0.01, that is, that 1 out of 100 strategies is true, hence $\theta = \frac{1}{99}$. Then, at the standard thresholds of $\alpha = 0.05$ and $\beta = 0.2$, researchers are expected to get approx. 58 positives out 1000 trials, where approx. 8 are true positives and approx. 50 are false positives. Under these circumstances, a *p*-value of 0.05 implies a false discovery rate of 86.09% (roughly $\frac{50}{58}$). For this reason alone, we should expect that most discoveries in financial econometrics are likely false.
- This argument leads to the same conclusion we reached earlier: *p*-values report a rather uninformative probability. **It is possible for a statistical test to have high confidence (low *p*-value) and low precision**.

Precision and Recall after Multiple Trials

- We saw earlier that the probability of making at least one type I error is

$$\alpha_K = 1 - (1 - \alpha)^K$$

- After one trial, the probability of making a type II error is β. After K independent trials, the probability of making a type II error on all of them is $\beta_K = \beta^K$.
- The precision and recall adjusted for multiple testing is

$$precision = \frac{(1 - \beta_K)\theta}{(1 - \beta_K)\theta + \alpha_K} = \frac{(1 - \beta^K)\theta}{(1 - \beta^K)\theta + 1 - (1 - \alpha)^K}$$

$$recall = 1 - \beta_K = 1 - \beta^K$$

Understanding Strategy Risk

Symmetric Payouts (1/2)

- Consider a strategy that produces *n* IID bets per year, where the outcome X_i of a bet $i \in [1, n]$ is a profit $\pi > 0$ with probability $\mathrm{P}[X_i = \pi] = p$, and a loss $-\pi$ with probability $\mathrm{P}[X_i = -\pi] = 1 - p$.
- Think of p as the precision of a binary classifier where a positive means betting on an opportunity, and a negative means passing on an opportunity: True positives are rewarded, false positives are punished, and negatives (whether true or false) have no payout.
- Since the betting outcomes $\{X_i\}_{i=1,\ldots,n}$ are independent, we compute the expected moments per bet:
  - The expected profit from one bet is $\mathrm{E}[X_i] = \pi p + (-\pi)(1 - p) = \pi(2p - 1)$.
  - The variance is $\mathrm{V}[X_i] = \mathrm{E}\left[X_i^2\right] - \mathrm{E}[X_i]^2$, where $\mathrm{E}\left[X_i^2\right] = \pi^2 p + (-\pi)^2(1 - p) = \pi^2$, thus $\mathrm{V}[X_i] = \pi^2 - \pi^2(2p - 1)^2 = \pi^2[1 - (2p - 1)^2] = 4\pi^2 p(1 - p)$.
- For *n* IID bets per year, the annualized Sharpe ratio (θ) is

$$\theta[p, n] = \frac{n\mathrm{E}[X_i]}{\sqrt{n\mathrm{V}[X_i]}} = \frac{2p - 1}{\underbrace{2\sqrt{p(1 - p)}}_{\substack{\mathrm{t-value\ of}\ p \\ \mathrm{under}\ \mathrm{H}_0 : p = \frac{1}{2}}}} \sqrt{n}$$

Symmetric Payouts (2/2)

- Note how π cancels out of the above equation, because the payouts are symmetric.
- Just as in the Gaussian case, $\theta[p, n]$ can be understood as a re-scaled t-value.
- This illustrates the point that, even for a small $p > \frac{1}{2}$, the Sharpe ratio can be made high for a sufficiently large *n.*
- **This is the economic basis for high-frequency trading, where p can be barely above .5, and the key to a successful business is to increase n.**
- **The Sharpe ratio is a function of precision rather than accuracy**, because passing on an opportunity (a negative) is not rewarded or punished directly (although too many negatives may lead to a small *n*, which will depress the Sharpe ratio toward zero).
  - For example, $p = .55 \Longrightarrow \frac{2p-1}{2\sqrt{p(1-p)}} = 0.1005$, and achieving an annualized Sharpe ratio of 2 requires 396 bets per year.

Asymmetric Payouts

- Consider a strategy that produces *n* IID bets per year, where the outcome X_i of a bet $i \in [1, n]$ is π_+ with probability $\mathrm{P}[X_i = \pi_+] = p$, and an outcome π_-, $\pi_- < \pi_+$ occurs with probability $\mathrm{P}[X_i = \pi_-] = 1 - p$.
  - The expected profit from one bet is $\mathrm{E}[X_i] = p\pi_+ + (1-p)\pi_- = (\pi_+ - \pi_-)p + \pi_-$.
  - The variance is $\mathrm{V}[X_i] = \mathrm{E}\left[X_i^2\right] - \mathrm{E}[X_i]^2$, where $\mathrm{E}\left[X_i^2\right] = p\pi_+^2 + (1-p)\pi_-^2 = (\pi_+^2 - \pi_-^2)p + \pi_-^2$, thus $\mathrm{V}[X_i] = (\pi_+ - \pi_-)^2 p(1-p)$.
- For *n* IID bets per year, the annualized Sharpe ratio (θ) is

$$\theta[p, n, \pi_-, \pi_+] = \frac{n\mathrm{E}[X_i]}{\sqrt{n\mathrm{V}[X_i]}} = \frac{(\pi_+ - \pi_-)p + \pi_-}{(\pi_+ - \pi_-)\sqrt{p(1-p)}}\sqrt{n}$$

and for $\pi_- = -\pi_+$ we can see that this equation reduces to the symmetric case:
$\theta[p, n, -\pi_+, \pi_+] = \frac{2\pi_+ p + \pi_+}{2\pi_+\sqrt{p(1-p)}}\sqrt{n} = \frac{2p-1}{2\sqrt{p(1-p)}}\sqrt{n} = \theta[p, n]$.

- For example, for $n = 260, \pi_- = -.01, \pi_+ = .005, p = .7$, we get $\theta = 1.173$.

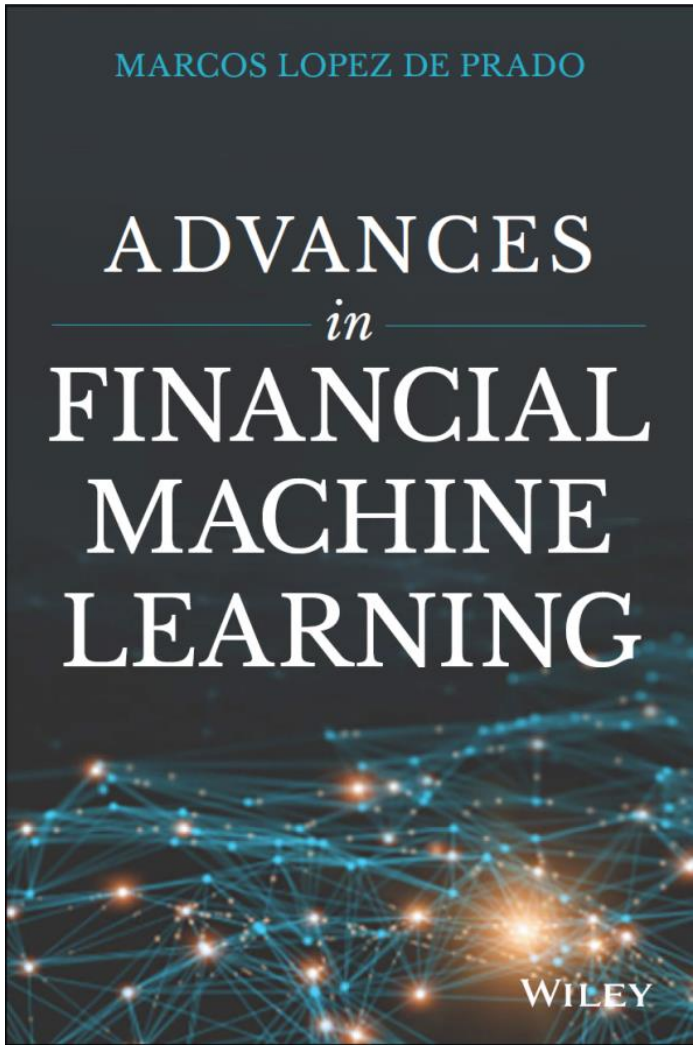
The Probability of Strategy Failure (1/2)

- In the example above, parameters
  - $\pi_- = -.01, \pi_+ = .005$ are set by the portfolio manager, and passed to the traders with the execution orders.
  - Parameter $n = 260$ is also set by the portfolio manager, as she decides what constitutes an opportunity worth betting on.
- The two parameters that are not under the control of the portfolio manager are p (determined by the market) and θ^* (the objective set by the investor). Because p is unknown, we can model it as a random variable, with expected value $\mathrm{E}[p]$.
- Let us define p_{θ^*} as the value of *p* below which the strategy will underperform a target Sharpe ratio θ^*, that is, $p_{\theta^*} = \max\{p | \theta \leq \theta^*\}$.
- For $p_{\theta^*=0} = \frac{2}{3}$, $p < p_{\theta^*=0} \Longrightarrow \theta \leq 0$. This highlights the risks involved in this strategy, because a relatively small drop in *p* (from $p = .7$ to $p = .67$) will wipe out all the profits. **The strategy is intrinsically risky, even if the holdings are not.**
- **That is a critical difference missing in most asset management textbooks: *Strategy risk* should not be confused with *portfolio risk*.**

The Probability of Strategy Failure (2/2)

- Firms and investors compute, monitor, and report portfolio risk without realizing that this tells us nothing about the risk of the strategy itself.
- Strategy risk is not the risk of the underlying portfolio, as computed by the *chief risk officer*.
- Strategy risk is the risk that the strategy will fail to succeed over time, a question of far greater relevance to the *chief investment officer*.
- The answer to the question "What is the probability that this strategy will fail?" is equivalent to computing $\mathrm{P}[p < p_{\theta^*}]$.

For Additional Details



The first wave of quantitative innovation in finance was led by Markowitz optimization. Machine Learning is the second wave and it will touch every aspect of finance. López de Prado's Advances in Financial Machine Learning is essential for readers who want to be ahead of the technology rather than being replaced by it.
— Prof. **Campbell Harvey**, Duke University. Former President of the American Finance Association.

Financial problems require very distinct machine learning solutions. Dr. López de Prado's book is the first one to characterize what makes standard machine learning tools fail when applied to the field of finance, and the first one to provide practical solutions to unique challenges faced by asset managers. Everyone who wants to understand the future of finance should read this book.
— Prof. **Frank Fabozzi**, EDHEC Business School. Editor of The Journal of Portfolio Management.

Disclaimer

- The views expressed in this document are the authors' and do not necessarily reflect those of the organizations he is affiliated with.
- No investment decision or particular course of action is recommended by this presentation.
-

How the Sharpe Ratio Died And Came Back to Life

Marcos López de Prado

Lawrence Berkeley National Laboratory
Computational Research Division

Key Points

- *Selection bias* under *multiple backtesting* makes it impossible to assess the probability that a strategy is false (Bailey et al. [2014]).
- Two implications:
  - "Most claimed research findings in empirical Finance are likely false" (Harvey et al. [2016])
  - Most quantitative firms invest in false positives
- This explains the high rate of failure among quantitative hedge funds: They do not have the technology to distinguish between a true strategy and a false strategy.
- The goal of this presentation is to introduce such technology, so that academic journals, regulators and investors may discard false strategies with confidence.
- My recent book discusses this subject at length: ***Advances in Financial Machine Learning***, Wiley (2018)

The Golden Age of the Sharpe Ratio (1966 – 2014)

Sharpe [1966]

- Consider an investment strategy with excess returns (or risk premia) $\{r_t\}$, $t = 1, \ldots, T$, which follow an IID Normal distribution,

$$r_t \sim \mathcal{N}[\mu, \sigma^2]$$

  where $\mathcal{N}[\mu, \sigma^2]$ represents a Normal distribution with mean μ and variance σ^2.
- The SR (non-annualized) of such strategy is defined as

$$SR = \frac{\mu}{\sigma}$$

- Because parameters μ and σ are not known, SR is estimated as

$$\widehat{SR} = \frac{\mathrm{E}[\{r_t\}]}{\sqrt{\mathrm{V}[\{r_t\}]}}$$

Lo [2002]

- Under the assumption that returns follow an IID Normal distribution, Lo [2002] derived the asymptotic distribution of $\widehat{SR}$ as

$$\left(\widehat{SR} - SR\right) \overset{a}{\rightarrow} \mathcal{N}\left[0, \frac{1 + \frac{1}{2}SR^2}{T}\right]$$

- Under the assumption that returns follow an IID non-Normal distribution, Mertens [2002] derived the asymptotic distribution of $\widehat{SR}$ as

$$\left(\widehat{SR} - SR\right) \overset{a}{\rightarrow} \mathcal{N}\left[0, \frac{1 + \frac{1}{2}SR^2 - \gamma_3 SR + \frac{\gamma_4 - 3}{4}SR^2}{T}\right]$$

where γ_3 is the skewness of $\{r_t\}$, and γ_4 is the kurtosis of $\{r_t\}$ ($\gamma_3 = 0$ and $\gamma_4 = 3$ when returns follow a Normal distribution).

