Deep Science Publishing
https://doi.org/10.70593/978-93-49910-19-5_8

Chapter 8: Predictive analytics for market volatility, trading algorithms, and liquidity risk management

8.1. Introduction to Predictive Analytics

Predictive analytics can be defined as a fact-based decision-making process to gain insights about expected future events and decision outputs. It involves building statistical models using data and implementing simulations and scenarios to predict outcome probabilities. It is one of the three perspectives of business analytics, the other two being performance or descriptive analytics and exploratory or prescriptive analytics (Zhang et al., 2005; Foucault et al., 2013). Predictive analytics basically answers the questions: What is likely to happen? What can be the reasonable outcomes for a decision? What are the expected probabilities for all possible alternative outcomes? Why do I need to predict?

The answer to the first question is easy. For any decision we make now, there will be consequences in the future. Many decisions in industries, such as financial services, insurance, healthcare and marketing depend on future events and their impact. For instance, to improve risk management, a loan officer might want to know the likelihood of default during the term of a loan before approving a loan application. To improve profitability on credit cards, a bank may want to know who is likely to use up their accumulated rewards points before expiration. For guiding capital allocation decisions, an insurance executive may want to know the expected claims for each insurance policyholder in next twelve months (Cont, 2001; Aldridge, 2013; Avramov et al., 2021).

8.1.1. Overview of Predictive Analytics in Financial Markets

This chapter deals with predictive analytics, which is the most discussed area in data science. We discover analytics in a broad and applied sense, which is motivated by the need and the demand for practical solutions. We discuss predictive analytics in the context of financial markets, and focus on regression analysis, which tries to figure out the relationship between two or more variables. Data exploration and description are also discussed, as this is the first step to predictive analytics. Predictive analytics is proposing that a significant cause-and-effect relationship exists which can actually produce valid predictions. As markets are very noisy and any relationship only exists for a certain time, we show an application for change in variance and benchmark against common filters. Thus, for predictive analytics at the upper management level it is recommended to frequently re-explore descriptive analytics and conditionally switch between active predictors to tune our model.

Predictive analytics is the ability to determine the likelihood of future outcomes based on historical data. Predictive analytics uses statistical techniques from data mining, machine learning and game theory to identify the likelihood of future outcomes based on historical data. Predictive analytics provides the business and business analyst functional user community. The success of an organization depends not only on analyzing what happened and why, but also on forecasting what will happen. Predictive analytics plays an important role in decision making and forecasting for their companies with better accuracy and confidence. Maintaining and training itself requires overheads and thus predictive analytics is mostly applied in areas with several time-series.

Fig 8 . 1 : Predictive Analytics for Market Volatility and Trading Algorithms

8.2. Understanding Market Volatility

Defining Market Volatility

Most financial practitioners define volatility as simply the variation in the price of securities in financial markets like stocks and derivatives. This definition is very broad as it encompasses all variations in price, both large and small, and does not stipulate any time frame. Close inspection, however, reveals that in much of the literature which makes use of the term without further explanation particularly in finance, the authors are generally referring to the tendency of price returns around a series of heftier price moves over time, of varying time frames. Of course, the above definition is not strictly correct. The price process is influenced by not only jump and continuance behavior; it can also experience periods of minimal movement. During these periods the process will practically track a horizontal line. However, the above definition does fit the popular usage. It is also important to home in on the time period that we are concerned about when we consider the behavior of volatility. Much financial modeling and option pricing is based on volatility characteristics observed on very short times frames.

The above definition of volatility draws attention to the word return. That is, absolute price movements in themselves do not constitute volatility. Only relative or percentage price movements actually entered in the definition do. One reason for drawing attention to this definition is that a number of analysts who are, quite rightly, baffled by the deficiencies of various price models attempt to model volatility behavior without recourse to the definition which relates it to return behavior. In general, one has to model the returns first - in particular, the correlation across time - before one can model volatility. Volatility is an input to derivative pricing formulae and risk measurement approaches. In that sense, volatility is just something to be predicted. However, the most common and well-accepted description of market dynamics is modeling the returns.

