Three Quant Lessons from COVID-19

Prof. Marcos López de Prado

Advances in Financial Machine Learning

ORIE 5256

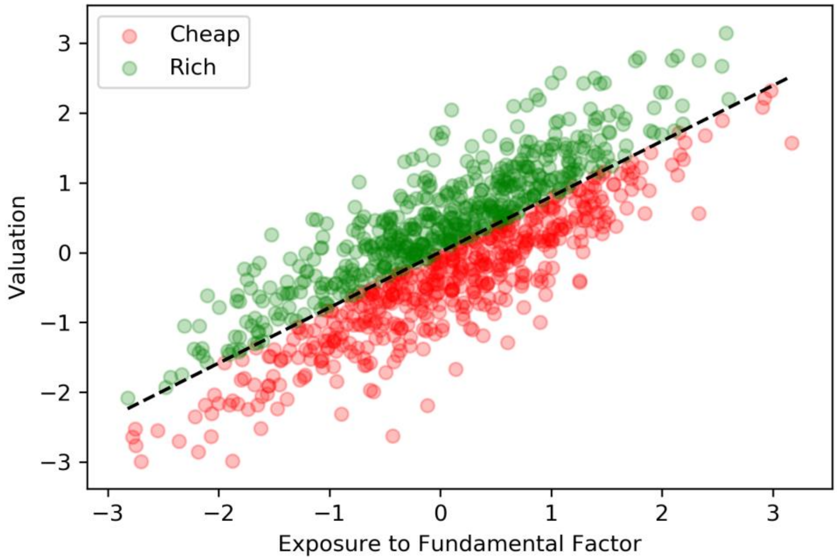
Background

- The SARS-CoV-2 virus was isolated and named on February 11, 2020
- Eight days later, the Standard and Poor’s 500 index reached an all-time close level at 3393.52
- Many quantitative firms have suffered substantial losses during the subsequent selloff
  - This is true even among funds that offer market neutral strategies
  - At the same time, market makers have enjoyed above-average profits during the selloff. Why?
- **What lessons can we learn amid this crisis?**
  1. **More nowcasting, less forecasting**
  2. **Develop theories, not trading rules**
  3. **Avoid all-regime strategies**
- The following observations are the result of joint work with Prof. Alexander Lipton

Lesson #1
More Nowcasting, Less Forecasting

Forecasting is the Past…

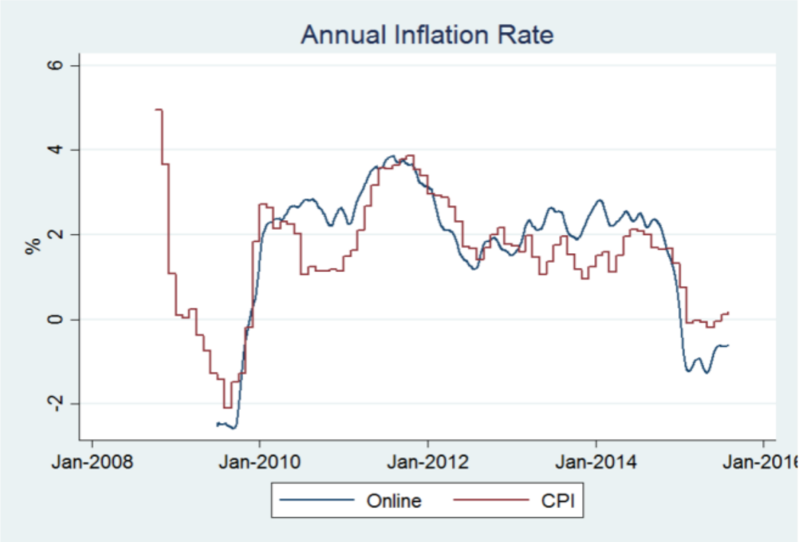
- Forecasting models use structured data to make long-range predictions
- Traditionally, quant strategies have focused on forecasting prices
  - based on price time-series dynamics (e.g., stat arb, CTAs)
  - based on cross-sectional data (e.g., asset pricing, factor investing)
- Forecasting relies on statistical relationships between lagged observations and future outcomes
  - These relationships do not always hold
  - Forecasting made a lot of sense years ago, when datasets were limited, and disclosures were few and infrequent



Factor investing strategies often determine the cheapness or richness of a security as a function of their exposure to a few fundamental factors, which are reported infrequently. **These models do not adjust quickly to changing market conditions**.

