JOURNAL OF ALGEBRAIC STATISTICS
Volume 14, No. 1, 2023, p.142-149
https://publishoa.com
ISSN: 1309-3452

Big Data Analysis in Finance Management

Surendranadha Reddy Byrapu
Northeastern University

Abstract: We explore the utilization of machine learning (ML) techniques within the realm of finance research. Initially, we highlight a crucial distinction: supervised and unsupervised learning, the two primary ML categories, tackle distinct challenges compared to the traditional econometric methods. Subsequently, we delve into the contemporary landscape of ML applications in finance, identifying three primary types: (i) the creation of more advanced and innovative metrics, (ii) the minimization of predictive errors, and (iii) the expansion of the conventional econometric toolkit. By organizing these applications into this framework, we provide insight into potential future avenues for both researchers and practitioners. Our findings underscore the numerous advantages of ML methodologies in contrast to traditional approaches, underscoring the immense potential of ML in shaping the future of financial research.

Introduction: Artificial intelligence has become an integral part of our daily lives, showcasing remarkable applications such as face recognition for secure and efficient airport procedures, voice recognition for seamless interactions with personal assistants on smartphones and smart home devices, and the widespread use of chatbots for swift customer support. Modern artificial intelligence interfaces with nearly everyone multiple times each day. At the heart of artificial intelligence lies machine learning (ML), a technology that empowers machines to perform intricate tasks like facial recognition, speech comprehension, and message responses. Given the robust capabilities of ML, a natural question arises: can ML methods find applications in fields beyond these everyday contexts? This paper explores the utilization of ML to address challenges in finance research.

Numerous comprehensive papers have underscored the potential of ML in the realm of finance. Varian (2014) describes ML as a suitable tool for analyzing vast datasets in economics, offering examples of ML methods and their applications in this field. He also hints at the prospective use of ML in econometrics. Mullainathan and Spiess (2017) identify prediction as the primary domain where ML excels in economics, presenting various categories of existing and potential future applications. Athey and Imbens (2019) shed light on the most pertinent ML techniques from an econometric perspective, while also providing an overview of ML's potential contributions beyond mere prediction, particularly in addressing relationships in economic inquiries.

Although the utilization of Machine Learning (ML) in finance research is still relatively nascent, there has been a remarkable surge in the number of applications harnessing the potential of ML in recent years. To put this growth into perspective, consider that in 2018, the number of ML-related publications more than tripled when compared to the yearly average from 2010 to 2017. In 2019, this increase became even more pronounced, surpassing a fivefold rise. By 2020, the growth was almost sevenfold, and in 2021, the number of publications employing ML was nearly eleven times greater than before. This exponential expansion of ML applications in finance research is undeniable; however, the precise areas and methods to effectively apply ML to address financial research problems remain largely ambiguous.

Methods

This paper makes a threefold contribution to the field. Firstly, it provides a comprehensive introduction to ML tailored for financial economists. We shed light on the various types of ML, their intended uses, functionalities, and the available techniques for each type. Given our focus on finance, we accentuate the distinctions between conventional econometric methods and ML. Furthermore, we demonstrate the advantages of ML over traditional linear methods, particularly in the context of prediction problems, by applying ML to a complex high-dimensional asset pricing challenge in finance. Our introductory section equips researchers in the finance domain with a quick grasp of essential ML concepts relevant to finance applications, requiring no prior knowledge of ML.

Secondly, we establish a taxonomy categorizing both current and prospective ML applications in finance. In light of the burgeoning volume of recent studies, previous classifications fall short in effectively encapsulating the breadth of existing applications. We meticulously survey the latest literature in the field and classify it into three distinct archetypes. This taxonomy serves multiple purposes: it enhances researchers' understanding of the current state of literature and how various contributions interrelate, and it provides valuable guidance for the future of ML applications in finance.

