



*Edited by*  
Thomas Walker · Frederick Davis  
Tyler Schwartz

---

# **Big Data in Finance**

## Opportunities and Challenges of Financial Digitalization

---

palgrave  
macmillan

# Big Data in Finance

Thomas Walker · Frederick Davis ·  
Tyler Schwartz  
Editors

# Big Data in Finance

Opportunities and Challenges of Financial  
Digitalization

palgrave  
macmillan

*Editors*

Thomas Walker  
Department of Finance  
Concordia University  
Montreal, QC, Canada

Frederick Davis  
Department of Finance  
Concordia University  
Montreal, QC, Canada

Tyler Schwartz  
Department of Finance  
Concordia University  
Montreal, QC, Canada

ISBN 978-3-031-12239-2      ISBN 978-3-031-12240-8 (eBook)  
<https://doi.org/10.1007/978-3-031-12240-8>

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer  
Nature Switzerland AG 2022

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Palgrave Macmillan imprint is published by the registered company Springer Nature Switzerland AG

The registered company address is: Gewerbstrasse 11, 6330 Cham, Switzerland

## PREFACE

Our global society is becoming increasingly data centric. The use of large, detailed datasets has revolutionized many fields, including—among others—medicine, biology, manufacturing, sports, marketing, and finance. With advances in how data can be collected and stored, a new phenomenon has emerged: big data. In simple terms, big data can be understood as a large amount of data that can be analyzed to understand past patterns better or predict future outcomes. This book examines the technical aspects of recent innovations surrounding big data in finance, as well as the benefits and risks associated with these developments. Moreover, the book sheds light on the ethical and privacy issues associated with big data, as well as the environmental footprint of collecting, storing, and analyzing big datasets.

The book features contributions from the international community of scholars and practitioners who work at the interface of artificial intelligence, big data, and finance. The authors review and critically analyze new developments at the intersection of big data and finance, and provide different perspectives on their impact on the financial sector and the way it operates. The book serves as a technical guide of these developments, exploring the theory and mechanisms behind the algorithms using big data, and exploring their use in a finance context. The contributors explain and demonstrate the predictive capabilities of big data in finance using different model types such as supervised, unsupervised, and semi-supervised learning. Moreover, because big data in finance has many

applications that extend beyond financial institutions, the book features contributions that explore possible policy and sustainability-oriented solutions and implications of the use of big data in finance.

Montreal, Canada  
Montreal, Canada  
Montreal, Canada

Thomas Walker  
Frederick Davis  
Tyler Schwartz

## ACKNOWLEDGMENTS

We acknowledge the financial support provided through the Jacques Ménard—BMO Centre for Capital Markets at Concordia University. In addition, we appreciate the excellent copy-editing and editorial assistance we received from Gabrielle Machnik-Kekesi, Victoria Kelly, Charlotte Frank, and Maya Michaeli.

# CONTENTS

## **Introduction**

<b>Big Data in Finance: An Overview</b>	<b>3</b>
Thomas Walker, Frederick Davis, and Tyler Schwartz	

## **Big Data in the Financial Markets**

<b>Alternative Data</b>	<b>13</b>
Vincent Grégoire and Noah Jepson	

<b>An Algorithmic Trading Strategy to Balance Profitability and Risk</b>	<b>35</b>
Guillermo Peña	

<b>High-Frequency Trading and Market Efficiency in the Moroccan Stock Market</b>	<b>55</b>
El Mehdi Ferrouhi and Ibrahim Bouabdallaoui	

<b>Ensemble Models Using Symbolic Regression and Genetic Programming for Uncertainty Estimation in ESG and Alternative Investments</b>	<b>69</b>
Percy Venegas, Isabel Britez, and Fernand Gobet	