… Nowcasting is the Future

- Nowcasting models use unstructured datasets to make
  - **direct measurements**: The target variable is directly observed (e.g. the basket of products used to estimate inflation)
  - **short-range predictions**: The target variable is not directly observed (e.g., parking lot occupancy used to estimate revenue)
- Advantages relative to forecasts:
  - Direct measurements always hold true (they do not rely on a statistical lead-lag relationship)
  - Short-range predictions are statistically more reliable than long-range predictions
  - In both cases, estimates involve millions of recent observations

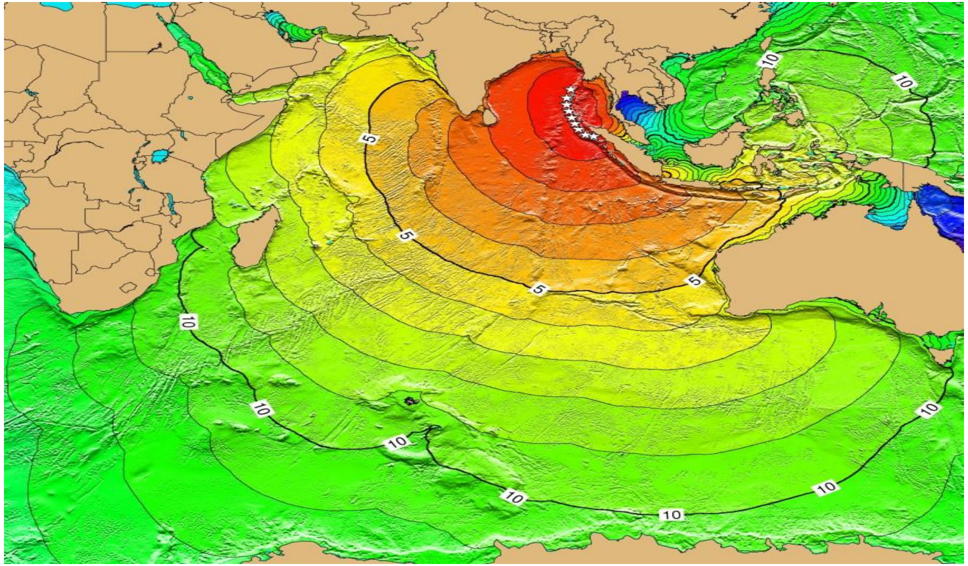


Source: [Cavallo and Rigobon [2016]]

The Billion Price Project collects daily price fluctuations associated with tens of millions of products sold by thousands of online retailers in almost 100 countries.

A Scientific Application of Nowcasting

- It is virtually impossible to predict the time and location of an earthquake
- Instead, scientists have developed early warning systems
  - Even a few seconds of warning can save thousands of lives
- Once the earthquake is detected, it is possible to determine with high accuracy
  - the cities that will be impacted by the shockwave
  - the coastal cities that will be impacted by tsunamis

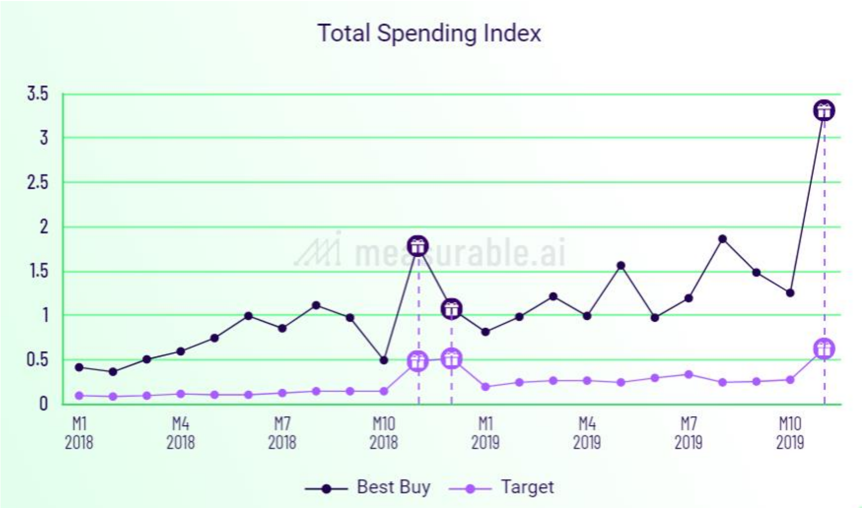


Source: NOAA

A 9.1 Mw earthquake occurred in Sumatra (Indonesia) on December 26, 2004. The ensuing tsunami killed 227,898 people and displaced over a million. Australia received a warning 5 hours ahead, and African countries 10 hours ahead.