8.2.1. Defining Market Volatility

Market volatility refers to the degree of variation in the financial prices of an asset over time and represents an important characteristic of financial time series. A high volatility environment usually makes investors feel uncomfortable because the willingness to pay of sellers is usually lower than the willingness to pay of buyers for the price to fluctuate too much. Therefore, volatility has a major effect on the evaluation of many derivative financial products, such as options and futures. Some empirical studies indicate that the increase in volatility has an effect on the usefulness of volatility for the actual prediction of returns, the estimation of the risk premium, and the prediction of stock market returns at future time periods shows no improvement over using returns volatility estimation applied on data with a period shorter than or equal to one month, and the time series of return volatility should be a prediction variable to obtain a superior forecast accuracy over several forecast horizons of the actual month-ahead returns in specific countries with developed stock markets.

Although in practice volatility can be defined in many different ways, the most common methodology for this purpose is by means of the standard deviation of financial returns. Thus, it is possible to define volatility as the fluctuations over time of a given price variable price, or the volatility of the financial return if one considers in time the first-differenced of a stochastic variable price that is assumed to follow a geometric random walk. Yet, the theoretical background of this type of estimation is not as solid as the one supporting the use of the average absolute or squared returns. The average squared return has become, for many years, the most common in academic and practitioner settings.

8.2.2. Historical Perspectives on Volatility

John Stuart Mill discussed in 1844 the idea of a return of events to a certain proportion bearing the relations of cause and effect relative to probable future events. W. S. Jevons applied this idea to the prices of commodities before 1800. Commenting on the statistical information contained in the early work of Pareto, Yule maintained that it enabled him to come to the surprising conclusion that, although the fluctuations were purely empirical laws, and therefore free from theoretical difficulties and reference to definite principles, it should be expected that future values of prices should display, in a certain measure, the same characteristics as historical values. Somewhat later, in 1873, the Swiss engineer Orest Chrevon stated that his harmonic analysis enables us to discover cycles even long before the data appeal to us as cyclical. More recently, Frisch, in 1932 and Yule in 1927 attempted to show that economic and sociological phenomena vividly illustrate the properties of stationary stochastic processes.

In the early 1960's, the consideration of the classical filtering theory stimulated considerable interest in the time-varying volatile systems by engineers and scientists. Indeed, it is a very natural conjecture that the variances are stochastic processes, the historic time span of the typical evidence being relatively short compared to the time span over which the variances are evolving, and the variances lacking the concept of a stationary distribution. Since, in the absence of a strong theoretical foundation which provides the framework within which we may test the presence of time-varying volatility, questions as to choice of confidence interval and of optimal Lag length for the construction of confidence intervals are important practical considerations in the application of the classical assumption of constant variance. Indeed, the design and use of a diagnostic test for constancy of variance assumes great importance.

8.3. The Role of Trading Algorithms

In today's financial markets, it is rare to find a trader executing his own operations, as individuals utilized to do in the past. Majority of trading in financial markets is nowadays conducted by machines and algorithms running on automated programs, shifting the focus of trading from the physical exchange floor to the electronic marketplace. By using advanced algorithms and relying almost entirely on the predictions produced by technologically advanced software, financial institutions have succeeded in improving the trading process altogether. The set of algorithms utilized in such automated trading is extensive and vast in its range of different functionalities, covering aspects such as trade generation, execution, portfolio optimization, trading scheduling or arbitrage exploitation, among others. Nevertheless, the most crucial role of algorithms, as the ones we are concerned with in this section, is as market traders. These algorithms are responsible for monitoring markets continuously, incurring in trades as soon as new opportunities arise. In a trading market populated by such state of the art algorithms, there are no subjective elements left in the trading decision process. Indeed, machines are thought to pursue the ultimate goal of maximizing profit through an automatic financial management command. But by seeking to maximize the banks and financial institutions' benefits over specific time periods, clear to the goals established by each company, what these algorithms are really achieving is to amplify the overall activity level in the markets, and thus amplify liquidity and volatility. By acting as instant market makers or speed bumps, they ensure a fast, efficient and dynamic trading process.