Bailey and López de Prado [2012] (1/2)

- Christie [2005] and Opdyke [2007] discovered that, in fact, the Mertens [2002] equation is also valid under the more general assumption that returns are stationary and ergodic (not necessarily IID).
- Bailey and López de Prado [2012] utilized those results to derive the Probabilistic Sharpe Ratio (PSR).
- PSR estimates the probability that an observed $\widehat{SR}$ exceeds SR^* as

$$\widehat{PSR}[SR^*] = Z\left[\frac{\left(\widehat{SR} - SR^*\right)\sqrt{T-1}}{\sqrt{1 - \hat{\gamma}_3\widehat{SR} + \frac{\hat{\gamma}_4 - 1}{4}\widehat{SR}^2}}\right]$$

where $Z[.]$ is the CDF of the standard Normal distribution, T is the number of observed returns, $\hat{\gamma}_3$ is the skewness of the returns, and $\hat{\gamma}_4$ is the kurtosis of the returns. Note that $\widehat{SR}$ is the non-annualized estimate of SR, computed on the same frequency as the T observations.

Bailey and López de Prado [2012] (2/2)

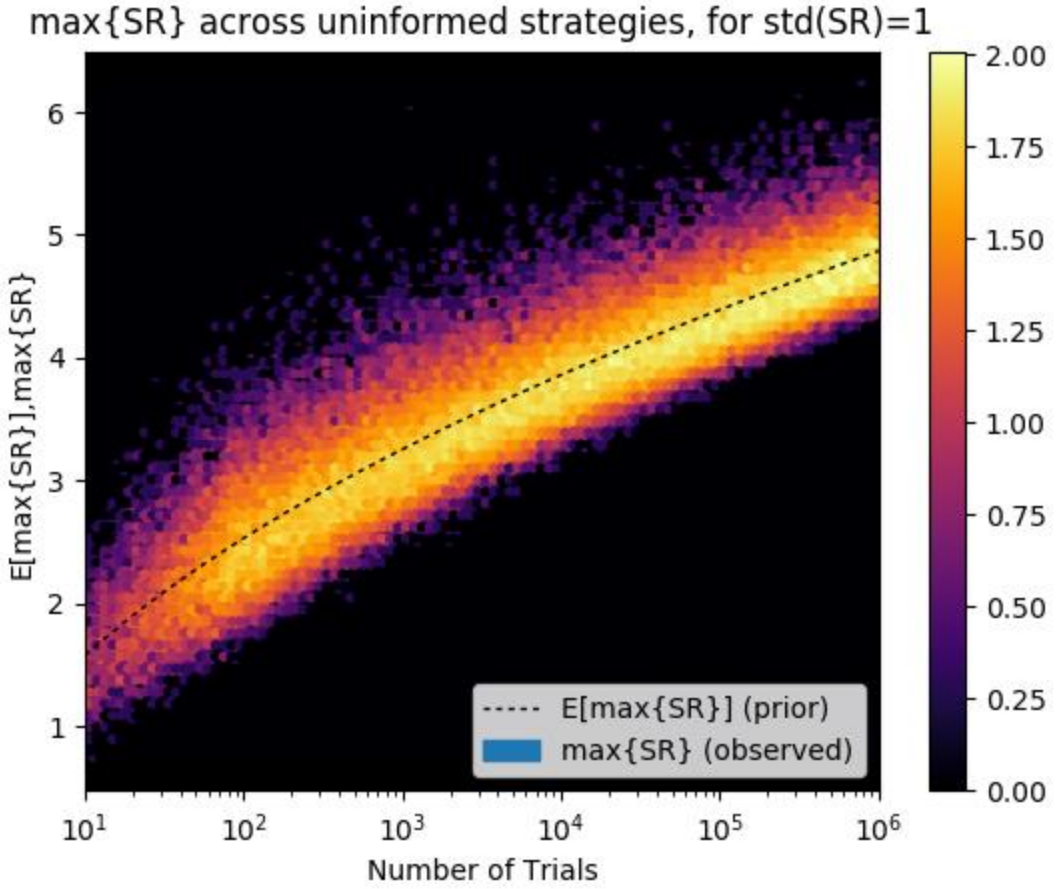
- For a given SR^*, $\widehat{PSR}$ increases with
  - greater mean returns ($\mathrm{E}[\{r_t\}]$)
  - lower variance of returns ($\mathrm{V}[\{r_t\}]$)
  - longer track records (T)
  - positively skewed returns ($\hat{\gamma}_3$)
  - thinner tails ($\hat{\gamma}_4$)
- This result also allows us to answer the question: *"How long should a track record be in order to have statistical confidence* $(1-\alpha)$ *that its estimated Sharpe ratio* $(\widehat{SR})$ *is above a given threshold* (SR^*)*"* (minimum track record length)

$$MinTRL = 1 + \left[1 - \hat{\gamma}_3 \widehat{SR} + \frac{\hat{\gamma}_4 - 1}{4}\widehat{SR}^2\right]\left(\frac{Z_\alpha}{\widehat{SR} - SR^*}\right)^2$$

where Z_α is the value of the Standard Normal CDF that leaves a probability α in the right tail.

The Death of the Sharpe Ratio (2014)

The Most Important Plot In Finance



The y-axis displays the distribution of the maximum Sharpe ratios (max{SR}) for a given number of trials (x-axis). A lighter color indicates a higher probability of obtaining that result, and the dash-line indicates the expected value.

For example, after only 1,000 independent backtests, the expected maximum Sharpe ratio (E[max{SR}]) is 3.26, even if the true Sharpe ratio of the strategy is 0!

The reason is Backtest Overfitting: When selection bias (picking the best result) takes place under multiple testing (running many alternative configurations), that backtest is likely to be a false discovery. **Most quantitative firms invest in false discoveries**.

The "False Strategy" Theorem [2014]

- Given a sample of IID-Gaussian Sharpe ratios, $\{\widehat{SR}_k\}$, $k = 1, \ldots, K$, with $\widehat{SR}_k \sim \mathcal{N}\left[0, \mathrm{V}\left[\{\widehat{SR}_k\}\right]\right]$, then

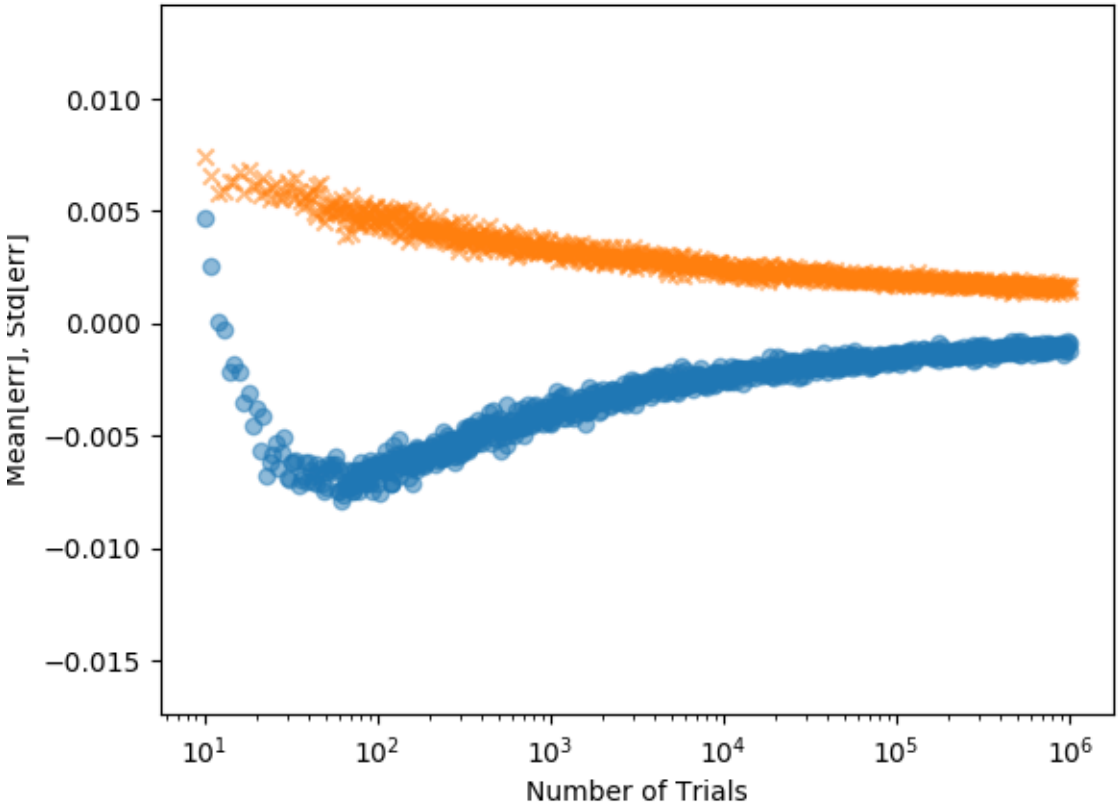
$$\mathrm{E}\left[\max_k\{\widehat{SR}_k\}\right]\left(\mathrm{V}\left[\{\widehat{SR}_k\}\right]\right)^{-1/2} \approx$$
$$(1-\gamma)Z^{-1}\left[1-\frac{1}{K}\right] + \gamma Z^{-1}\left[1-\frac{1}{Ke}\right]$$

  where $Z^{-1}[.]$ is the inverse of the standard Gaussian CDF, e is Euler's number, and γ is the Euler-Mascheroni constant.

- Corollary: Unless $\max_k\{\widehat{SR}_k\} \gg \mathrm{E}\left[\max_k\{\widehat{SR}_k\}\right]$, the discovered strategy is likely to be a *false positive*. But $\mathrm{E}\left[\max_k\{\widehat{SR}_k\}\right]$ is usually unknown, **ergo SR is dead**.

Source: López de Prado et al. (2014): "The effects of backtest overfitting on out-of-sample performance." *Notices of the American Mathematical Society*, 61(5), pp. 458-471.

Upper Boundary of Estimation Errors



Monte Carlo experiments confirm that the False Strategy theorem produces asymptotically unbiased estimates.

The blue circles report average errors relative to predicted values (y-axis), computed for alternative numbers of trials (x-axis). Only for K≈50, estimates exceed the correct value by approx. 0.7%.

The orange crosses report the standard deviation of the errors relative to predicted values (y-axis), computed for alternative numbers of trials (x-axis). The standard deviations are relatively small, below 0.5% of the forecasted values, and become smaller as the number of trials increases.