Moving forward, our investigation delves into the future outlook of ML applications in finance. We systematically analyze these applications, considering their varying degrees of success across distinct research fields, namely asset pricing, corporate finance, financial intermediation, and household finance, as well as different application types. Our findings not only suggest the substantial potential of ML applications as a whole but also provide researchers with valuable insights into the most promising avenues for future exploration.

In traditional econometrics, the primary objective is to offer causal explanations for economic phenomena by scrutinizing the relationships between economic variables. Conversely, ML empowers researchers to extract unique insights from high-dimensional data. There are two significant categories of high-dimensional data where ML outshines traditional methods like linear regression. First, ML excels at handling high-dimensional numerical data, which entails a surplus of variables relative to the number of observations. This scenario arises when a multitude of economically relevant variables exists or when nonlinearities and interaction effects significantly impact the analysis. ML methodologies harness the information embedded in such data to make predictions with minimal out-of-sample prediction errors. Second, unlike conventional methods, ML enables the exploration of unconventional data types, including text, images, or videos, which inherently possess high-dimensionality. ML techniques can extract economically pertinent information from such data, serving as a foundation for subsequent economic analyses.

ML shares a close connection with the concept of big data, which encompasses datasets featuring a large number of observations, a substantial number of variables, or both (Stock and Watson, 2020, p. 515). In general, datasets with a substantial number of observations enhance the precision of ML predictions, similar to how they improve the accuracy of parameter estimates in ordinary least squares (OLS) regressions.

In cases where datasets exhibit a high number of variables relative to observations, Machine Learning (ML) outperforms simpler, traditional methods such as linear regression. Applying ML to datasets with high numbers of observations and variables combines the advantages of yielding high prediction accuracy while surpassing the capabilities of traditional methods.

Drawing from our extensive review of the finance literature, we classify ML applications into three distinct categories:

1. Construction of Superior and Novel Measures: Researchers can employ ML to create advanced and innovative measures. For instance, when ML is applied to unconventional data, the extracted information can serve as a superior or novel measure of an economic variable. These improved ML measures often exhibit lower measurement error, enabling more precise estimates of economic relationships compared to traditional measures. Novel ML measures open up possibilities for analyzing previously unmeasurable economic variables.
2. Reduction of Prediction Error in Economic Predictions: ML can significantly reduce prediction error in various economic prediction problems. For instance, accurately pricing financial or real assets is a core challenge in finance, and ML's innate ability for prediction often outperforms traditional approaches in solving such economic prediction problems.
3. Extension of the Existing Econometric Toolkit: Econometric tools frequently involve prediction components. ML can enhance these existing tools by improving their predictive capabilities. Furthermore, some ML methods serve as entirely new econometric tools themselves. For instance, ML-based clustering methods expand the range of clustering techniques available in econometrics.

To demonstrate the superiority of ML over traditional methods in a typical prediction problem, we apply ML to real estate asset pricing, a context particularly relevant to household finance and real estate economics. Real estate asset pricing poses a high-dimensional challenge due to numerous property characteristics, nonlinearities, and interaction effects. We predict real estate asset prices in the German residential housing market using various ML methods that leverage the extensive individual property characteristics in our dataset. We compare these ML-based predictions to estimates obtained through traditional hedonic pricing (linear regression with the Ordinary Least Squares [OLS] estimator). Figure 1 illustrates our key findings, showing that, on average, ML-based price predictions closely align with actual prices, especially in the upper price range, where OLS estimates often deviate significantly from actual prices.

In the section of our paper, we conduct a bibliometric analysis, examining the publication success of articles in major finance journals from 2010 to 2021. We address several key questions:

1. Importance of ML in Finance Research: ML, although relatively new, has gained broad acceptance in the finance research community, with ML papers accounting for approximately 3%–4% of publications in top finance journals in 2021.
2. Methodological Purpose of ML: ML in finance research serves various methodological purposes beyond prediction, including constructing superior measures and novel variables.
3. Differences Across Subfields: Subfields in finance, such as financial markets/asset pricing and banking/corporate finance, leverage ML differently. Financial markets/asset pricing tends to apply ML to economic prediction problems, while banking and corporate finance often use ML to create superior and novel measures.