Examples of Nowcasting in Finance

- Forecasting is the mathematical analogue of guessing
  - Do not forecast what you can nowcast
- Examples of financial nowcasts include:
  - **Inflation**, based on web-scraping millions of online prices every day, are much more accurate than the forecasts derived from convoluted econometric models
  - **Liquidity conditions**, based on millions of daily FIX messages from market participants
  - **Earning of retailers**, based on satellite images of parking lot occupancy, and email receipts
  - **Industrial production**, based on engineering datasets, cargo shipments, auto production numbers, electricity consumption, etc.

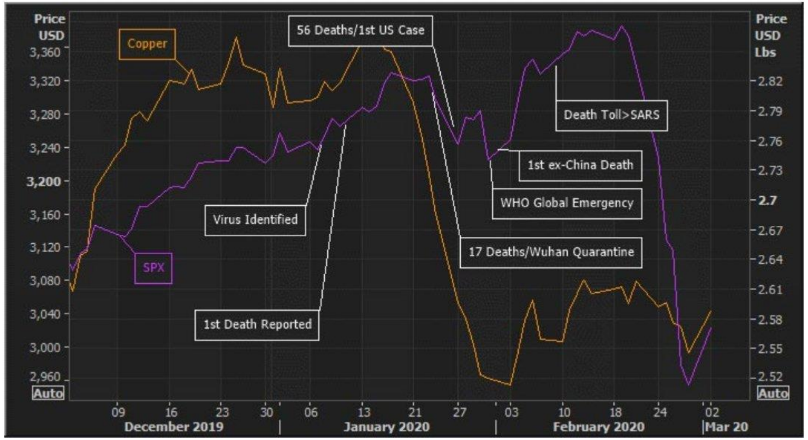


Best Buy experienced a large revenue increase in the 2019 holiday season, clearly outpacing Target. Measurable AI used e-mail receipts to report this information, months ahead of the regulatory filings for Q4 of 2019.

Nowcasting Black Swans

- Days before the COVID-19 selloff started, there were plenty of warning signs that the virus was disrupting critical supply chains in China
- Thanks to their nowcasting of orderflow imbalance, **very few market makers experienced losses during the selloff**
- This selloff may have been a **Black Swan** to market forecasters, but to market nowcasters, it was **a White Swan**
- It is time for quants to pay less attention to crystal balls and add nowcasting to their arsenal

Figure 1: Coronavirus (COVID-19) timeline and market (lack of?) response



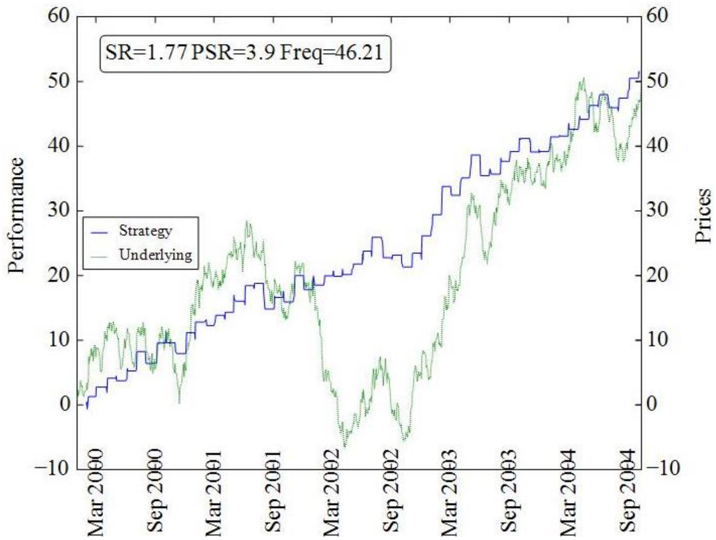
Source: Eikon

On January 23, 2020, the central government of China imposed a lockdown in Wuhan and other cities in Hubei province. The sudden drop in factory demand caused a selloff in various commodities, while U.S. stocks continued to rally.