**Big Data in Financial Services**

**Consumer Credit Assessments in the Age of Big Data** 95

Lynnette Purda and Cecilia Ying

**Robo-Advisors: A Big Data Challenge** 115

Federico Severino and Sébastien Thierry

**Bitcoin: Future or Fad?** 133

Daniel Tut

**Culture, Digital Assets, and the Economy:  
A Trans-National Perspective** 159

John Fan Zhang, Zehuang Xu, Yi Peng, Wujin Yang,  
and Haorou Zhao

**Case Studies and Applications**

**Islamic Finance in Canada Powered by Big Data: A Case  
Study** 187

Imran Abdool and Mustafa Abdool

**Assessing the Carbon Footprint of Cryptoassets: Evidence  
from a Bivariate VAR Model** 207

Hany Fahmy

**A Data-Informed Approach to Financial Literacy  
Enhancement Using Cognitive and Behavioral Analytics** 231

Prasanta Bhattacharya, Kum Seong Wan, Boon Kiat Quek,  
Waseem Bak'r Hameed, and Sivanithy Rathananthan

**Index** 265

## NOTES ON CONTRIBUTORS

**Abdool Imran** is the President of consultancy Blue Krystal Technologies and Business Insights. Currently, he lectures at the University of Western Ontario Richard Ivey School of Business. During the 2007/2007 Financial Crisis, Imran also served in the Assistant Deputy Minister's Office for Financial Sector Policy of the Government of Canada's Department of Finance. Imran's commentaries on the economy and finance has appeared in Canada's national media such as the Globe and Mail, the CBC, and the Toronto Star.

**Abdool Mustafa** is a Machine Learning Engineer at Airbnb, one of the largest travel accommodations platforms in the world. His work involves designing and implementing novel recommender and search systems for Airbnb products. Academic papers authored by Mustafa and his team have appeared in the SIGKDD Conference on Knowledge Discovery and Data Mining (KDD). Mustafa holds a graduate degree in Computer Science, with a specialization in Artificial Intelligence from Stanford University.

**Bhattacharya Prasanta** is a research scientist and innovation lead with the Social and Cognitive Computing department at the A\*STAR Institute of High Performance Computing. He also serves as adjunct Assistant Professor at the National University of Singapore (NUS) Business School. Prasanta holds a Ph.D. in Information Systems from the Department of Information Systems and Analytics, NUS, where he studied network science with a special focus on predictive and inferential methods in large

social networks. His current research aims at understanding the role of big data in emerging social and business applications from finance and education. Prasanta actively collaborates with major industry partners from around the world, and has presented his research in leading computer science, information systems, and marketing science venues.

**Bouabdallaoui Ibrahim** is a Ph.D. candidate in Big Data and Artificial Intelligence at the Mohammadia School of Engineering, Mohammed V University in Rabat, Morocco. He holds an MSc in Big Data Analytics from Sidi Mohamed Ben Abdellah University in Fez, Morocco. His research interests include applications of artificial intelligence in tourism, finance, medicine and insurance.

**Britez Isabel** holds a degree in International Relations from the Federal University of Santa Catarina (UFSC), with graduate degrees in Neurosciences and Behaviour from the Pontifical Catholic University of Rio Grande do Sul (PUCRS) (Brazil). She currently pursues a Master of Studies in Social Innovation at the Judge Business School, University of Cambridge (class of 2024). She is Editorial Assistant for the Journal on Policy and Complex Systems, and Founder of ESG Means, a boutique research firm specialized in ESG investment.

**Davis Frederick** has been a professor at the John Molson School of Business since 2011, having received his Ph.D. shortly prior from Queen's University. Prior to his academic career, he worked for several years in the government sector focusing on economic development for communities. His research interests include mergers and acquisitions, insider trading, and other aspects of corporate finance. He has published in the Journal of Corporate Finance, Journal of Business Finance & Accounting, European Financial Management, the International Review of Financial Analysis, and other high-quality journals.

**Fahmy Hany** is Associate Professor and the Finance Intellectual Lead of the School of Business at Royal Roads University in Victoria, British Columbia. His research interests include energy economics, energy finance, financial econometrics, climate finance and economics, and sustainability. Fahmy's work has been published in top tier academic journals such as Energy Economics, Applied Economics, and Economic Modeling. His research work has been accepted and presented at top domestic and international economic and finance conferences. Fahmy's work has also been featured in the local media and other news outlets.

He is an active member of the Canadian Economic Association (CEA) and the Canadian Sustainable Finance Network (CSFN). In addition to his academic appointments, Fahmy has over ten years of experience in financial and economic consulting. He has led and authored a number of market studies, feasibility studies, cost-benefit analyses, and project appraisals for government and business.