Notably, publications in the most prestigious journals exhibit a disproportionate use of ML for the purpose of constructing superior and innovative measures. This trend is particularly prominent in the realms of banking and corporate finance. Our findings underscore the significant potential of applying ML to unconventional data sources, paving the way for the development of superior and novel measures, especially concerning topics related to financial institutions and corporate finance.

In summary, our results paint a promising picture for the future of ML applications in finance. The numerous advantages that ML offers over traditional econometric methods, the consistent and robust growth in the number of ML publications in recent years, and the widespread adoption of ML by studies featured in the highest-ranked journals within the field leave little room for skepticism.

Our paper contributes to the expanding literature focused on ML applications in finance. Existing finance textbooks, for instance, either survey specific finance areas where ML techniques have gained prominence (e.g., Nagel, 2021, for asset pricing; De Prado, 2018, for asset management) or provide mathematical foundations for ML within quantitative finance (e.g., Dixon, Halperin, and Bilokon, 2020). These essential contributions aim to demonstrate how ML techniques can be carefully adapted to address the specific characteristics of certain subfields within finance, primarily focusing on financial markets. Our perspective on ML differs significantly, as our primary aim is to identify promising ML applications that extend beyond prediction problems, especially outside of financial markets.

Furthermore, we contribute to a smaller body of survey papers that review the applications of ML in finance. Distinct from these surveys, our approach doesn't rely on automated techniques like textual analysis or citation-based approaches. Instead, we manually review ML applications across various finance subfields, emphasizing applications beyond financial markets, with a focus on understanding their unique potential and contributions.

In this section, we lay the groundwork for subsequent chapters by providing a primer on Machine Learning (ML). Our primary objective is to delve into the mechanics of various ML types, delineate the problems for which ML excels, and introduce the methods commonly used in finance literature. Additionally, we emphasize the distinctions between ML and traditional econometric methods.

In empirical finance studies, the primary goal is to analyze economic relationships between different variables. A typical example involves investigating how specific factors influence capital structure or how regulatory changes impact the expectations of economic agents. Traditional econometric methods are typically employed to estimate βˆ, which provides insights into the direction and strength of these influencing factors.

Conversely, ML serves different purposes. Instead of directly elucidating the relationships between economic variables, ML primarily functions as a tool for prediction or data structure inference. Prediction methods utilize available observations to derive estimates for the dependent variable y of new, unseen observations based on their covariatesˆX. For instance, in the real estate market, observed property prices and their characteristics can be used to predict the prices of previously unobserved properties based on their attributes. The first major type of ML, known as supervised learning, encompasses techniques designed for making such predictions.

To illustrate the disparities between Machine Learning (ML) methods and traditional approaches, we apply ML to the task of predicting real estate prices. Real estate price prediction is an ideal example to highlight the advantages of ML in solving finance-related problems for several compelling reasons.

First, real estate stands as one of the most vital asset classes in the economy, with its total value in the United States being comparable to the combined size of equities and fixed income markets. For many households, real estate constitutes their primary source of wealth. The Global Financial Crisis of 2007/2008 exemplified how disruptions in the real estate sector can have far-reaching repercussions on economies worldwide. Hence, reducing prediction errors in real estate pricing carries significant economic importance.

Second, real estate assets exhibit a high level of heterogeneity, with each property being unique. This diversity complicates real estate pricing substantially.

Third, real estate pricing poses an inherently high-dimensional problem due to the numerous property characteristic variables and the potential presence of nonlinearities and interaction effects. In such cases, ML offers unique advantages over traditional methods.

The traditional approach for estimating prices of individual properties is known as hedonic pricing. Hedonic pricing initially regresses property characteristics against observed property prices using Ordinary Least Squares (OLS) to create a linear pricing model. This model is then used to predict prices for new, unobserved properties. However, hedonic pricing relies on an inherently linear model and does not explicitly account for nonlinearities and interaction effects. For instance, it may overlook important interactions between lot size and location. While specific effects can be manually added to the linear model, there might exist numerous unknown nonlinear and interaction effects that go unaccounted for. ML methods, in contrast, automatically consider these nonlinearities and interactions, potentially leading to more accurate price predictions.