Lesson #2
Develop Theories, Not Trading Rules

Backtest Overfitting Everywhere

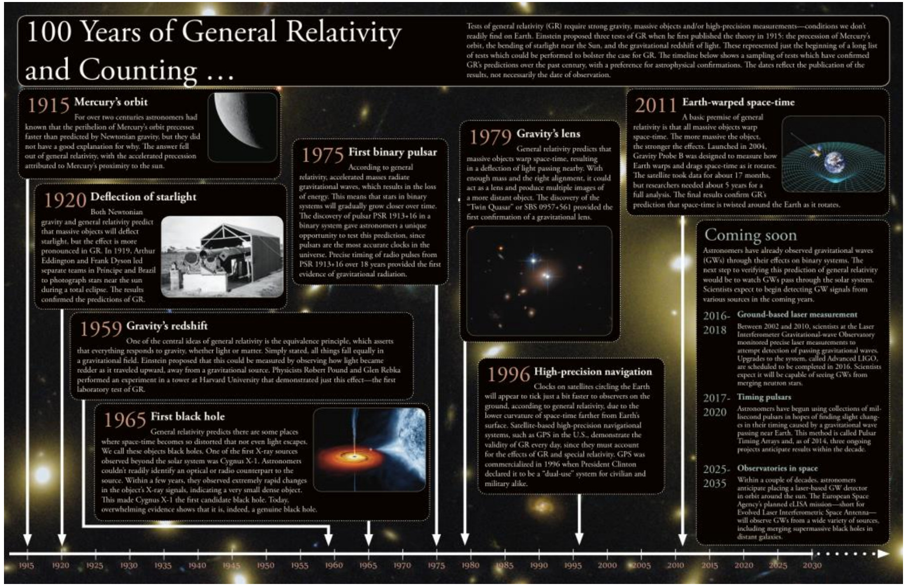
- When correctly done, backtesting is a useful *validation* tool
- It is common for academics and practitioners to run tens of thousands of historical backtests in order to identify a promising investment strategy
  - The best performing backtest is then reported as if a single trial had taken place, and selected for publication, or for launching a new fund
- As a result of this selection bias, [most published discoveries in finance are false]()
  - This fact explains why many funds have not performed as expected, including but not limited to the recent performance of quant funds during the COVID-19 crisis



It is trivial to produce a historical walk-forward backtest with a [high Sharpe ratio](), by trying thousands of alternative model specifications. Virtually no academic papers report the number of trials involved in a discovery.

Backtesting is Validation, Not Research

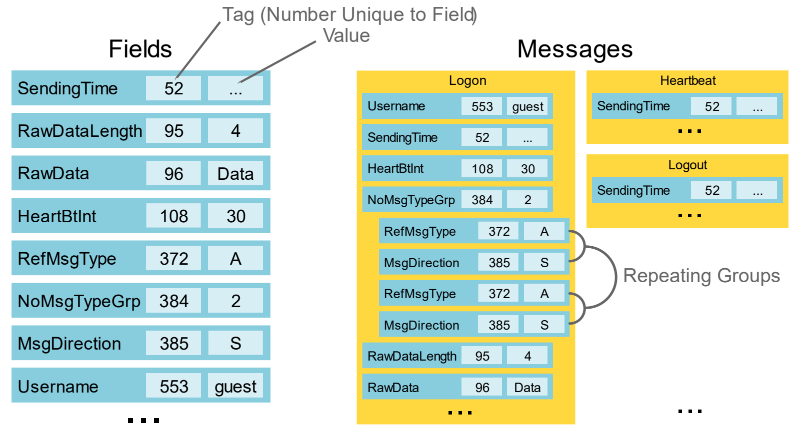
- In the scientific method, testing plays a critical role in attempting to *refute* a hypothesis
- In finance, however, **researchers have used backtesting for the opposite objective**, i.e., for *building* trading rules
- This misuse leads to a circular argument
  1. A researcher backtests thousands of trading rules (e.g., the factor zoo)
  2. The best performing rule is proposed as a hypothesis (e.g., buy low P/E stocks)
  3. The researcher publishes his hypothesis, and presents as evidence the same backtest that he used to find the hypothesis



Scientists can't bake their cake and eat it too. A scientist must formulate a hypothesis, so that colleagues can [independently test]() it. Scientists cannot pick how their theories will be tested (backtesting or otherwise).