8.3.1. Types of Trading Algorithms

Algorithmic trading has become routine, with thousands of firms and many more individual traders using so-called "trading algorithms" to facilitate their financial market transactions. Broadly speaking, trading algorithms can be categorized into three areas. First and foremost, there are those algorithms that employ academically established and implemented quantitative models and strategies, especially those implemented for trading securities; these rationale-based algorithms are called market making, pairs trading, statistical arbitrage and so on. They account for most of the trading volume on the various electronic exchanges around the world. In this section, we explore these rationale-based algorithms some more.

They are then joined by other types of algorithms. Some of these "solicitation algorithms" and their more advanced cousins, transaction cost algorithms help traders with experience and/or information on the target markets to solicit liquidity using price spikes or drift; these algorithms provide critical market functions and help keep price movements orderly during times of heavy trading. The price spikes or drift generated by the solicitation algorithms often provide visibility to forgettable buy/sell programs placed by flow traders with no market experience. The one-sided price drift or spikes signal algorithmically based long-term stock movement to the back office of other traders, who then come to the market committing their directional lists, assuming the side of the solicitation algorithm's participants. Although the solicitation type algorithms are critical to maintaining market stability, they account for a significantly smaller percentage of the overall traded volume.

The final family of algorithms is event-based algorithms. These algorithms reveal themselves when significant prices are invariably triggered by industrial finance news reports that report shocking news of mergers and acquisitions or senior management changes in a public corporation. These reports jumpstart event-related directional long or short trading in traded securities.

8.3.2. Algorithmic Trading Strategies

In general terms, one can consider two main algorithmic trading strategies, to which various combinations and specific alterations can be associated. Technically, an algorithm can perform as a market maker, acting passively by providing prices with a markup and waiting for someone to buy at the ask price or sell at the bid price, thus capturing the profit made from the spread. Market making is the earliest and most classic form of algorithmic trading and is the most direct way to take advantage of electronic markets' liquidity. Its science has been perfected by quantitative hedge funds or proprietary trading firms equipped with expensive technology and staffed with expert specialists in both mathematics and finance. Other algorithms can perform as buyers and sellers in the market, attempting to time the decision on when to enter and/or exit the trading process in order to maximize their profit. This timing is strictly related to the forecast of short-term price movements and consists of the following practical implications: Are short-term price movements predictable? If so, are prices likely to move upward or downward? When should short-term price forecasting begin in relation to actual buying or selling? When should the sequence of trading execution begin in relation to the price movement being forecast? And what trading volume should the profit-seeking trader intend to sell, in relation to the forecast price move? Finally, are the long-term price predictions, typically associated with price forecasting over longer time horizons (days, weeks, months, or years), initially obtained from pattern recognizers or at a very basic level from price filters, by which the positive and negative value signals used to extract excess returns are based on crossovers of longer- and shorter-term moving averages? Models can classify into systematic strategies that generate signals for quant funds and discretionary strategies that trigger trade actions for discretionary traders.

8.3.3. Impact of Algorithms on Market Dynamics

A serious issue is that the efficacy of the algorithms is changing the market behavior. This behavior changes the rules relating to arbitrary weights attributed to each limit order and corresponding costs are responsible for creating an order book. Algorithms, which react to limit orders and "hit" them creating liquidity instantly, are changing market dynamics in a permanent way. Thus, HFT liquidity creating instantly trades profit from what could be an edge, by extremely fast reactions, drawn from their capacity of quick sight does exist. Consequences of such behavior is always a doubt: could this be some kind of instability on a switch point? Consequently, if installed such a switch point, could the leading banks? The only answer could be Yes, it's possible. It has even been proven that if the order book is ruled by algorithms then such liquidity created at an infinitesimal cost, then a tsunami has also been predicted.

Another way of ordering has even found this affirmation using order flow for exchanges. But speaking of deterministic order flow, it would be better to speak of exchange internal order flow and use order book for exchanges in order to predict sudden impacts. This being the case, then algorithms should be protected against catastrophe creating algorithms. This is quite strange. It would also be a challenge for all risk management based on prediction of impact costs. But this reflection is only the tip of the iceberg. In other words, it is inexcusable not to make a proper risk analysis of creating algorithms internal system. Because what is observed is a collective behavior which could not be thought without speculation. Thus is trading algorithmic. It is this "day" behavior, characterized by curiosity, making an impact on fluctuation which is why of trading speculation. Otherwise said, if the term is defined as being a constant increase of trades activities, this assumption is based on capital gains.