In this study, we employ a comprehensive dataset comprising over four million residential real estate listings in Germany spanning from January 2000 to September 2020, sourced from major real estate online platforms and newspapers. This dataset encompasses offer prices and all relevant individual property characteristics such as floor area, number of rooms, construction year, location, lot size, etc. We utilize these data to train various ML models for predicting individual property prices, subsequently comparing these models with the linear OLS model derived from hedonic pricing.

The results, as shown in Panel A of Figure 4, are striking. ML methods substantially improve the accuracy of price predictions compared to the OLS baseline. Our top-performing ML model, boosted regression trees, elevates the out-of-sample R^2 to 77%, nearly doubling the explained price variation compared to OLS, which achieves 40%. On average, predictions from boosted regression trees deviate from actual prices by approximately 27%, while OLS exhibits a deviation of 44%. In monetary terms, this enhanced prediction performance corresponds to an average pricing error of about 94,000 EUR for ML, compared to 176,000 EUR for OLS. Given that the mean property price in our sample is 393,000 EUR, these improvements are not only statistically significant but also economically substantial.

Moreover, the advantages of ML become even more apparent at the upper end of the price range, as depicted in Panel B of Figure 4. Boosted regression trees outperform OLS across all price quintiles. In the highest price quintile, ML significantly reduces the average pricing error to 24%, compared to OLS's 50%. In terms of monetary units, this superior performance translates to an average pricing error reduction of over 240,000 EUR for boosted regression trees in the highest price quintile, where the average property price is approximately 884,000 EUR.

These findings underscore the relevance of nonlinearities and interaction effects in real estate pricing, especially for high-end properties. Our results demonstrate that ML can substantially reduce prediction errors in economic prediction problems compared to traditional linear regression with OLS. ML not only improves prediction accuracy in general but also excels, particularly for observations that pose challenges for traditional approaches.

Limitations and Considerations of Machine Learning

While our illustrative application of Machine Learning (ML) to real estate asset pricing highlights the advantages of ML over traditional methods for high-dimensional data problems, it's important to acknowledge the limitations, caveats, and drawbacks associated with ML. In the following, we delve into three crucial aspects in detail.

1. Low Interpretability: ML methods often exhibit low interpretability. While ML models can generate predictions with low prediction error, understanding how the algorithm arrived at these results is often not straightforward. Consequently, ML is generally less suited for problems that require a deep understanding of the economic determinants behind the prediction target. Nonetheless, the rapidly evolving field of interpretable ML is actively working on addressing the model interpretability challenge through various approaches.
2. Data Requirements: ML typically demands large datasets. Data size can be large in two dimensions: the number of relevant variables and the number of observations. ML provides advantages over traditional methods for prediction tasks when there is a large number of relevant variables compared to observations. However, ML generally delivers good prediction performance only if there's a substantial number of observations available for training an ML model. Unfortunately, large-scale data may not always be accessible for many research questions in finance. In some cases, researchers can leverage pre-trained ML models that have been trained on extensive, comparable data for common ML tasks like textual analysis or face recognition. This enables them to apply pre-trained models to specific problems, regardless of the data volume. Additionally, the ongoing trend of increasing data collection across various domains is expected to alleviate data scarcity concerns over time.
3. Computational Costs: Employing ML often comes with high computational costs. Compared to traditional methods such as linear regression, training ML models typically requires significantly more time and computing power. This challenge becomes more pronounced with more complex ML methods, especially neural networks with intricate architectures, which tend to have the highest computational requirements. Consequently, utilizing cloud computing services often becomes a necessity to address these computational demands effectively.