**Ferrouhi El Mehdi** is Associate Professor of Finance at the Faculty of Economics and Management, Ibn Tofail University, Morocco. He holds a Ph.D. from the Mohammed V University in Rabat, Morocco. His research interests include behavioral finance, banking, risk management, financial intermediation, and financial inclusion. He has published papers in many prestigious journals including the *Journal of African Business* and *Physica A*.

**Gobet Fernand** is Professorial Research Fellow in the Centre for Philosophy of Natural & Social Science at the London School of Economics. His research interests include the psychology of expertise and talent, the acquisition of language, scientific discovery, and computational modeling. He has authored and co-authored over 400 publications, including eleven books. His latest books are *The Psychology of Chess* (2018) and *Scientific Discovery in the Social Sciences* (2019).

**Grégoire Vincent** is a tenured Associate Professor of Finance at HEC Montréal, where he holds the Research Professorship in Financial Big Data Analytics. He holds a Ph.D. in Finance from the University of British Columbia and degrees in Computer Engineering and Financial Engineering from Université Laval in Quebec. His academic research interests are in information economics, market microstructure, and big data and machine learning applications in finance. His research has been published in leading academic journals such as the *Journal of Financial Economics*, the *Journal of Accounting Research*, and the *Journal of Financial and Quantitative Analysis*.

**Hameed Waseem Bak'r** majored in Psychology and holds a Bachelor of Social Sciences (Honours) degree from the National University of Singapore. His varied background includes working in educational psychology and conducting neuroscience research with the Agency for Science, Technology, and Research (A\*STAR). He has been Research Manager with Institute for Financial Literacy (IFL) since 2015.

**Jepson Noah** holds a B.B.A. in Finance and Economics from Bishop's University in Sherbrooke, Quebec where he was valedictorian of his graduating class. He is pursuing an MSc in Finance in Montreal. Noah's research interests are focused in the area of alternative data.

**Kiat Quek Boon** is a senior scientist and also director of the Social and Cognitive Computing Department at the A\*STAR Institute of High Performance Computing. His research interests include computational modeling and simulation of psychological and cognitive processes, social network modeling and analyses, and building network-based modeling tools for knowledge representation and reasoning.

**Peña Guillermo** achieved recognition as the second highest graduating student of Economics for 2009–2013 at the University of Zaragoza (Unizar: Zaragoza, Spain), where he was also awarded the Extraordinary Prize for the highest grade in the Master of Economics program. He was a research visitor of the University of Oslo and obtained a Cum Laude International Ph.D. in Economics in 2018. Peña was employed in the Ibercaja Bank Headquarters for two years, and is currently Assistant Professor at the Department of Economic Analysis at the University of Zaragoza and researcher of the Institute for Employment, Digital Economy, and Sustainability (IEDIS). He has published more than 25 papers in prestigious journals such as the National Tax Journal and Applied Economics. His research fields include, but are not limited to, urban economics, public economics, and finance.

**Purda Lynnette** is Professor and Associate Dean (Graduate Programs) and RBC Fellow of Finance at the Smith School of Business, Queen's University, Kingston, Ontario. She is a Chartered Financial Analyst (CFA) with previous investment banking experience. Lynnette conducts empirical research that is frequently interdisciplinary in nature, with journal publications spanning the areas of accounting, finance, international business, and law. Outside of academia, Lynnette has presented her work to policy makers and practitioners including the Bank of Canada, the Accounting Standards Oversight Council of Canada, and Financial Planning Canada. Lynnette has been a visiting researcher at the Bank of Canada and INSEAD Business School (France) and organized workshops at the Halbert Centre for Canadian Studies at the Hebrew University of Jerusalem. Lynnette is a past president of the Northern Finance Association and an award-winning teacher.

**Rathananthan Sivanithy** has twenty-five years of experience as a financial journalist with The Business Times. For the majority of those years he was the newspaper's dedicated stock market reporter, writing daily and weekly market reports. He has written extensively about movements in the stock market and various financial instruments such as bonds, real estate investment trusts, and company-issued and structured warrants. He has served as a judge for the Securities Investors Association of Singapore's Corporate Governance awards. He holds a Bachelor of Accountancy degree from the National University of Singapore, a Master of Science (Management Science) from the University of Manchester, and a Master in Business Administration from the University of New South Wales.

