# 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) $\{r_t\}$, $t = 1, \dots, T$, 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 $\mu$ and variance $\sigma^2$.
- The SR (non-annualized) of such strategy is defined as

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

- Because parameters $\mu$ and $\sigma$ 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

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

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

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

where $\gamma_3$ is the skewness of $\{r_t\}$, and $\gamma_4$ 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)

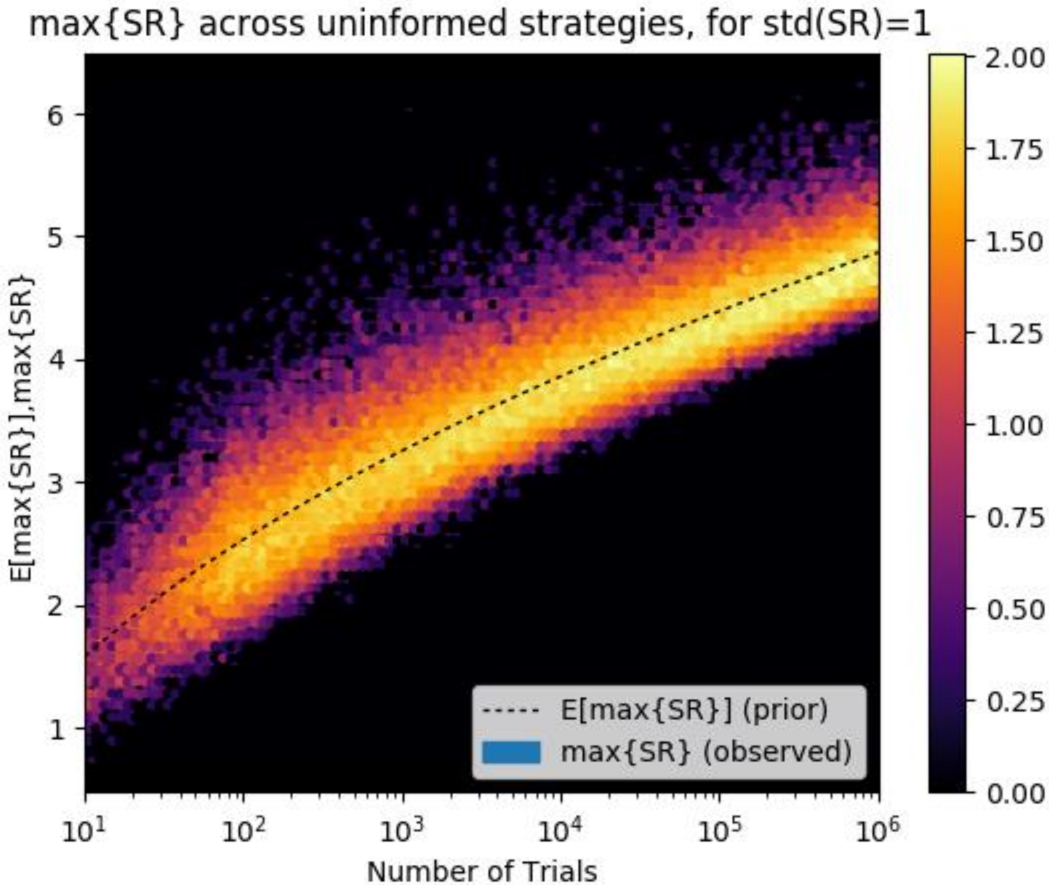
- For a given $SR^*$, $\widehat{PSR}$ increases with
  - greater mean returns ($\mathrm{E}[\{r_t\}]$)
  - lower variance of returns ($\mathrm{V}[\{r_t\}]$)
  - longer track records (*T*)
  - positively skewed returns ($\hat{\gamma}_3$)
  - thinner tails ($\hat{\gamma}_4$)
- This result also allows us 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 $Z_\alpha$ is the value of the Standard Normal CDF that leaves a probability $\alpha$ 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