Construction of Superior and Novel Measures

The first archetype of Machine Learning (ML) applications in finance involves the creation of superior and innovative measures. Studies in this category leverage ML to extract information from high-dimensional and unconventional data sources, such as text, images, or videos, to formulate a numerical measure for an economic variable. Traditionally, approaches for handling textual data relied on word counts based on domain-specific dictionaries. For image and video data, human assessments were the primary source of information. ML-based methods offer a more efficient and potent means to access and interpret information within unconventional data sources. Various types of ML methods can be applied, including predictions from supervised learning, data structure information from unsupervised learning, and outcomes from other ML techniques to formulate measures for economic variables.
These superior or novel measures subsequently serve as independent variables in the primary analysis of economic relationships. Utilizing superior measures (those with lower measurement error compared to existing measures) helps mitigate attenuation bias, leading to more precise estimates of parameters describing economic relationships. Novel measures facilitate the exploration of previously unmeasurable economic aspects. In the primary analysis, many studies that construct ML-based measures employ traditional econometric methods, such as linear regression with Ordinary Least Squares (OLS).
Below, we categorize a selection of studies that utilize ML to construct superior or novel measures into three groups: (1) measures of sentiment, (2) measures of corporate executives' characteristics, and (3) measures of firm characteristics.
1. Measures of Sentiment: These studies employ ML to gauge sentiment in various contexts, such as market sentiment, news sentiment, or investor sentiment. ML techniques are applied to analyze textual data, providing more nuanced and data-driven sentiment measures than traditional methods.
2. Measures of Corporate Executives' Characteristics: This category focuses on constructing measures related to corporate executives' characteristics, including their personality traits, behavior, or communication style. ML methods are utilized to extract insights from textual or other unconventional data sources to formulate these measures.
3. Measures of Firm Characteristics: Here, ML is used to create measures related to firm-level characteristics, often derived from unconventional data sources like textual information. These measures provide a richer understanding of the financial health and performance of firms.

Table 2 in the original document provides a detailed selection of studies that fall under these categories, showcasing the diversity of applications in the construction of superior and novel measures using ML in finance research.

Measures of Sentiment

Measures of sentiment are designed to capture people's beliefs or opinions, typically on a positive-to-negative scale. In this subcategory, most studies focus on constructing sentiment measures from textual data. There are several approaches to create one-dimensional sentiment measures (e.g., positive vs. negative) from text data.

One common approach is the dictionary method, as demonstrated by Loughran and McDonald (2011). This method involves counting the occurrence of positive and negative words based on a domain-specific word list. However, dictionary-based approaches have limitations as they may not consider the context of words within a sentence.

In contrast, flexible Machine Learning (ML) approaches can account for both word context within a sentence and how different sentences relate to each other. They offer a more nuanced and data-driven approach to sentiment analysis. For a comprehensive review of sentiment analysis using both traditional econometric and ML-based methods, Algaba et al. (2020) provide an extensive resource.

Sentiment measures can pertain to various topics and originate from diverse sources. In finance, the primary focus is often on aggregate market sentiment, such as in the stock market. Many studies utilize ML-based sentiment measures for stocks to investigate their impact on future stock returns and various financial reporting figures.

There are several studies that construct investor sentiment measures from social media platforms. For example:

1. Antweiler and Frank (2004): They use ML methods like naïve Bayes and Support Vector Machines (SVM) to classify user posts on the Yahoo Finance message board as positive or negative. These classifications are then aggregated to create a measure of stock market sentiment.
2. Renault (2017): This study classifies user posts on the finance-focused social network StockTwits to construct an investor sentiment measure.
3. Vamossy (2021): This research relies on StockTwits and extracts different emotional states from user posts using deep learning-based textual analysis to create a measure of investor emotions.
4. Studies by Sprenger et al. (2014), Bartov, Faurel, and Mohanram (2018), Giannini, Irvine, and Shu (2018), and Gu and Kurov (2020): These studies derive investor sentiment from user posts on Twitter.
5. Liew and Wang (2016): They apply ML to extract sentiment information from Twitter, but their focus is on pre-IPO sentiment.