**Schwartz Tyler** holds an MSc degree in Data Science and Business Analytics from HEC Montreal. He currently serves as a research assistant in the Department of Finance at Concordia University and is the co-author of an edited book collection on climate change adaptation. Tyler completed his undergraduate degree at the John Molson School of Business, for which he received an honors in finance, as well as a Concordia University Student Research Assistant (CUSRA) scholarship. His research interests include predictive modeling, bio-statistics, sports analytics, FinTech, and machine learning.

**Seong Wan Kum** is a Research Engineer for the Social and Cognitive Computing department at the A\*STAR Institute of High Performance Computing. His research interests include personality and individual differences, behaviour and motivation, society and culture, as well as research methods & design.

**Severino Federico** is Assistant Professor in the Department of Finance, Insurance, and Real Estate at Université Laval, and a researcher at CIRANO. After obtaining his Bachelor and Master degrees in Applied Mathematics (with a focus on mathematical finance), he earned a Ph.D. in Economics and Finance from Università Bocconi (Milano) in 2018. He has also been a Postdoctoral Fellow at Università della Svizzera Italiana (Lugano). His research interests include asset pricing, financial economics, and financial econometrics. The main goal of his research is the study and classification of economic and financial shocks according to the duration of their effects on the markets. His research has been presented at several international conferences. Federico's teaching

subjects include machine learning for business, portfolio management, corporate finance, asset pricing, and mathematics for economics.

**Thierry Sébastien** is a senior analyst at a major Canadian bank. He holds two Master's degrees: an M.B.A. in Finance from Université Laval and a Master of Science in Management from Grenoble École de Management. His expertise includes investments, econometrics, and corporate finance.

**Tut Daniel** holds a B.A. (Hons) in Economics from the University of Toronto and an M.A. in Economics and a Ph.D. in Finance from the Schulich School of Business at York University. Dr. Tut joined the Ted Rogers School of Business at Ryerson University in 2020. His research interests are in corporate finance, creditor rights, FinTech and cryptocurrencies, debt financing, and corporate governance. His work has been presented at some of the top finance conferences and he won the "Top Paper Award" at the 2021 Global Finance Conference. His work has been published in the *Journal of Financial Research*.

**Venegas Percy** is Chief Scientist at Economy Monitor, Editor-in-Chief at the *Journal on Policy and Complex Systems* (Policy Studies Organization, Washington, DC), and co-leader of Autonomous Hypotheses Generation for Investment Decision Making at the London School of Economics. He holds degrees in engineering, business, sustainability and a certification in Artificial Intelligence from Oxford University.

**Walker Thomas** is a full professor of finance, the director and academic lead of the Emerging Risks Information Center (ERIC), the inaugural director for the Jacques Ménard/BMO Center for Capital Markets, and the Concordia University Research Chair in Emerging Risk Management (Tier 1) at Concordia University in Montreal, Canada. He previously served as an associate dean, department chair, and director of Concordia's David O'Brien Centre for Sustainable Enterprise. Prior to his academic career, he worked for firms such as Mercedes Benz, KPMG, and Utility Consultants International. He has published over 70 journal articles and books.

**Xu Zehuang** is a graduate student of Business Analytics in the School of Business at Macau University of Science and Technology, Macau, where she also obtained her Bachelor of Applied Economics. His current research focuses on the application of machine learning methods in economics and finance.

**Yang Wujin** is a postgraduate student in Economics at the Hong Kong University of Science and Technology. Her undergraduate research in Applied Economics at Macau University of Science and Technology has given her a deep appreciation for advanced macroeconomics, advanced microeconomics, game theory, ESG investment, monetary policy, and information economics.

**Yi Peng** is a postgraduate student in the Department of Political Economy of King's College, London. She graduated from the Macau University of Science and Technology with a major in Applied Economics. During her undergraduate studies, her research involved combining economic theory with quantitative decision analysis methods to analyze the impact of the development of financial technology on the economies of different countries. During her postgraduate study, her research direction involves exploring the impact of fluctuations in the exchange rate of the RMB against the US dollar before and after the Sino-US trade war.