First Resurrection Attempt, and the Comatose SR (2014 – 2018)

Bailey and López de Prado [2014] (1/2)

- The Deflated Sharpe Ratio computes the probability that the Sharpe Ratio (SR) is statistically significant, after controlling for the inflationary effect of multiple trials, data dredging, non-normal returns and shorter sample lengths.

$$\widehat{DSR} \equiv \widehat{PSR}\left(\widehat{SR}_0\right) = Z\left[\frac{\left(\widehat{SR} - \widehat{SR}_0\right)\sqrt{T-1}}{\sqrt{1 - \hat{\gamma}_3\widehat{SR} + \frac{\hat{\gamma}_4 - 1}{4}\widehat{SR}^2}}\right]$$

where $\widehat{SR}_0$ is the estimate provided by the False Strategy theorem,

$$\widehat{SR}_0 = \sqrt{\mathrm{V}\left[\left\{\widehat{SR}_k\right\}\right]}\left((1-\gamma)Z^{-1}\left[1 - \frac{1}{K}\right] + \gamma Z^{-1}\left[1 - \frac{1}{Ke}\right]\right)$$

- DSR packs more information than SR, and it is expressed in probabilistic terms.

Bailey and López de Prado [2014] (2/2)

- The standard SR is computed as a function of two estimates:
  - Mean of returns
  - Standard deviation of returns
- DSR deflates SR by taking into consideration five additional variables (it packs more information):
  - The non-Normality of the returns $(\hat{\gamma}_3, \hat{\gamma}_4)$
  - The length of the returns series (T)
  - The amount of data dredging $\left(\mathrm{V}\left[\left\{\widehat{SR}_k\right\}\right]\right)$
  - The number of independent trials involved in the selection of the investment strategy (K)

The key to prevent selection bias is to record all trials, and determine correctly the number of effectively independent trials (*K*).

Unfortunately, $\mathrm{E}[K]$ and $\mathrm{V}\left[\left\{\widehat{SR}_k\right\}\right]$ are not directly observable … and so the first attempt to resurrect the Sharpe Ratio ended in a coma: Alive but inoperative.

The Second Coming of the Sharpe Ratio (2018)

Adding Meta-Research Variables

- Selection bias under multiple testing renders the Sharpe ratio useless:
  - The reason is, picking one strategy out of many discards information about the research process
  - These meta-research variables are crucial for evaluating the probability that the selected strategy is a false positive
- The "False Strategy" theorem tells us what meta-research variables are relevant, and how to use them.
- We are going to estimate these meta-research variables as follows:
  1. We cluster together strategies that are so highly correlated that we consider them duplicative
  2. Strategies in different clusters are so uncorrelated that we consider them effectively distinct (a proxy for independence)
  3. Once we have estimated what clusters are effectively independent, we can derive $\mathrm{E}[K]$ and $\mathrm{V}\left[\left\{\widehat{SR}_k\right\}\right]$

Extracting $\mathbf{E}[K]$

- What we have:
  - N strategies, with covariance matrix Σ, correlation matrix ρ
- What we want:
  - $K \ll N$ strategies that are "effectively independent"
- A solution:
  1. Define a local-distance between any two strategies

$$D_{i,j} = \sqrt{\frac{1}{2}(1 - \rho_{i,j})}$$

  2. Define a global-distance between any two strategies

$$\widetilde{D}_{i,j} = \sqrt{\sum_{k} \left(D_{i,k} - D_{j,k}\right)^2}$$

  3. Cluster together highly correlated strategies
  4. Form cluster-returns by applying the minimum variance allocation on intra-cluster strategies

Strategy Clustering (1/2)

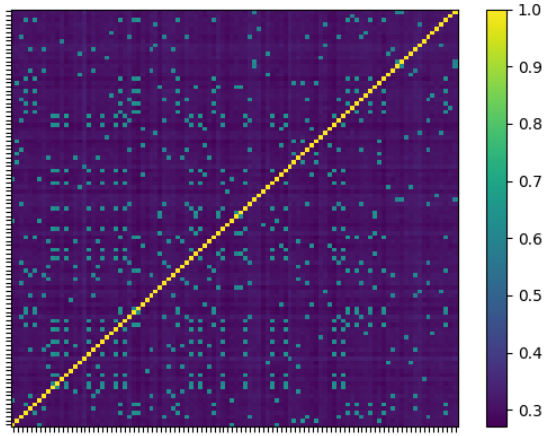
- Clustering steps:
  1. Base level clustering
  2. Recursive improvements to clustering

- Base Clustering:
  1. For each k in $2, \ldots, N-1$, and l in $1, \ldots, n_{init}$:
     - Apply Kmeans to extract *k* clusters using distance $\widetilde{D}$
     - Evaluate silhouette scores S_n, $n = 1, \ldots, N$, for the clustering
     - Evaluate quality score $q_{k,l} = \frac{\mathrm{E}[S_n]}{\sqrt{\mathrm{V}[S_n]}}$
  2. Choose optimal clustering, with $K = \mathrm{argmax}_k \left\{ \max_l \{q_{k,l}\} \right\}$

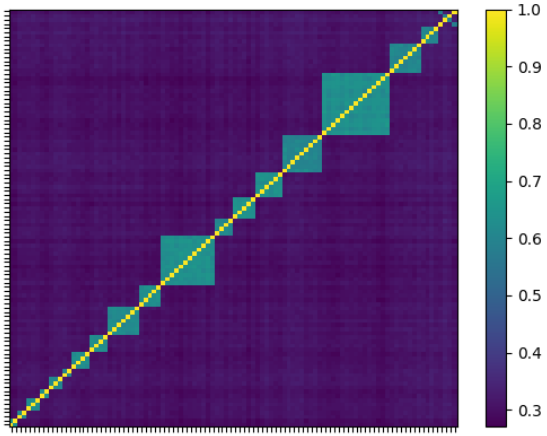
- Recursive improvement to clustering:
  1. Run Base Clustering to derive K
  2. Evaluate quality score q_k for each cluster $k = 1, \ldots, K$
  3. Fix clusters k where $q_k \geq \mathrm{E}[q_k]$
     - Recursively rerun Base Clustering for rest of clusters
     - Keep new clustering only if re-clustering improves average quality score

Strategy Clustering (2/2)

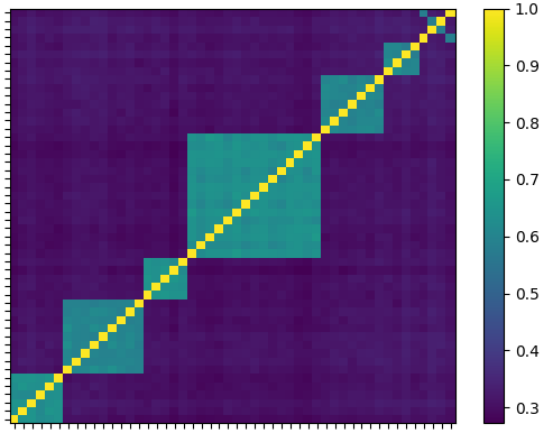
Initial Scrambled Random Block Covariance



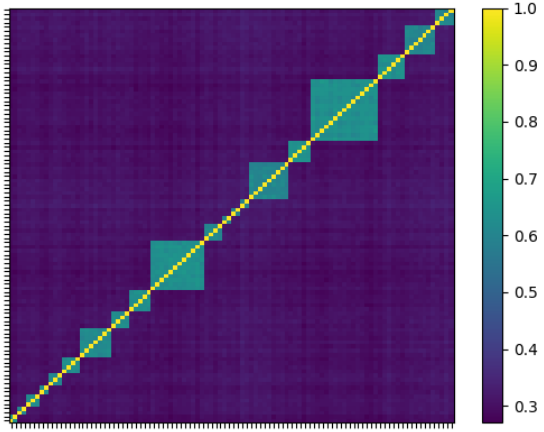
After Base Clustering



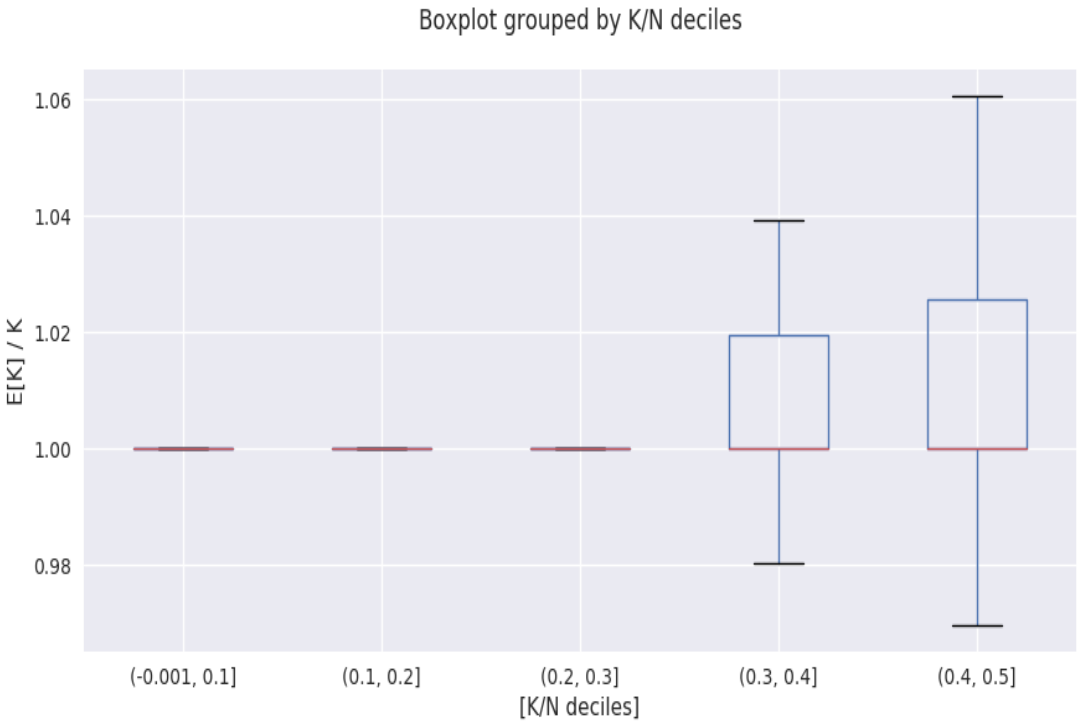
Recursively Re-evaluate Clustering



Final Clustering



Testing the Accuracy of the Clustering



- Create a random block covariance matrix
  - Size NxN
  - K block of random size
- Add global noise
- Run Clustering
- Compare expected cluster count K to number of estimated clusters

The boxplots show the results from these simulations. In particular, for $\frac{K}{N}$ in a given decile, we display the boxplot of the ratio of K predicted by the clustering to the actual $\mathrm{E}[K]$ predicted by clustering. Ideally, this ratio should be near 1. Simulations confirm that this clustering procedure is effective.

Cluster Returns

- Given Covariance matrix V_k for strategies $i \in C_k$ in a given cluster k, we compute the cluster returns using minimum variance weights

$$w = \frac{V_k^{-1}1}{1'V_k^{-1}1}; S_{k,t} = \sum_{i \in C_k} w_{k,i}\, r_{i,t}$$

- Note that, within a cluster, the strategies are highly correlated by design, so V_k may be numerically ill-conditioned.
- Two possibilities are:

1. Set $w_i \sim \frac{1}{\sigma_i^2}$. Stems from $V_k \approx V_{approx} = \begin{pmatrix} \sigma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_N^2 \end{pmatrix}$
2. Sherman-Morrison: Set $w_i \sim \frac{1}{\sigma_i^2} - \frac{\rho \sum_{j \in C_k} \frac{1}{\sigma_j}}{(1+(N-1)\rho)\sigma_i}$. Stems from $V_k \approx V_{approx} = \begin{pmatrix} \sigma_1^2 & \cdots & \rho\sigma_1\sigma_N \\ \vdots & \ddots & \vdots \\ \rho\sigma_1\sigma_N & \cdots & \sigma_N^2 \end{pmatrix} = (1-\rho)\begin{pmatrix} \sigma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_N^2 \end{pmatrix} + \rho\sigma\sigma', \sigma = \begin{pmatrix} \sigma_1 \\ \vdots \\ \sigma_N \end{pmatrix}$

Extracting $\mathbf{E}\left[\mathbf{V}\left[\left\{\widehat{SR}_k\right\}\right]\right]$ (1/2)

- So far, we have derived the strategy clusters, and the returns associated with those clusters. We can then compute the SR of each cluster, $\left\{\widehat{SR}_k\right\}_{k=1,\ldots,K}$.
- $\left\{\widehat{SR}_k\right\}$ are not comparable, as their frequency of trading may vary. To make them comparable, we must first annualize each. Accordingly, we calculate the frequency of trading as

$$Years_k = \frac{Last\ Date_k - First\ Date_k}{365.25\ days}$$

$$Frequency_k = \frac{T_k}{Years_k}$$

where T_k is the length of the $S_{k,t}$, and $First\ Date_k$ and $Last\ Date_k$ are the first and last dates of trading for $S_{k,t}$, respectively.