These studies demonstrate the growing interest in leveraging ML techniques to extract sentiment from social media and other textual sources, providing valuable insights into market sentiment and its potential impact on financial markets.

Measures of Corporate Executives' Characteristics

Corporate executives play a pivotal role in a firm's leadership, and their characteristics can have significant implications for various aspects of a company's operations. Within the finance literature, Machine Learning (ML) techniques have enabled the creation of advanced measures related to corporate executives' characteristics. While most measures in this category are derived from textual data, some studies also construct measures by analyzing images and videos.

Several studies focus on constructing ML-based measures of executives' personality traits:

1. Gow et al. (2016): This study employs ML to extract CEOs' Big Five personality scores (agreeableness, conscientiousness, extraversion, neuroticism, and openness to experience) from the question-and-answer portion of conference call transcripts. These extracted scores are then used to analyze the impact of personality traits on financing choices, investment decisions, and operating performance.
2. Hrazdil et al. (2020): The authors determine the Big Five personality scores of CEOs and CFOs using IBM Watson Personality Insights, a commercial service. From these scores, they create a novel measure of executives' risk tolerance and investigate its influence on audit fees.

Other studies focus on constructing measures related to executives' beliefs:

3. Du et al. (2019): This study applies ML to mutual fund managers' letters to shareholders to create a measure of managers' confidence in expressing their opinions. The primary analysis then explores the impact of confidence levels on future performance.

These studies highlight the use of ML to extract valuable insights from textual data related to corporate executives. By quantifying personality traits, beliefs, and other characteristics, researchers gain a deeper understanding of how executive attributes influence various financial and strategic decisions within organizations.

Reduction of Prediction Error in Economic Prediction Problems

The second archetype of ML applications in finance focuses on applying ML to minimize prediction error in economic prediction problems. While many economic problems involve identifying causal relationships between economic variables, others directly require accurate predictions. ML is particularly useful in the latter category, where it can outperform simpler approaches like linear regression with Ordinary Least Squares (OLS) in generating more precise predictions.

Predictions can be generated from various types of data, including numerical data and unconventional data sources such as text, images, or videos. Since the primary goal of ML in this archetype is to minimize prediction error in economic prediction tasks, supervised ML methods are primarily used. Researchers often employ a variety of ML techniques to determine which method works best for a given dataset. These supervised ML methods ultimately produce predictions for economic variables, contributing to the resolution of economic prediction problems.

Below are three categories of relevant studies within this archetype:

1. Prediction of Asset Prices and Trading Mechanisms: This category involves using ML to predict asset prices and trading mechanisms, which is crucial in financial markets. ML methods are employed to enhance the accuracy of price predictions and trading strategies.
2. Prediction of Credit Risk: ML techniques are applied to predict credit risk, a key concern in the financial industry. These studies aim to improve the assessment of borrowers' creditworthiness and the likelihood of loan defaults.
3. Prediction of Firm Outcomes and Financial Policy: In this category, ML is used to predict various outcomes related to firms and their financial policies. These studies aim to provide insights into the financial performance and decision-making processes of companies.

While the benefits of Machine Learning (ML) over traditional methods have been demonstrated, the relatively limited number of ML applications in finance suggests significant untapped potential for future research. However, several key questions remain unanswered: Will the usage of ML methods become widely popular within the finance community? Can ML applications find their way into the most prestigious finance journals, or are they more likely to be published in specialty journals? Furthermore, given the diverse application categories of ML and the wide array of research fields within finance, it's challenging to identify the most promising ML applications in finance research.

In this section, the existing finance literature that utilizes ML methods is systematically analyzed to provide indicative answers to these questions. The focus is on investigating the publication success of papers using ML and how it varies across different research fields and application types within finance. These findings not only offer insights into the future prospects of ML in finance but also provide guidance on where and how researchers can apply ML to maximize its potential impact.

This analysis aims to shed light on the trajectory of ML applications in finance, from their current state to their potential growth and acceptance within the broader finance research community.