8.4. Liquidity Risk Management

Liquidity risk is an important area of risk management. It relates to the way banks manage risks arising from their engagement in maturity transformation. Banks provide short-term liabilities while making long-term investments. By converting short-term liabilities into long-term assets, banks support investment in the economy and help provide consumers with access to goods and services. However, if a large number of consumers demand to withdraw funds from their bank because they are suddenly short of cash, the bank cannot process these demands, since it has invested deposits in illiquid assets. Liquidity risk occurs when the bank is not able to attract additional deposits at short notice to meet withdrawal demands and not able to liquidate loans because they cannot be sold for cash due to market risk, creating a correlation between market risk and liquidity risk. The traditional wisdom is that liquidity risk is unhedgeable. Banks are regulated to hold additional reserves to manage liquidity risk. Additionally, banks are supervised to perform regular liquidity stress tests as a function of the regulatory financial supervisory authority.

The regulatory requirement is to hold excess reserves to reduce the probability of encountering a liquidity crisis when consumer demand to withdraw deposits exceeds the short-term assets and cash reserves the bank holds. Various parameters affect the cash reserves, including the expected withdrawal demand, the bank's ability to liquidate risky long-term assets, and how long it will take for the bank to get additional cash from external sources. The bank's holdup in the value and risk of its short-term assets, expressed in cash reserves, significantly limits its capacity to manage liquidity risk, not only since having only monetary assets means having low net income, but also since having unstable cash reserves makes it unlikely to provide the expected service.

8.4.1. Defining Liquidity Risk

Liquidity is defined as the ease of converting an asset into cash, without incurring a loss. A risk associated with liquidity is that an organization might not have enough liquid assets to meet its obligations. In case, liquid reserves are not sufficient to meet short-term obligations, it needs to sell liquid securities. For any organization on an ordinary day, selling liquid securities should incur little loss. The distinction of the situation causing liquidity risk to become a serious issue is that the organization's securities are not sufficiently liquid, so it ends up incurring a loss on the sale of the securities. Hence liquidity risk arises from having insufficient liquid reserves, and the organization's securities are not sufficiently liquid. Theoretically as a rule of thumb, at least 10% of total long-term expenditure should be maintained in liquid form, on an average basis.

Ensuring that enough relatively liquid bank deposits are available for the organization is usually not a difficult task except in individual crisis situations for banks and other financial intermediaries. However, generally, market liquidity conditions are overlooked in corporate liquidity management, despite the potential rise of very costly market liquidity problems, particularly during times of stress. The problem is that few statistical measures and forecasting models are available, which could help identify vulnerable periods. For loans, the liquidity risk taken depends on the following characteristics: the term of the loan, the currency in which the loan is denominated, prepayment conditions in the loan agreement, and amortization patterns over the life of the loan.

8.4.2. Measuring Liquidity Risk

This section focuses on how to assess liquidity risk, which refers to how difficult it would be to exit, that is to sell, a certain position over a certain time horizon at a given cost.

Unlike other financial risks, a liquidity shortage does not have to come from the assumed position. For example, one can measure the cost of selling a large volume of risky assets by looking at the price impact of executing large volume orders over short time horizons. Thus, an immediate cost of exiting a risky position is expected to be very high when the risk of the position is high and expected to fall toward zero when risk is low. Furthermore, even when one considers the issue of exiting a risky position, during a financial crisis the cost of liquidating that position in a short time horizon becomes very high because other market participants would be unwilling to participate.

Liquidity risk measures are driven by a principle suggesting that the cost of selling or buying large quantities of unpleasant-to-hold assets increases with the amount of such assets held by the investor. Typically, traditional models calculate the cost of having to use market orders for optimal strategic trading, which can be solved by a stochastic dynamic optimization problem. Some advances in this research area have come from the models considering the entire distribution of as yet untaken investment actions, rather than focusing on its mean or variance. These attempts often find justifiable models for calculating estimated market impact costs.