Extracting $\mathbf{E}\left[\mathbf{V}\left[\left\{\widehat{SR}_k\right\}\right]\right]$ (2/2)

- We estimate the annualized Sharpe Ratio (aSR) as

$$\widehat{aSR}_k = \frac{\mathrm{E}\left[\left\{S_{k,t}\right\}\right] Frequency_k}{\sqrt{\mathrm{V}\left[\left\{S_{k,t}\right\}\right] Frequency_k}} = \widehat{SR}_k \sqrt{Frequency_k}$$

- With these now comparable $\widehat{aSR}_k$, we can estimate the variance of clustered trials as

$$\mathrm{E}\left[\mathrm{V}\left[\left\{\widehat{SR}_k\right\}\right]\right] = \frac{\mathrm{V}\left[\left\{\widehat{aSR}_k\right\}\right]}{Frequency_{k^*}}$$

where $Frequency_{k^*}$ is the frequency of the selected strategy.

- That's it! We can now apply the "False Strategy" theorem and compute DSR.

Implications

Implications for Academics

- Most discoveries in empirical finance are false (Harvey et al. [2016]).
- Selection bias may invalidate the entire body of work performed for the past 100 years. Finance cannot survive as a discipline unless we solve this problem.
- Unless we learn to prevent them, investors and regulators have no reason to trust the value added by researchers and asset managers.
- We believe that providing practical solutions to this problem is in the best interest of the entire community of academics and practitioners.
- In this paper we apply the False Strategy theorem, first proved in Bailey et al. [2014], to the prevention of false positives in finance. This requires the estimation of two meta-research variables that allow us to discount for the likelihood of "lucky findings."
- Given that this method appears to be accurate and relatively easy to implement, academic journals should cease to accept papers that do not control for selection bias under multiple testing.
- In particular, papers must report the probability that the claimed financial discovery is a false positive.

Implications for Regulators

- Before the Food and Drug Administration (FDA) was created, adulteration and mislabeling of food and drugs caused frequent episodes of mass poisoning, birth defects and death. Such calamities only stopped through the enforcement of minimum research quality standards that prevented false positives.
- Every year, financial firms engaging in backtest overfitting defraud investors for tens of billions of dollars. It is, perhaps, the greatest fraud in financial history. It will only worsen as more powerful computers allow for an ever-larger number of trials. The financial firms of today are the pharmaceutical firms of 100 years ago.
- We hope that the machine learning tools presented in this paper will empower the Securities and Exchange Commission (SEC) and other regulatory agencies worldwide to take a more active role in stopping this rampant financial scam.
  - The SEC could demand that, going forward, quantitative firms that promote new investments must certify the probability that the proposed advice is simply bogus (false positive probability)
  - Quantitative firms should be required to store all trials involved in a discovery, so that a *post-mortem* analysis can be conducted after an investment fails to perform as advertised

Implications for Investors

- Many financial firms pray on the public's trust in science.
- They promote pseudo-scientific products arguments as scientific.
- Investors must understand that investment products based on award-winning journal articles are not necessarily scientific.
  - The authors never reported the number of trials involved in a discovery, and therefore we must assume the discovery is false
  - Firms have all the incentive to promote those journal articles, and make a fortune by charging fees (agency problem)
- One cynical argument is this: If the original author has not become rich with the discovery, what are my chances I will? The firm will make money regardless.
- For every financial product or investment advice, investors must demand that firms report the results of all trials, not only the best-looking ones.
- Investors should consult databases of investment forecasts, and assess the credibility of gurus and financial firms, based on all outcomes from past predictions investment funds (control for survivorship bias).

THANKS FOR YOUR ATTENTION!

References (1/3)

- American Statistical Association (2016): "Ethical guidelines for statistical practice." Committee on Professional Ethics of the American Statistical Association (April). Available at http://www.amstat.org/asa/files/pdfs/EthicalGuidelines.pdf
- Bailey, D., J. Borwein, M. López de Prado, and J. Zhu (2014a): "Pseudo-mathematics and financial charlatanism: The effects of backtest overfitting on out-of-sample performance." *Notices of the American Mathematical Society*, Vol. 61, No. 5, pp. 458–471. Available at http://ssrn.com/abstract=2308659
- Bailey, D., J. Borwein, M. López de Prado, and J. Zhu (2017): "The Probability of Backtest Overfitting." *Journal of Computational Finance*, Vol. 20, No. 4, pp. 39-70. Available at http://ssrn.com/abstract=2326253
- Bailey, D. and M. López de Prado (2012): "The Sharpe ratio efficient frontier." *Journal of Risk*, Vol. 15, No. 2, pp. 3–44.
- Bailey, D. and M. López de Prado (2014): "The deflated Sharpe ratio: Correcting for selection bias, backtest overfitting and non-normality." *Journal of Portfolio Management*, Vol. 40, No. 5, pp. 94-107.
- Christie, S. (2005): "Is the Sharpe Ratio Useful in Asset Allocation?" MAFC Research Paper No. 31, Applied Finance Centre, Macquarie University.

References (2/3)

- Harvey, C., Y. Liu and C. Zhu (2016): "…and the Cross-Section of Expected Returns." *Review of Financial Studies*, 29(1), pp. 5-68. Available at https://ssrn.com/abstract=2249314
- Lo, A. (2002): "The Statistics of Sharpe Ratios." *Financial Analysts Journal* (July), pp. 36-52.
- López de Prado, M. (2016a): "Building Diversified Portfolios that Outperform Out-of-Sample." *Journal of Portfolio Management*, Vol. 42, No. 4, pp. 59-69.
- López de Prado, M. (2016b): "Mathematics and Economics: A reality check." *Journal of Portfolio Management*, Vol. 43, No. 1, pp. 5-8.
- López de Prado, M. (2017): "Finance as an Industrial Science." *Journal of Portfolio Management*, Vol. 43, No. 4, pp. 5-9.
- López de Prado, M. (2018): Advances in Financial Machine Learning. 1st edition, Wiley. https://www.amazon.com/dp/1119482089
- Mertens, E. (2002): "Variance of the IID estimator in Lo (2002)." Working paper, University of Basel.
- Opdyke, J. (2007): "Comparing Sharpe ratios: So where are the p-values?" *Journal of Asset Management*, Vol. 8, No. 5, pp. 308–336.

References (3/3)

- Rousseeuw, P. (1987): “Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis.” *Computational and Applied Mathematics*, Vol. 20, pp. 53–65.
- Sharpe, W. (1966): “Mutual Fund Performance”, *Journal of Business*, Vol. 39, No. 1, pp. 119–138.
- Sharpe, W. (1975): “Adjusting for Risk in Portfolio Performance Measurement", *Journal of Portfolio Management*, Vol. 1, No. 2, Winter, pp. 29-34.
- Sharpe, W. (1994): “The Sharpe ratio”, *Journal of Portfolio Management*, Vol. 21, No. 1, Fall, pp. 49-58.

Notices

- This presentation is based on the paper:
  - López de Prado, M. and M. Lewis (2018): "Detection of False Investment Strategies Using Unsupervised Learning Methods." Working Paper. Available at https://ssrn.com/abstract=3167017
- The views expressed in this document are the author's and do not necessarily reflect those of the organizations he is affiliated with.
- No investment decision or particular course of action is recommended by this presentation.
-

A Practical Solution to the Multiple-Testing Crisis in Financial Research

Marcos López de Prado

Lawrence Berkeley National Laboratory
Computational Research Division

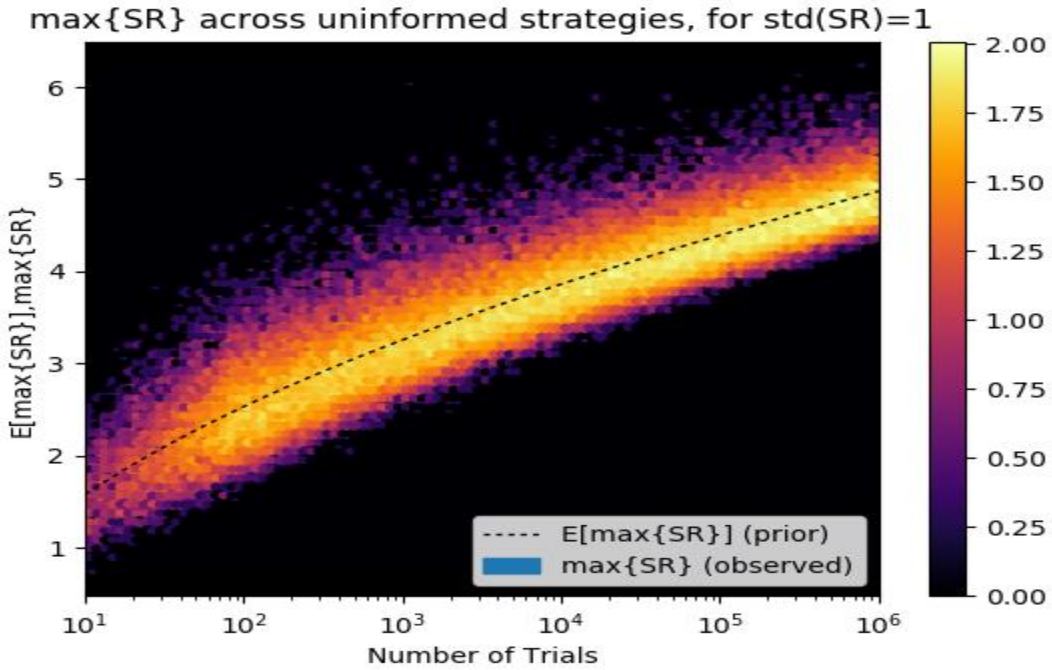
Key Points

- Most discoveries in empirical finance are false, as a consequence of selection bias under multiple testing.
- This may explain why so many hedge funds fail to perform as advertised or as expected, particularly in the quantitative space.
- These false discoveries may have been prevented if academic journals and investors demanded that any reported investment performance incorporates the false positive probability, adjusted for selection bias under multiple testing.
- In this presentation, we demonstrate how this adjusted false positive probability can be computed and reported for public consumption.

The full paper can be downloaded at https://ssrn.com/abstract=3177057

Section I
The Problem

The Most Important Plot In Finance



The y-axis displays the distribution of the maximum Sharpe ratios (max{SR}) for a given number of trials (x-axis). A lighter color indicates a higher probability of obtaining that result, and the dash-line indicates the expected value. For example, after only 1,000 independent backtests, the expected maximum Sharpe ratio (E[max{SR}]) is 3.26, even if the true Sharpe ratio of the strategy is zero!