Conclusion

In this study, we have delved into the utilization of ML technology within the realm of finance research. Initially, we highlighted the distinctive problem-solving capabilities of various ML techniques when compared to traditional linear regression with OLS. While OLS is prized for its explanatory abilities, supervised ML emerges as the superior choice for prediction tasks, as illustrated through our real estate asset pricing prediction example where ML-based predictions significantly outperformed OLS in terms of pricing accuracy.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 14, No. 1, 2023, p.142-149
https://publishoa.com
ISSN: 1309-3452
The subsequent section of this paper saw the development of a taxonomy for ML applications in finance, categorized as follows: 1) the construction of advanced and novel measures, 2) the reduction of prediction errors in economic prediction problems, and 3) the extension of the existing econometric toolkit. This taxonomy serves a multifaceted purpose, facilitating a systematic review of existing ML literature in finance, enhancing comprehension of novel contributions and their integration into the existing body of work, and ultimately guiding researchers in identifying potential applications, thereby fostering new ML studies in finance.
In the final part, we provided insights into the future prospects of ML applications in finance by scrutinizing ML papers published in major finance journals. Notably, we observed a robust growth in the number of ML applications in finance over recent years, with many of these applications finding their place in the most prestigious journals of the profession. Our findings suggest a likelihood of further proliferation of ML in finance research in the forthcoming years. Additionally, there appears to be substantial untapped potential for ML in constructing advanced and novel measures, particularly in the realms of corporate finance and governance. The fields of behavioral and household finance also hold promising avenues for future ML research endeavors.

References

[1] Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.
[2] Brogaard, J., Hendershott, T., & Riordan, R. (2014). High-frequency trading and price discovery. Review of Financial Studies, 27, 2267–2306.
[3] Brown, G. W., & Cliff, M. T. (2005). Investor sentiment and the near-term stock market. Journal of Empirical Finance, 12, 49–78.
[4] Brown, J. R., Farrell, K. A., & Weisbenner, S. J. (2015). Decision-making approaches and the propensity to default: Evidence and implications. The Journal of Finance, 70, 691–724.
[5] Brown, J. R., & Ivković, Z. (2021). Financial market pricing of external political risks: Evidence from the Brexit vote. Journal of Financial Economics, 139, 1–29.
[6] Bubb, R., & Pavan, A. (2017). Environmental externalities and optimal taxation. Econometrica, 85, 569–600.
[7] Buchner, A., Heinemann, F., & Krahnen, J. P. (2018). The relevance of personality traits for
[8] economic preferences. Journal of Economic Psychology, 69, 40–55.
[9] Bühlmann, P., & van de Geer, S. (2011). Statistics for high-dimensional data: Methods, theory and applications (Vol. 32). Springer Science & Business Media.
[10] Bühlmann, P., Kalisch, M., Meier, L., & Meinshausen, N. (2014). High-dimensional statistics with a view toward applications in biology. Annual Review of Statistics and Its Application, 1, 255–278.
[11] Bumann, S., & Weigert, F. (2020). Do buy-side analyst reports reveal contrarian sentiment?
[12] European Financial Management, 26, 462–496.
[13] Cai, J., Cai, N., & Keasey, K. (2019). Overinvestment, CEO entrenchment, and voluntary disclosure. The Review of Financial Studies, 32, 2837–2870.
[14] Cai, J., & DeAngelo, L. E. (2017). Board heterogeneity and managerial entrenchment: Evidence from corporate payout policies. Journal of Financial Economics, 125, 497–519.
[15] Cai, J., & Kim, W. S. (2011). CEO option compensation and systemically important banks.
[16] Journal of Financial Economics, 99, 116–138.
[17] Cai, J., & Walkling, R. A. (2011). Shareholders' say on pay: Does it create value? Journal of Financial Economics, 103, 61–81.
[18] Campi, L., Garlappi, L., & He, H. (2018). Director networks and takeovers. Management Science, 66, 477–493.
[19] Campbell, J. L. (2003). Understanding the formulas of inclusion: How financial analysts determine recommended capital structures. The Journal of Business, 76, 465–491.