

Review of the book
Advances in Financial Machine Learning
by Marcos Lopez de Prado, Wiley 2018

Reviewed by Peter Schwendner, Zurich University of Applied Sciences

Marcos Lopez de Prado defines his book as “a research manual for teams, not for individuals”. Hopefully, not only quant specialists of investment and trading companies will read the book, but also their senior managers will adapt its main findings. Already in the first chapter – the “preamble” – de Prado lays the cornerstones of a successful structure of a quant research setup and outlines the most common management mistakes of implementing machine learning across the investment process.

In his line of thoughts, the most notorious management mistake is to treat quant specialists like discretionary managers. Unlike discretionary managers, who are usually kept in competing silos in order not to dilute their specific edge, quants should specialise at a certain task of the whole value chain and validate and support each other to enable a professional production line. The reason for this different approach is the increasing complexity of the various steps due to the trend to include more complex data. In a market of decreasing alpha, this is a competitive advantage.

The suggested “production chain” of quantitative strategies consists of data curators, feature analysts, strategists, backtesting specialists, a deployment team and a separate portfolio oversight responsibility. Notably, colleagues from sales and marketing are not part of this production chain, although they will be responsible for raising assets for the client product. This process setup contrasts the current trend in financial firms of moving from investment-centric to client-centric businesses.

De Prado has a strong view of the use of backtests in the financial industry. He points out that each tested strategy variation is a hypothesis, and that a not strictly planned loop of sequential testing and “improving” a strategy leads to a significant overfitting bias that will likely render the result as worthless. The more careful and diligent the strategy development process is, the more realistic (read: lower) the shown expected value of its performance will be. Therefore, under severe time and performance pressure, even a very highly qualified, but isolated PhD will likely fail to develop a successful quant-driven investment process. This flawed R&D process concerns colleagues working in the financial industry as well as in academia.

Machine learning is not simply a more casual alternative to econometrics without proper hypotheses testing but can help to professionalise the production chain. The author presents concrete tools and algorithms that make it possible to improve the structure and labelling of

data, to find meaningful features, to generate synthetic data for additional out-of-sample testing, to overcome unstable covariance matrices, and to quantify the damned overfitting bias. To do so, de Prado relies not only on own well-known scientific contributions like hierarchical risk parity (HRP) and the “probability of backtest overfitting”, but refers to a full selection of the most important literature of this field at the end of each chapter and even points to specific helpful Python libraries.

The book offers a rigorous mathematical presentation and compact code snippets that profit from the rigorous mathematical education and in-depth quantitative finance experience of the reader. In other words: this is not an introductory book¹. For readers with the necessary background, it delivers on the promise to “cross the proverbial divide that separates academia and the industry” written on the flap. We are glad to have already spotted several third-party Github repositories and blogs that explore, explain and apply the rich material of this book. If used as intended – as a workbook for teams, and then communicated to senior management – it can have a substantial impact on the seriousness of investment product development, and on investors’ confidence in the resulting products.

¹ As an introduction to the statistical problems of developing investment strategies, we recommend David Aronson's book “Evidence-Based Technical Analysis: Applying the Scientific Method and Statistical Inference to Trading Signals” (Wiley, 2006).