The reason is *Backtest Overfitting*: When selection bias (picking the best result) takes place under multiple testing (running many alternative configurations), that backtest is likely to be a false discovery. **Most quantitative firms invest in false discoveries.**

The "False Strategy" Theorem [2014]

- Given a sample of IID-Gaussian Sharpe ratios, $\{\widehat{SR}_k\}$, $k = 1, \ldots, K$, with $\widehat{SR}_k \sim \mathcal{N}\left[0, \mathrm{V}\left[\{\widehat{SR}_k\}\right]\right]$, then

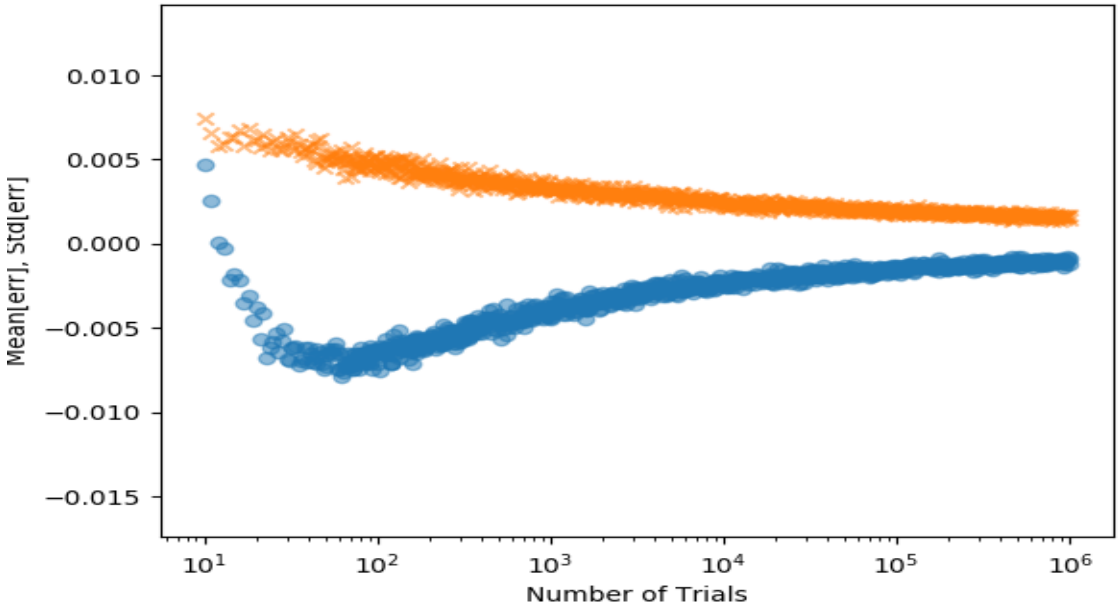
$$\mathrm{E}\left[\max_k\{\widehat{SR}_k\}\right]\left(\mathrm{V}\left[\{\widehat{SR}_k\}\right]\right)^{-1/2} \approx$$
$$(1-\gamma)Z^{-1}\left[1-\frac{1}{K}\right] + \gamma Z^{-1}\left[1-\frac{1}{Ke}\right]$$

where $Z^{-1}[.]$ is the inverse of the standard Gaussian CDF, e is Euler's number, and γ is the Euler-Mascheroni constant.

- Corollary: Unless $\max_k\{\widehat{SR}_k\} \gg \mathrm{E}\left[\max_k\{\widehat{SR}_k\}\right]$, the discovered strategy is likely to be a *false positive*.

Source: López de Prado et al. (2014): "The effects of backtest overfitting on out-of-sample performance." *Notices of the American Mathematical Society*, 61(5), pp. 458-471.

Upper Boundary of Estimation Errors



Monte Carlo experiments confirm that the False Strategy theorem produces asymptotically unbiased estimates.

The blue circles report average errors relative to predicted values (y-axis), computed for alternative numbers of trials (x-axis). Only for K≈50, estimates exceed the correct value by approx. 0.7%.

The orange crosses report the standard deviation of the errors relative to predicted values (y-axis), computed for alternative numbers of trials (x-axis). The standard deviations are relatively small, below 0.5% of the forecasted values, and become smaller as the number of trials increases.

Section II
The Solution

The Deflated Sharpe Ratio (1/2)

- The Deflated Sharpe Ratio computes the probability that the Sharpe Ratio (SR) is statistically significant, after controlling for the inflationary effect of multiple trials, data dredging, non-normal returns and shorter sample lengths.

$$\widehat{DSR} \equiv \widehat{PSR}\left(\widehat{SR}_0\right) = Z\left[\frac{\left(\widehat{SR} - \widehat{SR}_0\right)\sqrt{T-1}}{\sqrt{1 - \hat{\gamma}_3\widehat{SR} + \frac{\hat{\gamma}_4 - 1}{4}\widehat{SR}^2}}\right]$$

where

$$\widehat{SR}_0 = \sqrt{\mathrm{V}\left[\left\{\widehat{SR}_k\right\}\right]}\left((1-\gamma)Z^{-1}\left[1 - \frac{1}{K}\right] + \gamma Z^{-1}\left[1 - \frac{1}{Ke}\right]\right)$$

- DSR packs more information than SR, and it is expressed in probabilistic terms.

The Deflated Sharpe Ratio (2/2)

- The standard SR is computed as a function of two estimates:
  - Mean of returns
  - Standard deviation of returns.
- DSR deflates SR by taking into consideration five additional variables (it packs more information):
  - The non-Normality of the returns $(\hat{\gamma}_3, \hat{\gamma}_4)$
  - The length of the returns series (T)
  - The amount of data dredging $\left(\mathrm{V}\left[\{\widehat{SR}_k\}\right]\right)$
  - The number of independent trials involved in the discovered strategy (K)

DSR can be used to determine the probability that a discovered strategy is a **False Positive**. The key is to record all trials, and determine correctly the clusters of effectively independent trials.

Section III
A Numerical Example

Selected Strategy

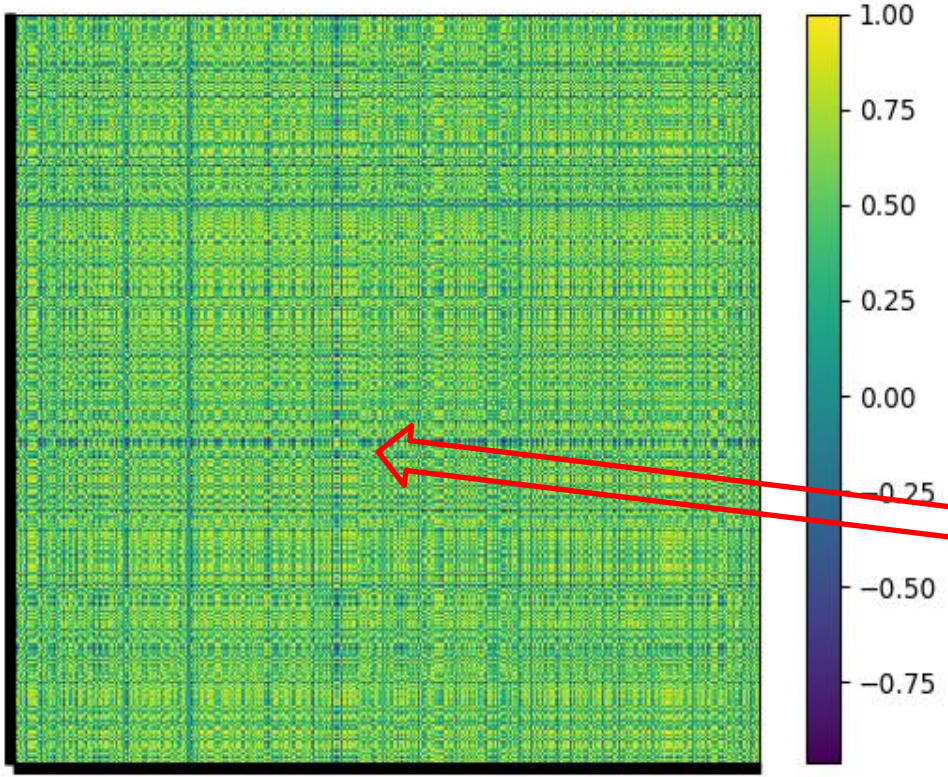
Statistic	Strategy
Start date	1/21/2010
End date	5/1/2018
aRoR (Total)	9.35%
Avg AUM (1E6)	1506.43
Avg Gini	0.88
Avg Duration	0.08
Avg Default Prob	1.58%
An. Sharpe ratio	2.00
Turnover	5.68
Efficient Number	186.26
Correl to Ix	0.48
Drawdown (95%)	2.89%
Time Underwater (95%)	0.20
Leverage	3.59

Consider a long-short investment strategy on high grade corporate bonds, with the performance statistics reported on this table.

That annualized Sharpe ratio is 2.00, with an annualized return of 9.35% and an average duration of only 0.08. The 95-percentile of the drawdowns is only 2.89%, and the 95-percentile of the time underwater is only 0.2 years.

Most investors would agree that this appears to be a good investment strategy…

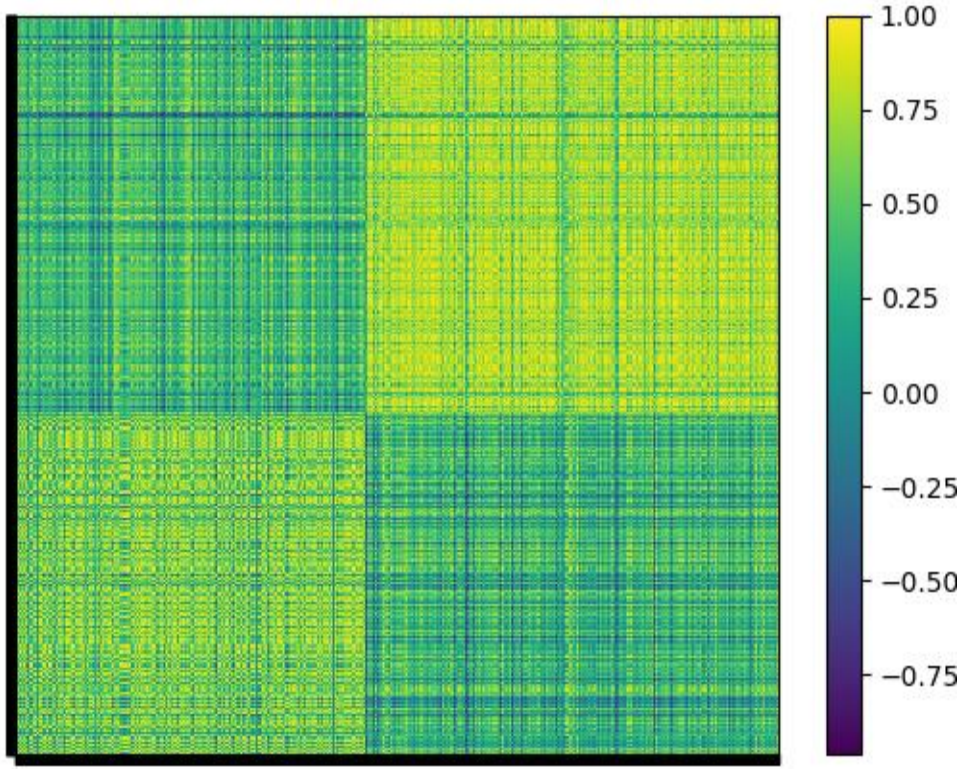
Disclosure of All Trials



The problem is, this is the best strategy out of 6,385 trials conducted by the investment firm on this investment universe. The left plot represents the heat map of the correlations between trials' return series.