**Ying Cecilia** is a Ph.D. candidate in Analytics at the Smith School of Business at Queen's University. Her research explores the application of large language models in natural language processing tasks and their potential societal impacts. In addition, she is interested in machine learning and AI model explainability, biases, and corrections. Her work has been presented at leading computer science and management science conferences, including Neural Information Processing Systems, the Association for Computational Linguistics, and INFORMS. Cecilia is the recipient of the SSHRC Joseph-Armand Bombardier Canada Graduate Scholarship and the Smith School of Business New Ph.D. Student Research Excellence Award. Previously, Cecilia worked as a director of Credit Risk Analytics in Global Risk Management at Scotiabank.

**Zhang John Fan** currently serves as an Assistant Professor at Macau University of Science and Technology. John gained his Ph.D. in finance at Auckland University of Technology in 2018. His research interest lies in the area of multinational finance, the cross-listing of emerging market firms, organizational diversity, and the impact of culture on financial and economic aspects. He has published academic papers in the *Review of Corporate Finance*, *International Research in Business and Finance*, and *China & World Economy*. John is a Chartered Financial Analyst (CFA) and Microsoft Certified Solutions Expert (MCSE). John is also a member



of the Christian Finance Faculty Association and treasurer of his local church.

**Zhao Haorou** is a Master of International Economics and Finance candidate at Newcastle University in the UK. She graduated from the Macau University of Science and Technology with a Bachelor of Applied Economics. In 2019, she participated in the Business Immersion program of The University of Hong Kong and the Global Business Case Competition. Her current research interests are in the area of macroeconomics, monetary policy, and forecasting.

# LIST OF FIGURES

## **An Algorithmic Trading Strategy to Balance Profitability and Risk**

Fig. 1	Correlations between the three main indicators: indicator of unitary profit (IUP), mobile-ratio (MR), and balanced investment indicator (BII)	43
Fig. 2	Correlations between economic activity and the balanced investment indicator (BII)	44
Fig. 3	Relationship between the mobile-ratio (MR) and balanced investment indicator (BII) (left) and between the deviation from the MR average and BII (right)	46
Fig. 4	Relationship between the mobile-ratio (MR) and balanced investment indicator (BII) for the sample of worldwide stock exchanges	50

## **Ensemble Models Using Symbolic Regression and Genetic Programming for Uncertainty Estimation in ESG and Alternative Investments**

Fig. 1	'ESG' (environmental, social, and corporate governance) searches by month	72
Fig. 2	Accuracy-complexity trade-off across modeling rounds	82
Fig. 3	Ensemble prediction plots, modelling rounds comparison for publicly traded private equity stocks	82
Fig. 4	Ensemble prediction and accuracy-complexity trade-off plots for sustainable exchange-traded funds (ETFs)	83

Fig. 5	Error residuals for the ensemble one model in the first round of modeling for publicly traded private equity stocks	84
Fig. 6	Error residuals for the ensemble two model in the second round of modeling publicly traded private equity stocks	85
Fig. 7	Error residuals for the variable ‘EBITDA’ (earnings before interest, taxes, depreciation, and amortization) in the ensemble for the second round of modeling for publicly traded private equity stocks	86

**Robo-Advisors: A Big Data Challenge**

Fig. 1	The development of robo-advisors	118
--------	----------------------------------	-----

**Bitcoin: Future or Fad?**

Fig. 1	Price, volume, and market capitalization of Bitcoin	136
Fig. 2	Price volatility of Bitcoin	137
Fig. 3	Tether coin versus U.S. dollar (USD) index	139
Fig. 4	Value and volume of Stablecoins	140
Fig. 5	Comparing the price of Bitcoin, the volatility index (VIX), and the price of gold	141
Fig. 6	Blockchain technology mining, blocksize, duration, and transactions time series	142
Fig. 7	<b>a</b> Bitcoin versus Standard and Poor’s 500 (S&P500) returns. <b>b</b> Kernel density of daily Bitcoin prices and returns	145
Fig. 8	Trades per minute on cryptocurrency exchanges	146
Fig. 9	The liquidity of Bitcoin	147
Fig. 10	Comparing the returns for Bitcoin and gold	148