Develop Theories, Not Trading Rules

- Researchers should **develop theories** without backtesting
  - E.g., with feature importance analysis methods that are robust to overfitting
- A functional theory explains a phenomenon by exposing **a precise cause-effect mechanism**
- The validity of this cause-effect mechanism must be tested through
  - Backtesting (adjusted for selection bias), and
  - Collecting evidence against the ultimate implications of the proposed theory
- **Backtesting trading rules is not enough**



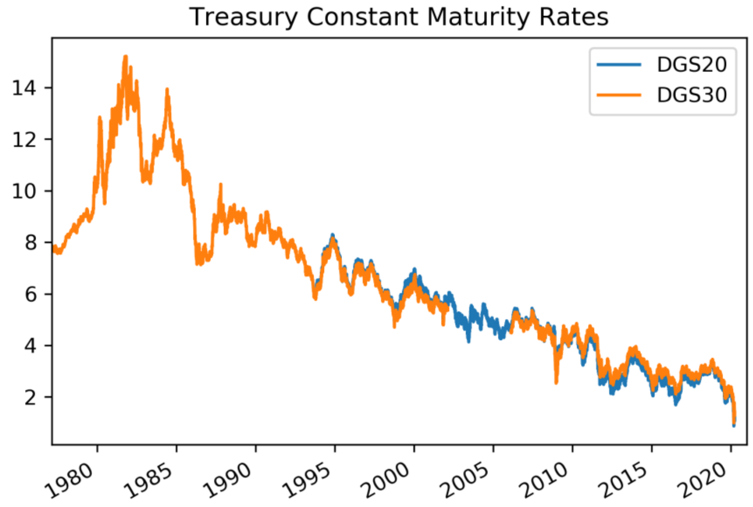
Source: CodeOfView.com

Suppose a theory that explains the COVID-19 selloff as a market panic, similar to the Flash Crash of 2010. Backtesting a strategy that profited from both events does not validate the theory. A better test is to search for evidence of panic behavior in the FIX messages of both days (e.g., market makers becoming like liquidity takers).

Lesson #3
Avoid All-Regime Strategies

The Futile Search for The Holy Grail

- Academics and practitioners usually search for investment strategies that would have performed well across all market regimes
- The likelihood that genuine “all-regime” strategies exist is rather slim, because
  - markets are adaptive
  - investors learn from mistakes
- Even if all-regime strategies existed, they are likely to be a rather insignificant subset of the population of strategies that work across one or more regimes

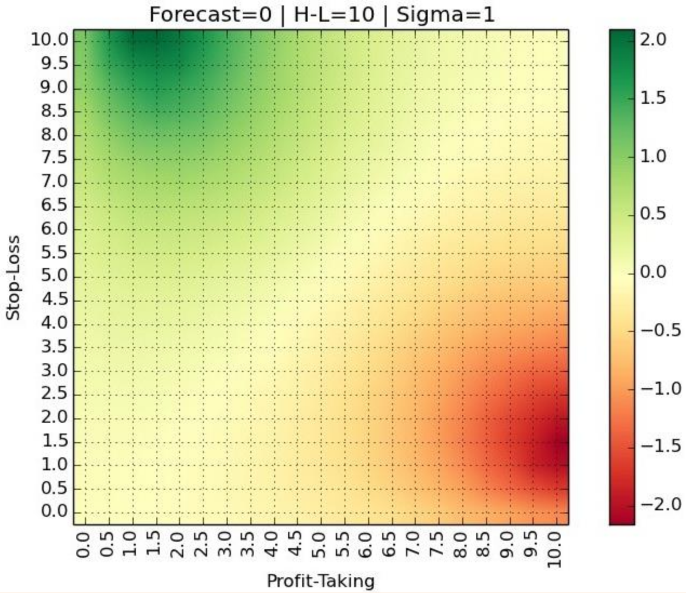


Source: Federal Reserve Bank of St. Louis. [Treasury Constant Maturity Rates]

Most strategies are backtested over decades (sometimes even centuries!) to imply that they work under all market regimes. And yet, the current zero-rate environment makes those backtests unrepresentative.