Statistic	iBoxxIG	Strategy
Start date	1/21/2010	1/21/2010
End date	5/1/2018	5/1/2018
aRoR (Total)	4.90%	9.35%
Avg AUM (1E6)	1000.00	1506.43
Avg Gini	0.29	0.88
Avg Duration	7.88	0.08
Avg Default Prob	1.36%	1.58%
An. Sharpe ratio	0.99	2.00
Turnover	0.64	5.68
Efficient Number	1034.87	186.26
Correl to Ix	1.00	0.48
Drawdown (95%)	3.17%	2.89%
Time Underwater (95%)	0.23	0.20
Leverage	1.00	3.59

Clustering of Trials (1/10)



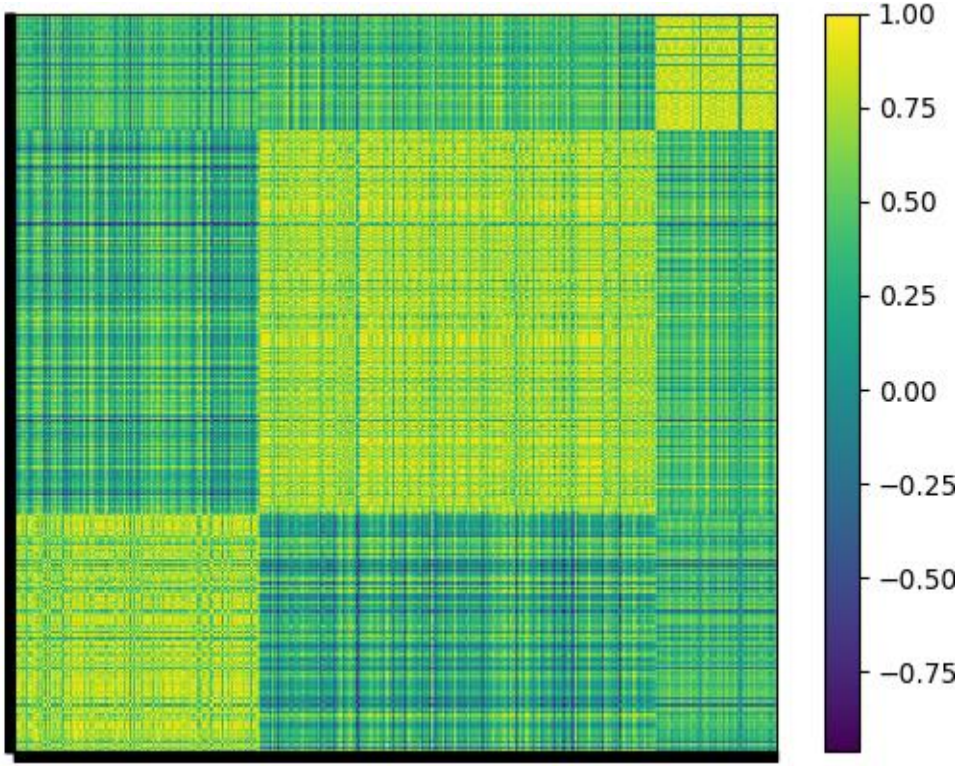
Fortunately, the 6,385 trials are not independent. The left plot represents the heat map of the correlations between trials' returns, grouped into 2 clusters.

The *quality* of this clustering, measured in terms of

$$q = \frac{\mathrm{E}[\{S_i\}]}{\sqrt{\mathrm{V}[\{S_i\}]}} = 2.3274$$

where S_i is the silhouette score of bond *i*.

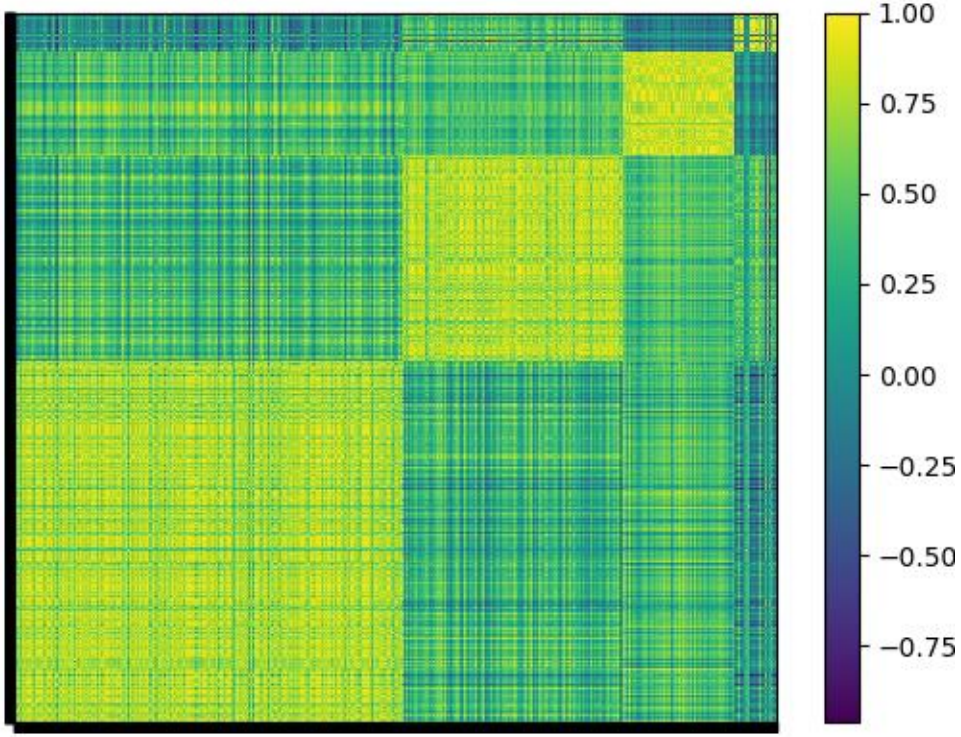
Clustering of Trials (2/10)



When we split the trials into 3 clusters, the quality of the clustering increases from $q = 2.3274$ to $q = 2.7068$.

Interestingly, what happens is that one of the two previous clusters is further split into two, leading to a greater contrast between the block diagonal color and the off-diagonal colors.

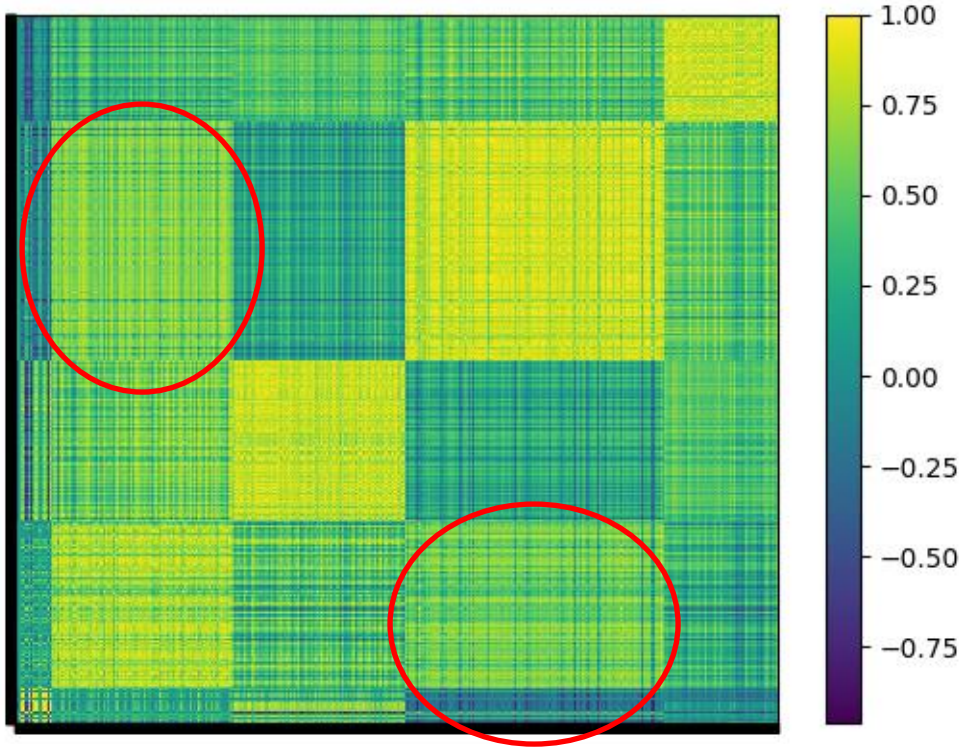
Clustering of Trials (3/10)



When we split the trials into 4 clusters, the quality of the clustering increases from $q = 2.7068$ to $q = 2.7281$, a relatively minor improvement.

Out of the original two clusters, one has now been split into three.

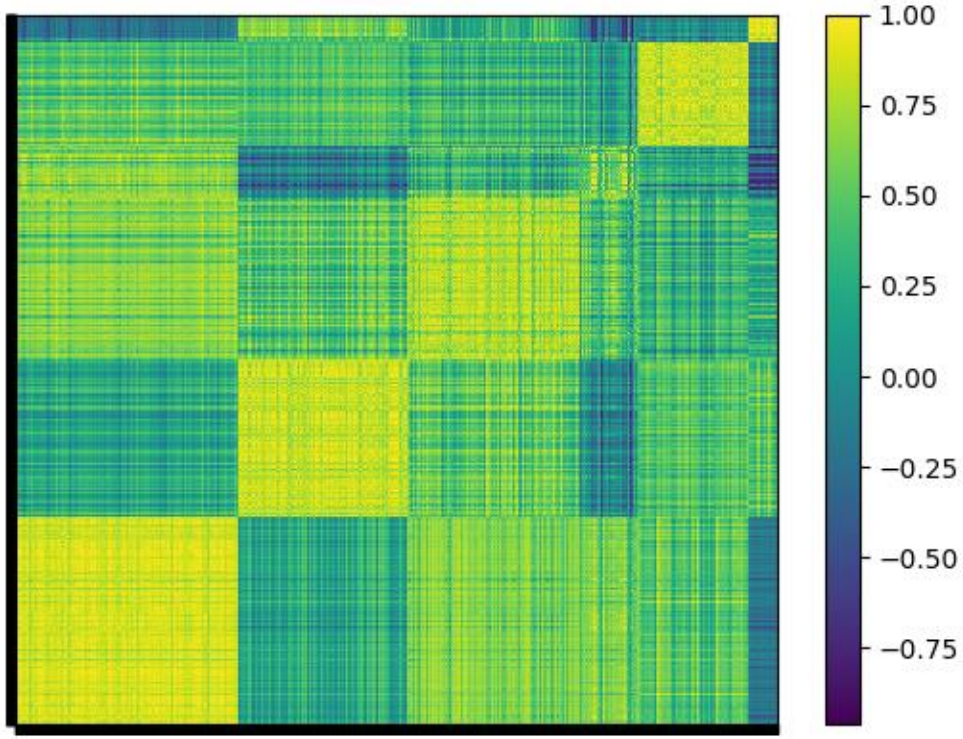
Clustering of Trials (4/10)



When we split the trials into 5 clusters, the quality of the clustering decreases from $q = 2.7281$ to $q = 2.6517$.

This clustering quality is worse than what we observed for 4 and 3 clusters. This worsening can be visualized in the lighter off-diagonal blocks, and it is the consequence of splitting one of the two initial clusters into 2.

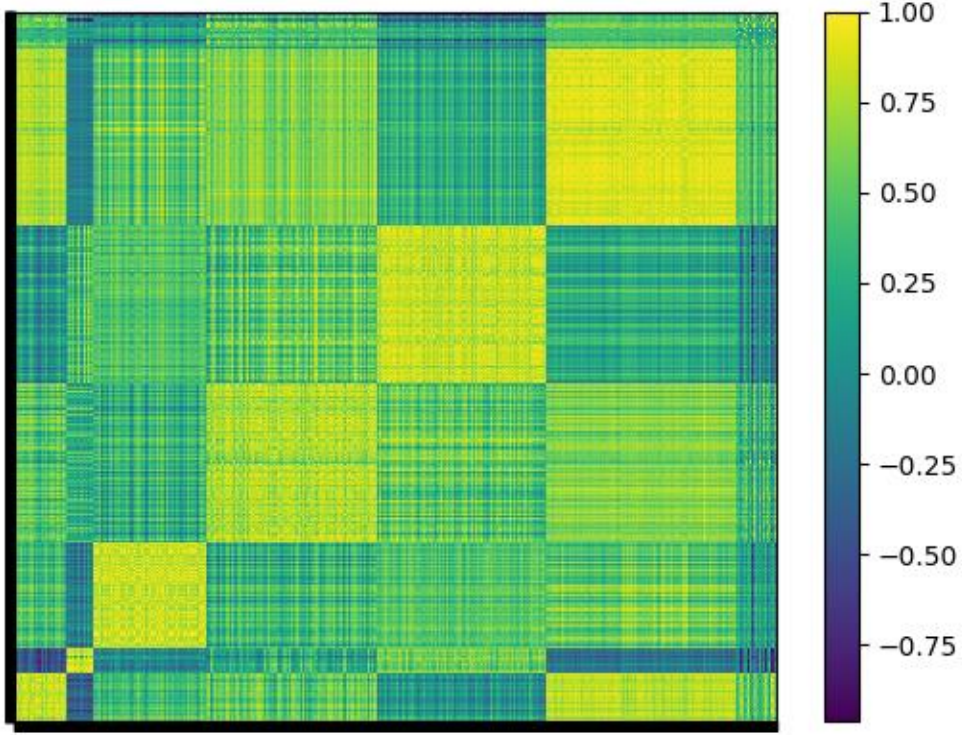
Clustering of Trials (5/10)



When we split the trials into 6 clusters, the quality of the clustering decreases from $q = 2.6517$ to $q = 2.4919$.