8.4.3. Liquidity Risk Mitigation Strategies

Bayes' theorem tells us that we can compute the probability of uncertain event E having occurred given some evidence or information event I, which is denoted as P(E|I). In financial markets, it is known that we can condition the probabilities of price moves on observed investor behavior such as the volume of buy and sell orders, the observed spreads between bid and ask prices or by the expectations of the underlying distributions. This essentially is the statistical foundation for all risk management models, including liquidity risk.

Equally important to the measurement of liquidity risk are the strategies to mitigate the negative effect of the liquidity risk on the investor's portfolio return and risk. Some of the major liquidity boundary parameters are the position size, overnight holding period, whether the trade is likely to be unidirectional or two-sided, and the investor's specific policy of dealing with temporary marketmaker who can provide liquidity for a limited time period. These boundary conditions differ at any point of time for every investor depending on the investor's expected return and temperament to deal with the variation around that expected return arising from short-term deviation from the long-term asset value, which is essentially the source of risk. Market prices can deviate from the asset value not only due to risk, but also due to market liquidity.

In designing liquidity risk mitigation strategies, the first step is to identify the ratio of liquidity to risk and projected trade path for the strategy being contemplated. For example, investors who voluntarily take on a greater liquidity risk, intending to realize maximum benefits, may be limited in the size of trade in a higher risk, lower liquidity market. Their trade may take longer to execute. The evidence of a long damped trade path is indicated by repeated price anomalies.

8.5. Predictive Models for Market Volatility

Understanding and forecasting volatility dynamics is at the core of most option pricing models even as volatility is unobservable and need to be estimated from the observed prices. Models of this popular form are called log-linear volatility or generalized autoregressive conditional heteroskedasticity predictor, which are popular estimators of both ex ante and ex post volatility. In the GARCH, the conditional variance depends on lags of the process itself and lags of error terms. The estimation is carried out using restricted maximum likelihood constrained for the coefficients’ sum to be unity.

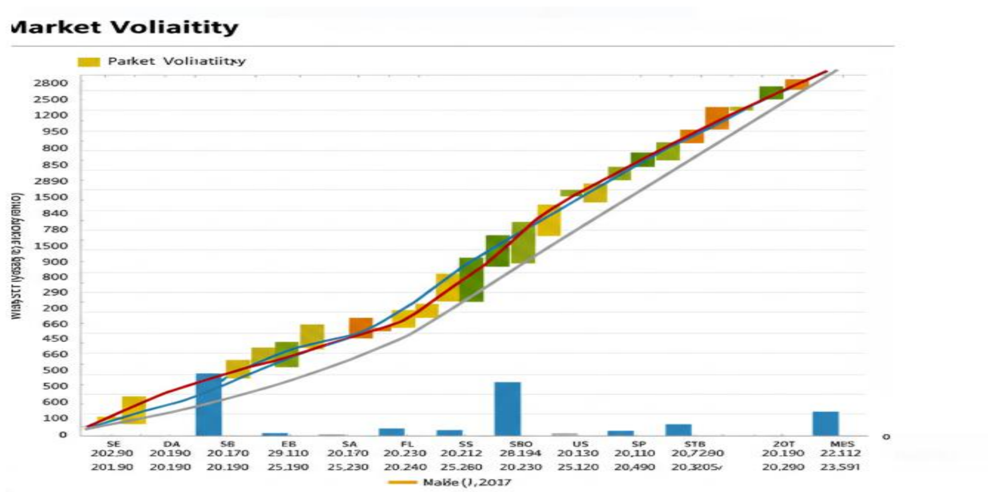

Fig 8 . 2 : Market Volatility

Generalizations exist for both the mean and variance processes with the mean process allowing for a deterministic component, regression on other exogenous variables, non-Gaussian distribution, time varying coefficient model structure and also allowing for long memory features. These dynamic regression models can also be run in a structural form to allow for correlations in prediction error terms stemming from two or more related time series. It is also not necessary for the mean effects to be stationary for the forecasting. Volatility prediction models have since then continued to largely utilize GARCH, SV, DLM and other extensions for accuracy.