**Islamic Finance in Canada Powered by Big Data: A Case Study**

Fig. 1	A mathematical depiction of a single neuron as used in deep neural networks	194
Fig. 2	Deep neural networks stacking architecture	194
Fig. 3	Recurrent neural network (RNN) architecture using long short-term memory neural networks (LSTMs)	198
Fig. 4	Example of how a dense embedding can be used in conjunction with a recurrent neural network (RNN)	201
Fig. 5	Two-stage processing for associating words and sentences with an event signal	203

## Assessing the Carbon Footprint of Cryptoassets: Evidence from a Bivariate VAR Model

Fig. 1	Monthly Cambridge Bitcoin Electricity Consumption Index (BCECI) and the monthly average trading volume (ATV) of Bitcoin, Ethereum, Ripple, Stellar, and Litecoin	214
Fig. 2	Orthogonalized response of Bitcoin's electricity consumption to a one standard deviation impulse (shock) of Bitcoin's trading volume	221
Fig. 3	Orthogonalized response of Bitcoin's electricity consumption to one standard deviation impulse (shock) of the monthly average trading volume (ATV) of Ethereum, Ripple, Stellar, and Litecoin	222

## A Data-Informed Approach to Financial Literacy Enhancement Using Cognitive and Behavioral Analytics

Fig. 1	Demographic distribution of participants	239
Fig. 2	Distribution of financial knowledge and financial confidence scores pre- and post-workshop	243
Fig. 3	Distribution of financial knowledge and financial confidence difference (increase/decrease) scores post- and pre-workshop	244
Fig. 4	Distribution of post-workshop financial intention	245
Fig. 5	Mean pre- and post-workshop financial confidence scores and mean number of interest topics indicated, across age groups	246
Fig. 6	Mean financial confidence scores pre- and post-workshop across topics	247
Fig. 7	3-cluster solution visualization	252
Fig. 8	User profiling and intervention analysis framework for financial literacy enhancement	257

# LIST OF TABLES

## Alternative Data

Table 1	Main characteristics of alternative data	15
---------	--	----

## An Algorithmic Trading Strategy to Balance Profitability and Risk

Table 1	Average value of the indicators for the three samples	44
Table 2	Countries with the highest balanced investment indicator (BII) in the sample	44
Table 3	Countries with the lowest mobile-ratio (MR) in the sample	45
Table 4	Countries with the highest mobile-ratio (MR) in the sample	45
Table 5	Stocks with the highest balanced investment indicator (BII) on the IBEX-35 exchange	46
Table 6	Investing-selling decisions according to Peña's (2020) proposal	47
Table 7	Investing-selling decisions according to current practices	48
Table 8	Investing-selling decisions according to the current proposal	48
Table 9	Summary of investing-selling decisions using the three different methods	49
Table 10	Stock exchanges with the lowest mobile-ratio (MR) in the sample	50

## High-Frequency Trading and Market Efficiency in the Moroccan Stock Market

Table 1	Descriptive statistics of Moroccan All Shares Index (MASI) price tick changes by different time-frequencies	62
Table 2	Variance ratio test results	63

## Ensemble Models Using Symbolic Regression and Genetic Programming for Uncertainty Estimation in ESG and Alternative Investments

Table 1	Data dictionary	88
---------	-----------------	----

## Robo-Advisors: A Big Data Challenge

Table 1	Main advantages and disadvantages of robo-advisors relative to human advisors	123
---------	---	-----

## Culture, Digital Assets, and the Economy: A Trans-National Perspective

Table 1	Summary statistics	168
Table 2	Use of digital assets at financial institutions	170
Table 3	National culture and the use of digital assets	171
Table 4	Use of digital assets at financial institutions while controlling for cultural collectivism (Panel A) and trust (Panel B)	174

## Assessing the Carbon Footprint of Cryptoassets: Evidence from a Bivariate VAR Model

Table 1	Summary statistics	213
Table 2	Cross-correlations between the Cambridge Bitcoin Electricity Consumption Index (CBECI) and the average trading volumes of the top five cryptocurrencies	215
Table 3	Goodness of fit of the bivariate vector autoregression (VAR) models	219
Table 4	Results of Granger Causality tests	220
Table 5	Global hash rate by country	223
Table 6	Total electricity net generation from all sources and sectors in the United States	224