Now we can observed multiple light-colored off-diagonal clusters.

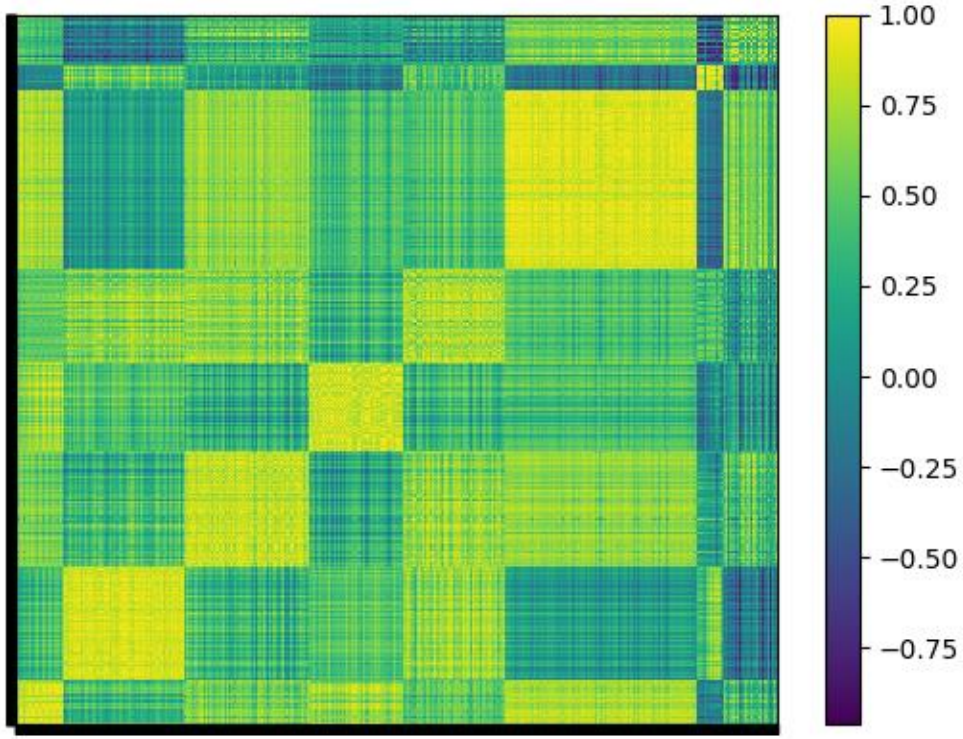
Clustering of Trials (6/10)



When we split the trials into 7 clusters, the quality of the clustering decreases from $q = 2.4919$ to $q = 2.3650$.

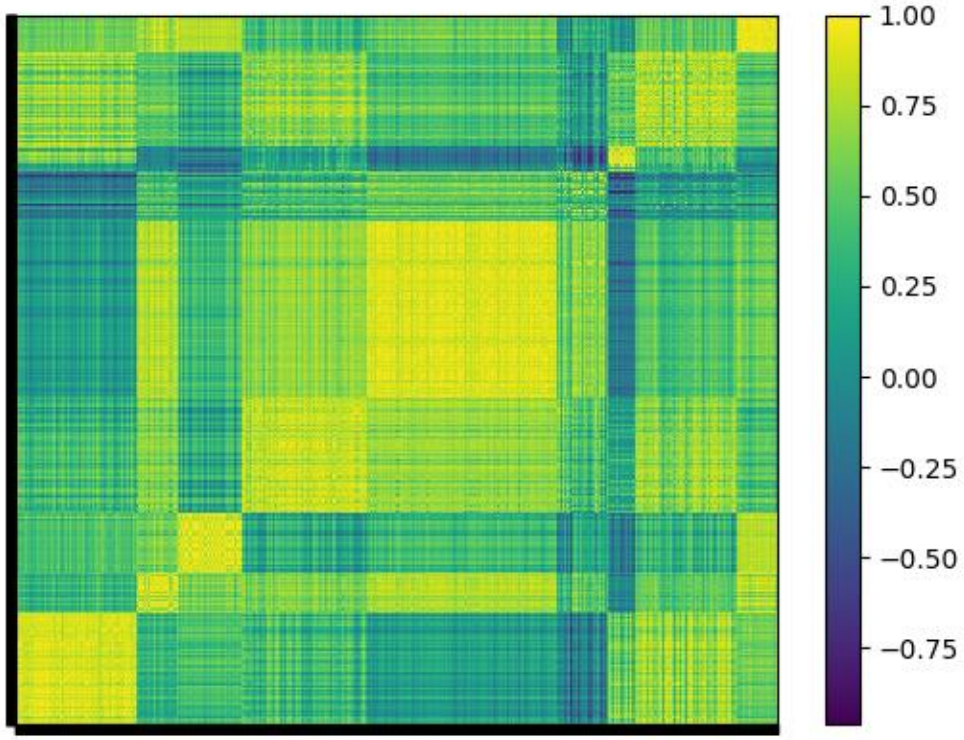
This clustering is only slightly better than the one with two clusters.

Clustering of Trials (7/10)



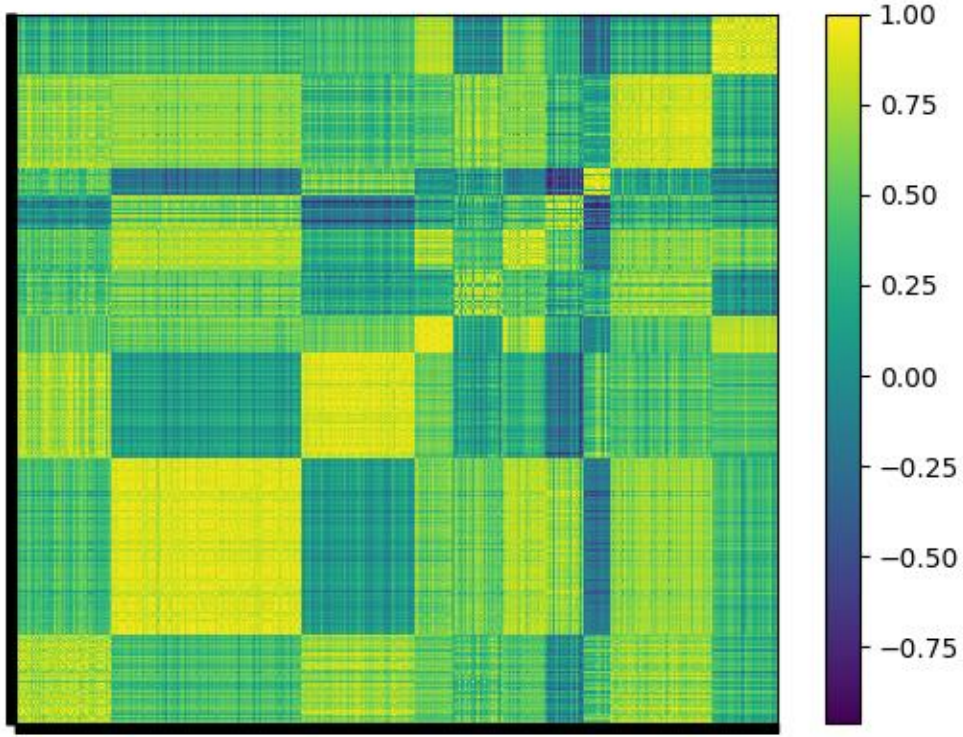
When we split the trials into 8 clusters, the quality of the clustering decreases from $q = 2.3650$ to $q = 2.2822$.

Clustering of Trials (8/10)



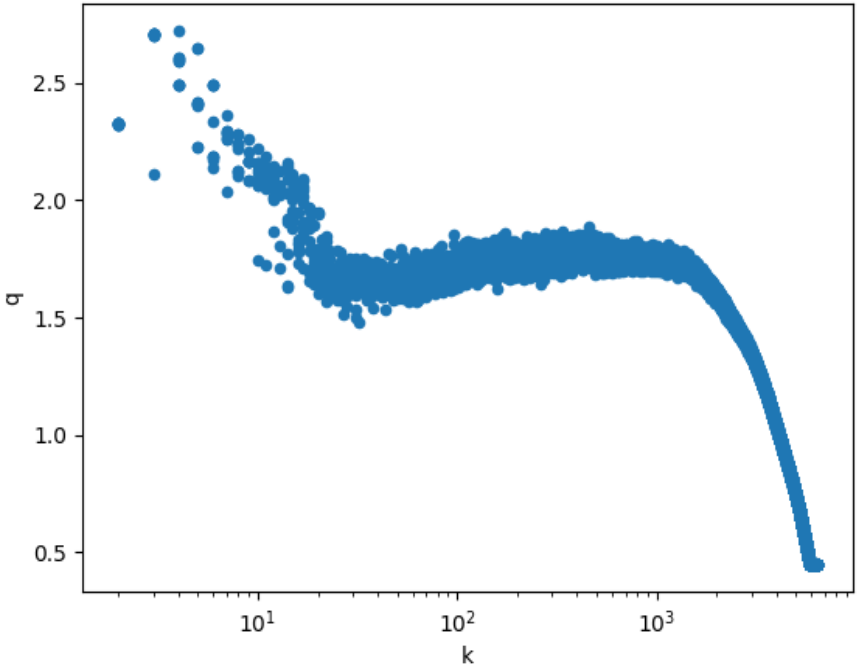
When we split the trials into 9 clusters, the quality of the clustering decreases from $q = 2.2822$ to $q = 2.2594$.

Clustering of Trials (9/10)



When we split the trials into 10 clusters, the quality of the clustering decreases from $q = 2.2594$ to $q = 2.2211$.

Clustering of Trials (10/10)



The scatter plot shows the clustering quality (y-axis) as we increase the number of clusters (x-axis, in polynomial scale) from 2 to 6,384.

As we observed earlier, the maximum quality is achieved for 4 clusters. Clustering quality decays abruptly after 10 clusters.

In conclusion, the firm has conducted 4 substantially uncorrelated trials on this investment universe.

False Positive Probability

Stats	Cluster 0	Cluster 1	Cluster 2	Cluster 3
Strat Count	3265	1843	930	347
aSR	1.5733	1.4907	2.0275	1.0158
SR	0.0974	0.0923	0.1255	0.0629
Skew	-0.3333	-0.4520	-0.4194	0.8058
Kurt	11.2773	6.0953	7.4035	14.2807
T	2172	2168	2174	2172
StartDt	2010-01-04	2010-01-04	2010-01-04	2010-01-04
EndDt	2018-05-01	2018-04-25	2018-05-03	2018-05-01
Freq	261.0474	261.0821	261.1159	261.0474
sqrt(V[SR_k])	0.0257	0.0256	0.0256	0.0257
E[max SR_k]	0.0270	0.0270	0.0270	0.0270
DSR	0.9993	0.9985	**1.0000**	0.9558

For each cluster, we can compute the returns associated with an inverse variance allocation to the trials involved. Then we can compute the performance statistics *per cluster*. This prevents that outliers may significantly bias results. The strategy selected in slide 10 belongs to Cluster 2, which contains 930 trials. The non-annualized Sharpe ratio is 0.1255.

For 4 clusters with a standard deviation across cluster SRs of 0.0256, the "False Strategy" theorem expects a maximum SR of 0.027. Taking into account the number of cluster returns, their skewness and kurtosis, the probability that the estimated SR is significantly greater than E[max{SR}] (a.k.a. DSR) is almost 1. **The probability of a false positive is virtually zero**.

Robustness of Results (1/4)

Stats	Cluster 0	Cluster 1
Strat Count	2937	3448
aSR	1.7707	1.6023
SR	0.1096	0.0992
Skew	-0.5780	-0.3351
Kurt	6.5878	11.3212
T	2174	2172
StartDt	2010-01-04	2010-01-04
EndDt	2018-05-03	2018-05-01
Freq	261.1159	261.0474
sqrt(V[SR_k])	0.0074	0.0074
E[max SR_k]	0.0038	0.0038
DSR	**1.0000**	1.0000

We can analyze the robustness of the previous result by repeating the calculations under alternative clusterings. For 2 clusters, the selected strategy belongs to Cluster 0, where the non-annualized SR is 0.1096. The expected maximum SR is 0.0038, and the DSR is again virtually 1.