8.5.1. Statistical Methods

Models of financial time series emerged within the overwhelming progress of the empirical research dedicated to return distributions' anomaly detection during the late 1980s. Since then, volatility modeling has been a hot research topic in quantitative finance. Returns which follow a non-normal distribution are the starting point of the large field of empirical studies on return distributions, referred to as the stylized facts on financial returns. Further studies on volatility estimation by Autoregressive Conditional Heteroscedasticity (ARCH) are proposed, as well as the Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model, which became the first empirical framework to treat the systematic factor affecting volatility as causal. The systemic approach is consolidated through the Multivariate GARCH (BEKK) model proposal, which merges the analysis on security systemic relationships with volatility pattern modeling.

8.5.2. Machine Learning Approaches

The existence of a nonlinear component in the volatility structure is evidenced by the success of different neural network architectures in volatility modeling. Generally, neural networks are used to correct residuals of other models' fit. Also, these correcting neural networks can be used in sequence in the outputs of two models to capture both linear and nonlinear effects. NNs can simultaneously capture nonlinear effects and correct different situations that originate imperfect descriptive functions of the underlying market mechanisms because of its capacity for self-adaptation in nonstationary environments. NNs have been expressed as universal approximators. There are several reports of big successes in variance and volatility predictions with NNs. We consider all the mentioned papers with daily and intraday horizons, but we put more emphasis on volatility predictions with daily horizons.

The obvious questions now are: what is the essence of all these superior results? What is the reason for these particular first best scenarios? The available answer is that the NNs may even model properly the noise of the process represented by the model prediction errors. The problem that we face in choosing a machine learning architecture, to predict volatility of returns, is that the predictive function must satisfy the nonnegativity constraint. The volatility predicted values must also capture the extreme values verified in the historical empirical volatility series. There are several machine learning predictions that equipped with modeling corrections may help model the volatility of financial returns. These predictions differ in architecture and in the modeling concepts they apply.

8.5.3. Comparative Analysis of Predictive Models

The following models were examined to compare their predicative ability on forex time series: Random Walk, ARIMA, Adaptive ARIMA, GARCH, ANN, SVM, EN, GMDH, and RBF. The empirical study covers a prediction interval of 1 to 10 days ahead in the EUR/USD, GBP/USD, CHF/USD, AUD/USD, and NZD/USD exchange rate markets. In general, it is rich in forecasting time series, prediction time horizons, input variables, and model explanations. It uses three patterns of model evaluation averaging. The relative error metrics are MSE, MAE, and RMSE. The best performance model statistics are pooled and presented for comparison in the conclusion. The choice of error reduction averages impacts consolidation results. Consider forecasting as a large-scale task; we prefer predictive power-focusing error metrics.

Most evaluations pool errors across prediction horizons. In this case, assumptions need to be made as to horizon sensitivity. To avoid biasing the model prediction power, there are two ways to compare models. The first is to take the model that performs best across models evaluated, and the second is to assume that prediction variability across models exists according to a common error distribution. In this case, using an error ratio metric is imperative due to the so-called score synchronization problem. For score pooling across experience patterns to come into effect the prediction error distribution has to be the same for all the scaling errors. Additive errors are varied serially and across scaled tasks. A second approach is choosing a single model based on its prediction performance. In this case, we prefer using a cross-validation procedure to select tuning parameters before application.

8.6. Integrating Predictive Analytics with Trading Algorithms

At a specific level of accuracy, too high or too low, the signals become costly for the investment companies. The trade decisions must consider the risk, fees, and their impact on the financial result so that the signals can become a reliable tool for the investment decision. A trading algorithm is used for trading operation research and optimization. Similar to predictive analytics, the trading algorithm signals need to be optimized for a specific company's strategy. Every trading algorithm complexity requires their signals to be at a specific level of that, too high or too low. The optimization characteristic makes it possible to generalize each approach. The purpose of the strategy is to increase the investor's wealth by increasing the Marey ratio. The integration of the predictive model with the investment strategy is characterized by three levels of data and signal use during the continuous decision-making process.