Regime-Specific Investment Strategies

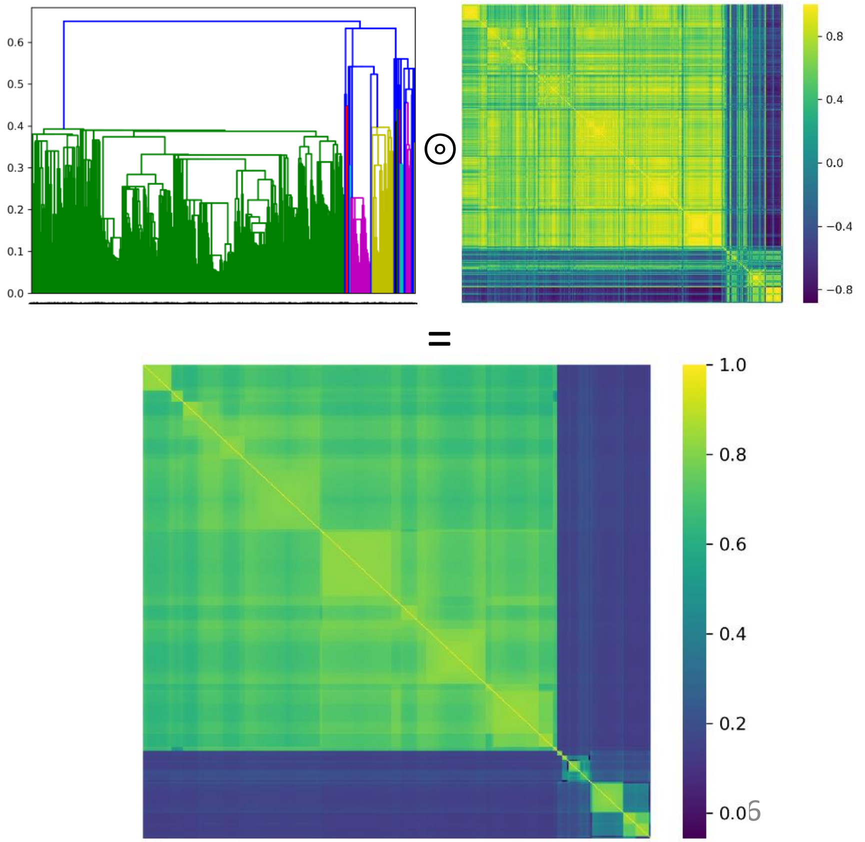
- Asset managers should focus their efforts on searching for investment strategies that perform optimally under [specific market regimes]
  - Each regime is characterized by a particular data generating process (DGP)
  - We can nowcast the probability that current observations are being drawn from each DGP, and use those probabilities to build an ensemble portfolio of those optimal strategies
- This approach allows funds to adapt as market conditions change
  - E.g., from risk-on to risk-off strategies in the advent of COVID-19



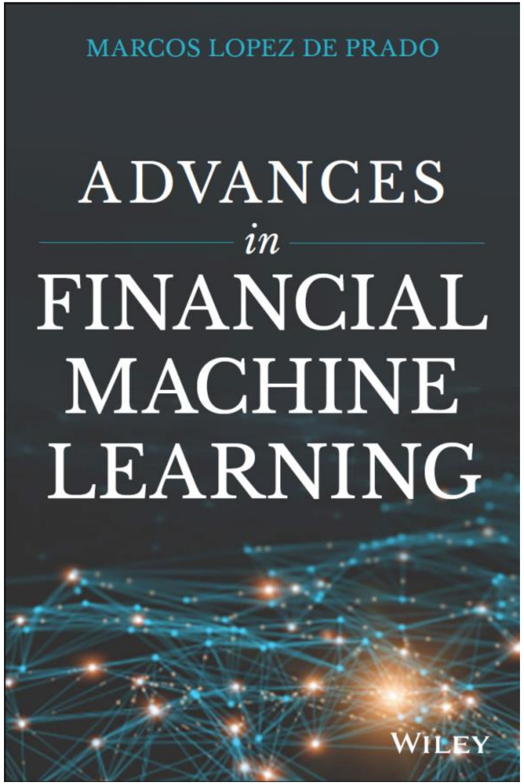
Regime-specific investment strategies are common among market makers. It allows them to adapt to new market conditions quickly. Thanks to nowcasting, funds can apply the same approach to strategy deployment.

Knowledge Graphs

- A knowledge graph links companies based on multiple criteria, such as
  - business ties, supply chain, competitors, industrial sectors, shared ownership, etc.
- Knowledge graphs encode forward-looking information about a regime
- We can use this information to derive [theory-implied correlation matrices]() (TICs)
- A TIC blends theoretical (forward-looking) codependence structure with empirical observations drawn from history
  - This allows for risk models to be instantly updated, with reduced noise



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.
-

www.QuantResearch.org