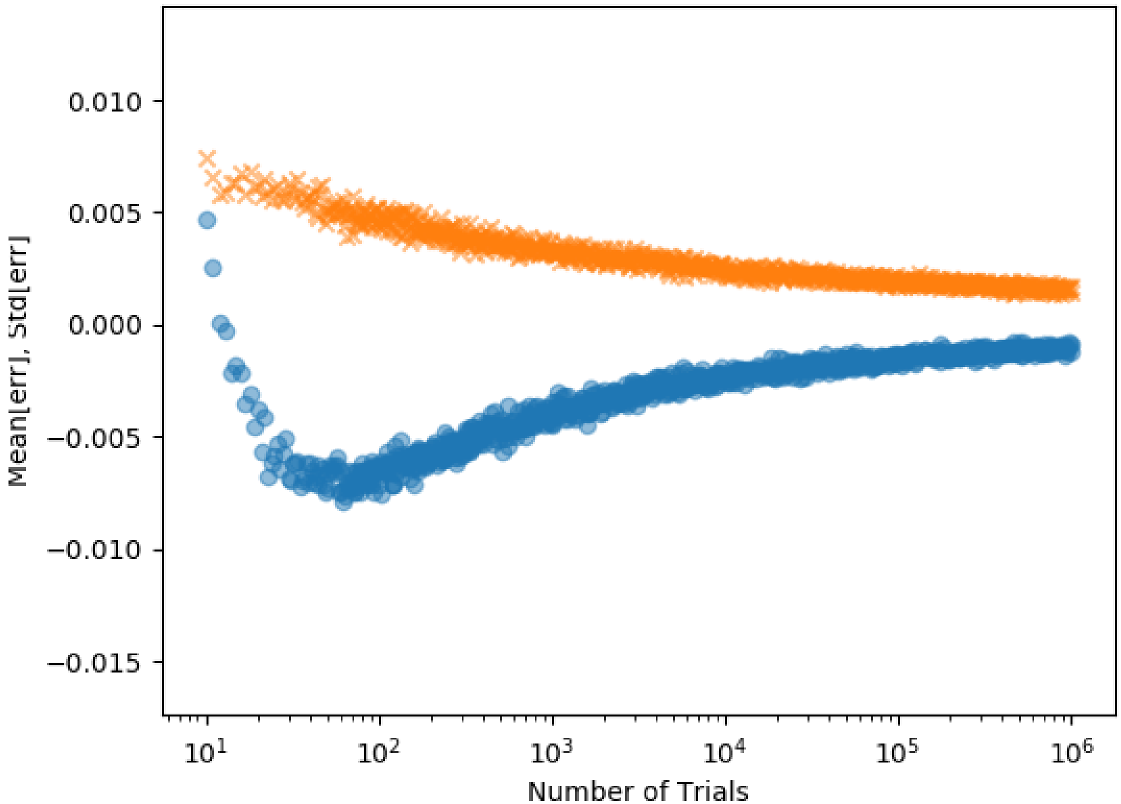
$$\mathrm{E}\left[\max_k\{\widehat{SR}_k\}\right]\left(\mathrm{V}\left[\{\widehat{SR}_k\}\right]\right)^{-1/2} \approx$$
$$(1-\gamma)Z^{-1}\left[1-\frac{1}{K}\right] + \gamma Z^{-1}\left[1-\frac{1}{Ke}\right]$$

  where $Z^{-1}[.]$ is the inverse of the standard Gaussian CDF, $e$ is Euler's number, and $\gamma$ 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 $\Sigma$, correlation matrix $\rho$
- 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)