The features used to build the predictive model can be data of a different type provided by different data sources, e.g., fundamental factors, news, estimates of experiment analysts, social networks, and transaction volume. However, specific factors are considered to be better indicators, e.g., open price, close price, high price, low price, transaction volume, and the publicly available insider information. Predictive model results are of the "yes" or "no" type. At the same time, the investing entropy risk is a market risk the sensitivity of which to microstructure variables is defined by analysis of volatility. Predictive model tests can be performed using data of a certain period. The confirmation of the predictive model with the real market should be done continuously. Predictive analytics signals can be used for pricing or index construction. Predictive modeling is part of the trading algorithm's decision-making process.

8.6.1. Data Sources for Predictive Analytics

New predictive analytics algorithms for forecasting financial time series will not be very effective unless they are provided with high quality data delivered at high speed. Effective predictive timing of financial market is a practical problem requiring realistic solutions. This means that one must consider the technology of doing such predictions not the fundamental theory of finance and economics.

For example, our main focus is on the practical side of applying machine learning to predicting the future for financial investment. The investment process is a tiny sector of economics involving specialized participants who play according to their own rules. Therefore it does not require a general theory for financial economics to be successful.

Clearly the prediction must be made with data from the immediate past. It is important to identify which data has been actually available at the time of the prediction, not data which has subsequently become available at a later time. For example the published closing price is generally available for trading algorithms to use. In contrast, the published daily high and low prices have been adjusted subsequently to delete erroneous outliers.

The key issues are what real-time data to consider, how to process such data, and ultimately how to use the information to motive rules for market timing predictions. The question of what data sources is vital. In particular, high speed delivery of high quality data is crucial for quantum trading where trades are executed once every second.

8.6.2. Real-time Data Processing

Unlike data collection for predictive analytics projects, which is done in a relatively academic fashion, integration of those models into trading systems requires the implementation of tools capable of processing information in real-time. Practical considerations such as minimizing latency while implementing a processing engine capable of ingesting the massive flow of news hits and corporate disclosures as they become public require the specification of the type of data to be processed, the complexity of the transformations required, and constraints on the algorithm components that process each individual piece of information. In our case, trading-system processing speed has priority over advanced functionalities, such as incremental learning of new topics and/or event types through online supervised classification, knowledge extraction, relation detection, and inference making. The primary purpose of the processing system is to transform unstructured text data into structured inputs for the trading system allowing real-time decision-making.

Centralized or distributed processing capabilities need to be implemented to aggregate factors across thousands of features, which we process in real-time. A scalable data collection framework must leverage on a cloud-enabled architecture using either major providers capable of handling massive datasets or cost-effective open-source distributed processing clusters. Advanced automated hardware virtualization technologies, coupled with a regenerable processing architecture would ensure a cost-effective solution for managing large temporal data potentially subject to exponential usage growth. Consideration must also be given both to collaborative environments for notifications and alerts, and to the dashboards and reporting specifications required for monitoring the processing and its use of cloud resources.

8.6.3. Backtesting Trading Algorithms

A trading algorithm is evaluated by trading it on historical data before being traded live. A common method for backtesting a trading algorithm calculates its fixed-length position over the backtest sample and uses it as if it traded during the sample. Trading decisions made along the backtest may invoke a number of model bias issues, resulting in net-excess performance that is possibly not realizable in a live trading setup. For example, estimates of predictive accuracy may be formed from the same sample on which the trades are based, inflating the realized net-excess returns. Predictive analytics estimates may also be reconsidered as the trader observes the performance of the strategy over time, whether directly or through a live trading implementation.

If a strategy trades more than once in the sample, realized returns can be further compromised. Trading during lunch in markets that are characterized by occasional bursts of volatility risks realizable returns not matching backtested returns that are predicated on the magnitude of these bursts matching the schedule of trades. Furthermore, trading strategies that are of longer duration will likely determine a trader’s sales at far in advance of execution. Backtest estimates of success for the strategy will disproportionately suffer from the delays that are managed only for discrete trades in the shorter-duration scenarios.

Trading strategy backtesting is usually less complex than for signal backtesting, simply because of the crude implementation of trades regardless of any outcome expectation and timing precision for their execution that one may be trying to exploit. Accounting for transaction costs, slippage, and market impact has a significant effect on a trading algorithm's performance implementation in practice. Trades initiated by many signals are impractical for trading since they cannot be filled at exact signal price points. Hence, in practice, strategies might be employed for only a small fraction of signal trades, the most significant ones.