Stats	Cluster 0	Cluster 1	Cluster 2
Strat Count	2063	3329	993
aSR	1.4411	1.5780	2.0638
SR	0.0892	0.0977	0.1277
Skew	-0.4310	-0.3357	-0.4137
Kurt	5.8606	11.2267	7.3681
T	2170	2172	2174
StartDt	2010-01-04	2010-01-04	2010-01-04
EndDt	2018-04-27	2018-05-01	2018-05-03
Freq	261.1507	261.0474	261.1159
sqrt(V[SR_k])	0.0202	0.0203	0.0202
E[max SR_k]	0.0173	0.0173	0.0173
DSR	0.9995	0.9999	**1.0000**

For 3 clusters, the selected strategy belongs to Cluster 2, where the non-annualized SR is 0.1277. The expected maximum SR is 0.0173, and the DSR is again virtually 1.

Robustness of Results (2/4)

Stats	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Strat Count	317	1524	1434	2169	941
aSR	0.9690	1.4664	1.4065	1.5272	2.0319
SR	0.0600	0.0907	0.0870	0.0945	0.1257
Skew	2.2161	-0.3286	-0.4864	-0.4086	-0.4172
Kurt	41.2726	9.7988	5.4162	12.1809	7.4552
T	2172	2170	2168	2172	2174
StartDt	2010-01-04	2010-01-04	2010-01-04	2010-01-04	2010-01-04
EndDt	2018-05-01	2018-04-27	2018-04-25	2018-05-01	2018-05-03
Freq	261.0474	261.1507	261.0821	261.0474	261.1159
sqrt(V[SR_k])	0.0234	0.0234	0.0234	0.0234	0.0234
E[max SR_k]	0.0279	0.0279	0.0279	0.0279	0.0279
DSR	0.9418	0.9979	0.9964	0.9987	**1.0000**

For 5 clusters, the selected strategy belongs to Cluster 4, where the non-annualized SR is 0.1257. The expected maximum SR is 0.0279, and the DSR is again virtually 1.

Stats	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Strat Count	1873	1418	1447	476	935	236
aSR	1.5205	1.4034	1.4580	1.3853	2.0296	0.4322
SR	0.0941	0.0869	0.0902	0.0857	0.1256	0.0267
Skew	-0.4254	-0.4872	-0.3458	0.5432	-0.4188	0.1344
Kurt	13.0185	5.4077	9.9281	16.1401	7.4308	5.6976
T	2170	2168	2170	2172	2174	2170
StartDt	2010-01-04	2010-01-04	2010-01-04	2010-01-04	2010-01-04	2010-01-04
EndDt	2018-04-27	2018-04-25	2018-04-27	2018-05-01	2018-05-03	2018-04-27
Freq	261.1507	261.0821	261.1507	261.0474	261.1159	261.1507
sqrt(V[SR_k])	0.0321	0.0321	0.0321	0.0321	0.0321	0.0321
E[max SR_k]	0.0417	0.0418	0.0417	0.0418	0.0417	0.0417
DSR	0.9909	0.9797	0.9862	0.9807	**0.9999**	0.2421

For 6 clusters, the selected strategy belongs to Cluster 4, where the non-annualized SR is 0.1256. The expected maximum SR is 0.0417, and the DSR is again virtually 1.

Robustness of Results (3/4)

Stats	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Strat Count	443	232	940	1436	1418	1591	325
aSR	1.4985	0.4229	2.0314	1.4566	1.4034	1.4816	1.2380
SR	0.0927	0.0262	0.1257	0.0901	0.0869	0.0917	0.0766
Skew	-0.4098	0.1355	-0.4174	-0.3447	-0.4872	-0.4488	10.2898
Kurt	10.4565	5.6820	7.4499	9.9064	5.4077	13.8743	295.3934
T	2170	2170	2174	2169	2168	2170	2172
StartDt	2010-01-04	2010-01-04	2010-01-04	2010-01-04	2010-01-04	2010-01-04	2010-01-04
EndDt	2018-04-27	2018-04-27	2018-05-03	2018-04-26	2018-04-25	2018-04-27	2018-05-01
Freq	261.1507	261.1507	261.1159	261.1164	261.0821	261.1507	261.0474
sqrt(V[SR_k])	0.0298	0.0298	0.0298	0.0298	0.0298	0.0298	0.0298
E[max SR_k]	0.0413	0.0413	0.0413	0.0413	0.0413	0.0413	0.0413
DSR	0.9901	0.2403	**0.9999**	0.9868	0.9807	0.9884	0.9799

Stats	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7
Strat Count	411	1021	1037	794	846	1606	228	442
aSR	1.8643	1.3267	1.4133	1.9881	1.5228	1.4607	0.3817	1.3586
SR	0.1154	0.0821	0.0875	0.1230	0.0942	0.0904	0.0236	0.0841
Skew	-0.2217	-0.4884	-0.3657	-0.4156	-0.3822	-0.4481	0.1270	1.6051
Kurt	13.2850	5.1541	10.3922	6.7874	7.4346	12.7538	5.3075	34.8674
T	2170	2167	2169	2174	2168	2170	2170	2172
StartDt	2010-01-04	2010-01-05	2010-01-04	2010-01-04	2010-01-04	2010-01-04	2010-01-04	2010-01-04
EndDt	2018-04-27	2018-04-25	2018-04-26	2018-05-03	2018-04-25	2018-04-27	2018-04-27	2018-05-01
Freq	261.1507	261.0477	261.1164	261.1159	261.0821	261.1507	261.1507	261.0474
sqrt(V[SR_k])	0.0298	0.0298	0.0298	0.0298	0.0298	0.0298	0.0298	0.0298
E[max SR_k]	0.0435	0.0435	0.0435	0.0435	0.0435	0.0435	0.0435	0.0435
DSR	0.9994	0.9606	0.9772	**0.9998**	0.9895	0.9829	0.1774	0.9754

For 7 clusters, the selected strategy belongs to Cluster 2, where the non-annualized SR is 0.1257. The expected maximum SR is 0.0413, and the DSR is again virtually 1.

For 8 clusters, the selected strategy belongs to Cluster 3, where the non-annualized SR is 0.1230. The expected maximum SR is 0.0435, and the DSR is again virtually 1.

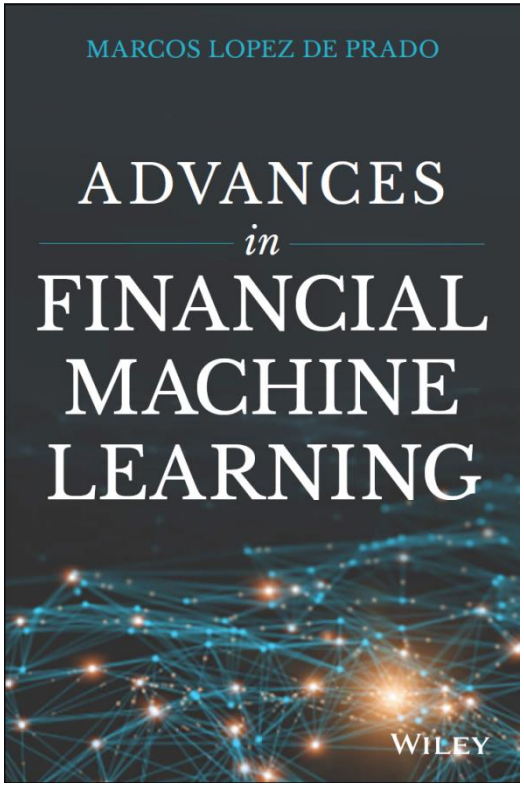
Robustness of Results (4/4)

Stats	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8
Strat Count	1021	352	536	1037	1593	440	228	846	332
aSR	1.3267	1.8185	1.8971	1.4133	1.4578	1.3482	0.3817	1.5228	1.9497
SR	0.0821	0.1125	0.1174	0.0875	0.0902	0.0834	0.0236	0.0942	0.1207
Skew	-0.4884	-0.2077	-0.3769	-0.3657	-0.4467	2.2752	0.1270	-0.3822	-0.4008
Kurt	5.1541	13.3085	6.1852	10.3922	12.7629	49.3210	5.3075	7.4346	10.0715
T	2167	2170	2160	2169	2170	2172	2170	2168	2171
StartDt	2010-01-05	2010-01-04	2010-01-22	2010-01-04	2010-01-04	2010-01-04	2010-01-04	2010-01-04	2010-01-04
EndDt	2018-04-25	2018-04-27	2018-05-03	2018-04-26	2018-04-27	2018-05-01	2018-04-27	2018-04-25	2018-04-30
Freq	261.0477	261.1507	260.9792	261.1164	261.1507	261.0474	261.1507	261.0821	261.0131
sqrt(V[SR_k])	0.0290	0.0290	0.0290	0.0290	0.0290	0.0290	0.0290	0.0290	0.0290
E[max SR_k]	0.0441	0.0441	0.0441	0.0441	0.0441	0.0441	0.0441	0.0441	0.0441
DSR	0.9580	0.9990	**0.9995**	0.9755	0.9813	0.9736	0.1696	0.9886	0.9997

Stats	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9
Strat Count	806	1596	948	332	409	353	327	227	851	536
aSR	1.5222	1.4586	1.3083	1.9497	1.3378	1.8174	1.2172	0.3787	1.4057	1.8971
SR	0.0942	0.0903	0.0810	0.1207	0.0828	0.1125	0.0753	0.0234	0.0870	0.1174
Skew	-0.3953	-0.4461	-0.4847	-0.4008	-0.1356	-0.2065	4.5167	0.1274	-0.4064	-0.3769
Kurt	6.9109	12.7512	5.1189	10.0715	7.4999	13.3321	108.1831	5.3035	10.9871	6.1852
T	2168	2170	2167	2171	2170	2170	2172	2170	2169	2160
StartDt	2010-01-04	2010-01-04	2010-01-05	2010-01-04	2010-01-04	2010-01-04	2010-01-04	2010-01-04	2010-01-04	2010-01-22
EndDt	2018-04-25	2018-04-27	2018-04-25	2018-04-30	2018-04-27	2018-04-27	2018-05-01	2018-04-27	2018-04-26	2018-05-03
Freq	261.0821	261.1507	261.0477	261.0131	261.1507	261.1507	261.0474	261.1507	261.1164	260.9792
sqrt(V[SR_k])	0.0278	0.0278	0.0278	0.0279	0.0278	0.0278	0.0278	0.0278	0.0278	0.0279
E[max SR_k]	0.0438	0.0438	0.0439	0.0439	0.0438	0.0438	0.0439	0.0438	0.0438	0.0439
DSR	0.9889	0.9819	0.9544	0.9997	0.9636	0.9990	0.9483	0.1706	0.9748	**0.9995**

Same conclusion for 9 and 10 clusters. Beyond 10 clusters, the quality of the clustering is so poor that we can consider them unrealistic scenarios

For Additional Details



The first wave of quantitative innovation in finance was led by Markowitz optimization. Machine Learning is the second wave and it will touch every aspect of finance. López de Prado's Advances in Financial Machine Learning is essential for readers who want to be ahead of the technology rather than being replaced by it.

— Prof. **Campbell Harvey**, Duke University. Former President of the American Finance Association.

Financial problems require very distinct machine learning solutions. Dr. López de Prado's book is the first one to characterize what makes standard machine learning tools fail when applied to the field of finance, and the first one to provide practical solutions to unique challenges faced by asset managers. Everyone who wants to understand the future of finance should read this book.

— Prof. **Frank Fabozzi**, EDHEC Business School. Editor of The Journal of Portfolio Management.

THANKS FOR YOUR ATTENTION!

Disclaimer

- The views expressed in this document are the authors' and do not necessarily reflect those of the organizations he is affiliated with.
- No investment decision or particular course of action is recommended by this presentation.
-