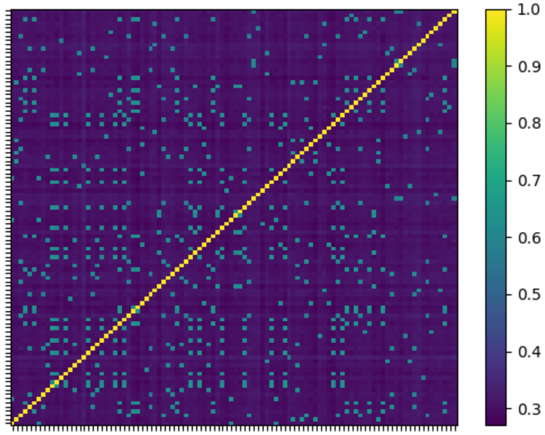
- Clustering steps:
  1. Base level clustering
  2. Recursive improvements to clustering

- 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 $K = \mathrm{argmax}_k \left\{ \max_l \{ q_{k,l} \} \right\}$

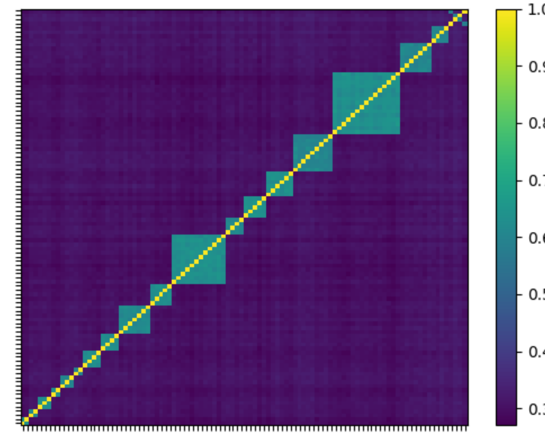
- Recursive improvement to clustering:
  1. Run Base Clustering to derive $K$
  2. Evaluate quality score $q_k$ for each cluster $k = 1, \ldots, K$
  3. Fix clusters $k$ where $q_k \geq \mathrm{E}[q_k]$
     - Recursively rerun Base Clustering for rest of clusters
     - Keep new clustering only if re-clustering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

# Strategy Clustering (2/2)

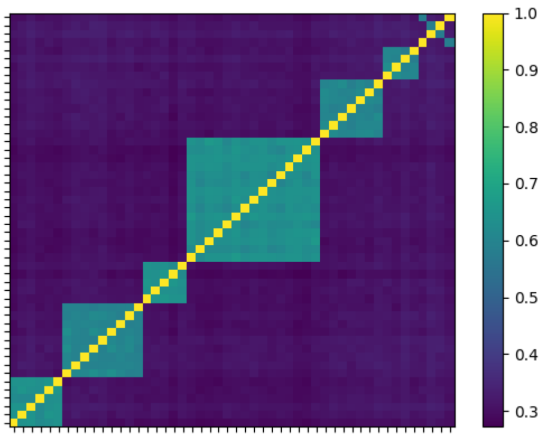
Initial
Scrambled
Random
Block
Covariance