Fig 8 . 3 : Financial Market Forecasting with Predictive Analytics

8.7. Conclusion

To summarize, many predictive analytics methods have a long history of use in financial market prediction. However, only a few methods that we describe are being considered for real-world applications, like Bayesian DDM, which is widely used in risk control of FX transactions. There are many challenges to developing better predictive analytics methods to respond to those challenges. First, the performance of many statistical and machine learning methods expands widely based on the configuration and parameter selection for the employed method. However, finding the best configuration is a tedious task, especially for real-world applications. The second challenge is related to model overfitting. To mitigate overfitting risks in the training of predictive methods, it is essential to formulate a strategy for dividing the dataset into short-term training, development, and validation periods, mainly driven by the market conditions. A relevant third challenge is an accurate estimation of the uncertainty of the prediction, especially for critical applications. Improving the prediction uncertainty estimation is essential for many practical applications, especially in financial markets.

The reason behind the above challenges relates to financial markets being highly dynamic systems where market conditions constantly change. The pipeline of development and implementation of predictive approaches for financial market prediction is frequently not streamlined due to the external conditions that affect the prediction and the system's requirements. The concrete and unique nature of the financial prediction task implies that an a priori setup of the development pipeline for the prediction is not possible. Thus, the development of better predictive methods of financial markets using predictive analytics will always require an element of human intervention. The human development effort allows the analyst to properly assess and fine-tune every step of the prediction pipeline to achieve better financial risk control. Better predictive analytics methods will directly correlate with the financial market prediction research. An important area of exploration in future works will be how to further improve the human intervention in predictive analytics applied within financial prediction problems.

8.7.1. Key Takeaways and Future Directions in Predictive Analytics

In the recent years, disruptive changes in financial markets have changed the ecosystem of predictive analytics in the financial domain. First, low interest rates and moderation of economic expansion have made it harder to make money by trading equity indexes and in bonds. As a result, equity traders have started using increasingly complex trading strategies in order to generate excessive return from trading small and midcap stocks.

That has made the returns from investing in such stocks highly volatile and stochastic, leading to corresponding volatility of returns from trading in the corresponding stock indexes. Second, synergies between crypto assets and traditional market have led to low correlation of returns from such asset classes. This has opened up the scope for enhanced excess return from traditional stock indexes by tracking or investing in conjunction with crypto indexes.

High frequency finance has started becoming a must in time series predictive analytics. Strikingly, models have been developed recently that use day-ahead returns from multiple assets in highly correlated environments to accurately predict the timeliness of arrival of various high-impact economic events, which can in turn help in predicting asset volatility surrounding the event time. Predictive signal of such models can also potentially be used to optimize the hedging strategy of an asset sensitive to the arrival of a high-impact economic event. We look forward to the expansion of synthetic data generated alongside the information containing the event of interest which can thus help in development of the predictive algorithms based on the convenient architecture of deep learning models.

References

Cont, R. (2001). Empirical Properties of Asset Returns: Stylized Facts and Statistical Issues. Quantitative Finance, 1(2), 223–236. https://doi.org/10.1080/713665670

Foucault, T., Pagano, M., & Röell, A. (2013). Market Liquidity: Theory, Evidence, and Policy. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199936243.001.0001

Zhang, L., Mykland, P. A., & Aït-Sahalia, Y. (2005). A Tale of Two Time Scales: Determining Integrated Volatility With Noisy High-Frequency Data. Journal of the American Statistical Association, 100(472), 1394–1411. https://doi.org/10.1198/016214505000000169

Avramov, D., Cheng, S., & Metzker, L. (2021). Predicting Market Volatility: Machine Learning vs. Economic Variables. Journal of Financial Economics, 139(3), 758–775. https://doi.org/10.1016/j.jfineco.2020.07.010

Aldridge, I. (2013). High-Frequency Trading: A Practical Guide to Algorithmic Strategies and Trading Systems. Wiley. https://doi.org/10.1002/9781118662865