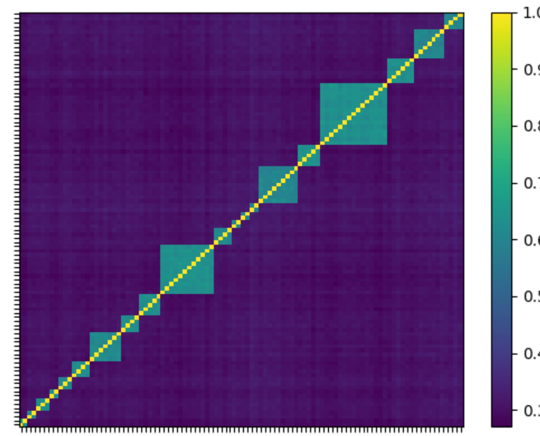

After Base
Clustering



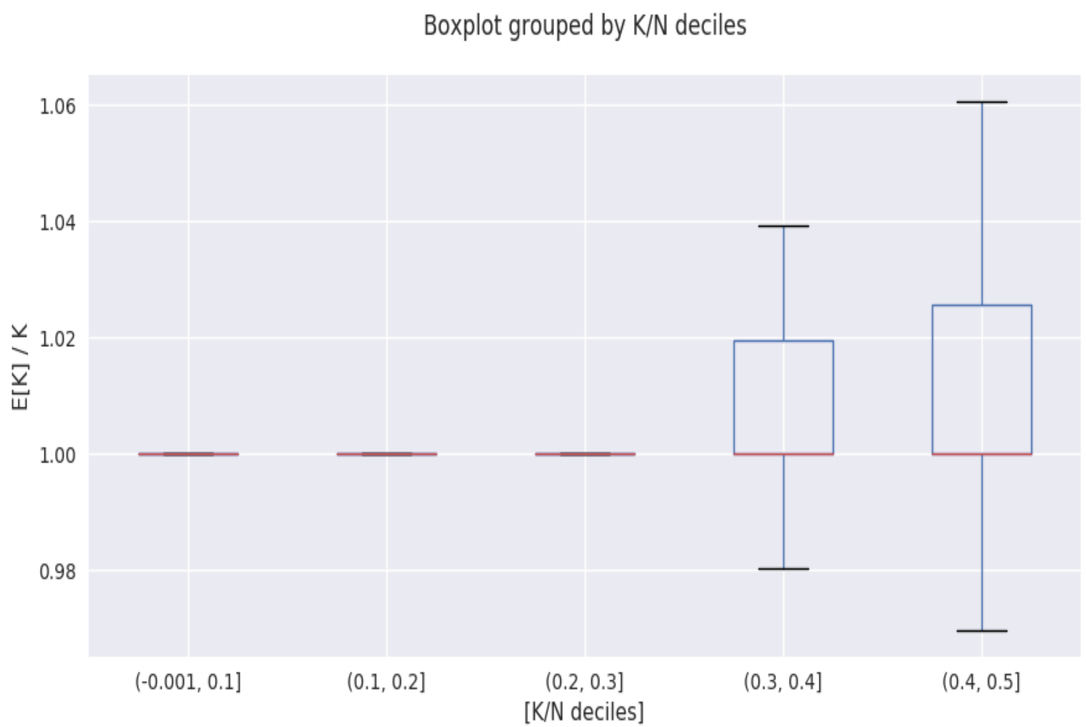

Recursively
Re-evaluate
Clustering




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 $\mathbf{E}\left[\mathbf{V}\left[\left\{\widehat{\boldsymbol{SR}}_{\boldsymbol{k}}\right\}\right]\right]$ (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 $\mathrm{E}\left[\mathrm{V}\left[\left\{\widehat{SR}_k\right\}\right]\right]$ (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.

# Bio

Dr. Marcos López de Prado is the chief executive officer of *True Positive Technologies*. He founded *Guggenheim Partners*' Quantitative Investment Strategies (QIS) business, where he developed high-capacity machine learning (ML) strategies that consistently delivered superior risk-adjusted returns. After managing up to $13 billion in assets, Marcos acquired QIS and spun-out that business from Guggenheim in 2018.

Since 2010, Marcos has been a research fellow at *Lawrence Berkeley National Laboratory* (U.S. Department of Energy, Office of Science). One of the top-10 most read authors in finance (SSRN's rankings), he has published dozens of scientific articles on ML and supercomputing in the leading academic journals, and he holds multiple international patent applications on algorithmic trading.

Marcos earned a PhD in Financial Economics (2003), a second PhD in Mathematical Finance (2011) from *Universidad Complutense de Madrid*, and is a recipient of Spain's National Award for Academic Excellence (1999). He completed his post-doctoral research at *Harvard University* and *Cornell University*, where he teaches a Financial ML course at the School of Engineering. Marcos has an Erdős #2 and an Einstein #4 according to the *American Mathematical Society*.

# 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.
- 