Table 7	Average proportion of total electricity net generation from fossil fuels, nuclear power, and renewable energy between 2016 and 2020 in the United States	225
Table 8	One year prediction of Bitcoin’s electricity consumption	227

### **A Data-Informed Approach to Financial Literacy Enhancement Using Cognitive and Behavioral Analytics**

Table 1	Summary of participation in the Institute for Financial Literacy (IFL) workshops and talks	237
Table 2	Correlation matrix for workshops (by pairwise deletion)	241
Table 3	Post-hoc pairwise comparison of financial confidence scores post- and pre-workshop by module topics	248
Table 4	Regressing financial intention on key factors after controlling for module evaluation, demographics, and topics	250
Table 5	Profile summary of 3-cluster solution	253
Table 6	Designing behavior change interventions	258

# Introduction





# Big Data in Finance: An Overview

*Thomas Walker, Frederick Davis, and Tyler Schwartz*

## 1 INTRODUCTION

In the financial sector, the big data movement refers to the analysis of vast amounts of data with the goal of making better informed investment decisions, improving corporate operations, and enhancing decision-making processes on both the buy and supply sides of transactions (Hasan et al., 2020). Big data analysis frequently draws on artificial intelligence (AI) models and has created a paradigm shift in the operations of financial institutions (Nobanee, 2021; Sun et al., 2019). Big data and its impact on the financial industry now benefit from scholarly interest; many publications, reports, books, and conventions focus on dissecting and comparing different applications and techniques within the domain,

---

T. Walker (✉) · F. Davis · T. Schwartz  
Department of Finance, Concordia University, Montreal, QC, Canada  
e-mail: [twalker@jmsb.concordia.ca](mailto:twalker@jmsb.concordia.ca)

F. Davis  
e-mail: [frederick.davis@concordia.ca](mailto:frederick.davis@concordia.ca)

T. Schwartz  
e-mail: [tyler.schwartz@mail.concordia.ca](mailto:tyler.schwartz@mail.concordia.ca)

© The Author(s), under exclusive license to Springer Nature  
Switzerland AG 2022

T. Walker et al. (eds.), *Big Data in Finance*,  
[https://doi.org/10.1007/978-3-031-12240-8\\_1](https://doi.org/10.1007/978-3-031-12240-8_1)

with researchers envisioning what the movement means for the future of finance (Nobanee, 2021). The two main subtopics of interest are big data itself as well as the AI techniques used to analyze that data.

AI is best understood as computer programming that automates tasks that have traditionally relied on human intelligence (Russell, 2010). A popular branch of AI is machine learning, which uses statistical techniques to allow machines to “learn” through experience by correcting previous errors (Alzubi et al., 2018). Spacecraft engineering, pattern recognition, entertainment, biology, finance, and medicine, among others, all rely on machine learning (El Naqa & Murphy, 2015).

More recently, deep learning, a sub-domain of machine learning, has proven to be effective in extracting patterns from big data (Chen & Lin, 2014). Deep learning uses neural networks, which are inspired by the neurological basis of brain functioning, to extract patterns from big data (LeCun et al., 2015). Primary applications of deep learning include image and voice recognition, as well as natural language processing (NLP) tasks such as text summarization and classification (LeCun et al., 2015; Najafabadi et al., 2015).

In the financial industry, AI has been successfully employed by financial institutions in such areas as automated lending, portfolio construction and management (robo-advising), risk management, fraud detection, quantitative and high-frequency trading, as well as customer support (Bahrammirzaee, 2010; Buchanan, 2019; Buchanan & Wright, 2021). As the complexity of algorithms increases with new developments in deep learning and other technological advancements, the role of AI in finance is certain to become more important as well, making the field a rapidly changing frontier in the financial technology (FinTech)/banking nexus and making space not only for opportunities, but also for substantial ethical, social, legal, and economic risks (Buchanan, 2019).

Big data, as a discipline, is what allows AI to learn and develop complex pattern-detecting algorithms used for a variety of purposes (O’Leary, 2013). The troves of information collected in finance, for example, include customer information on banking, historical prices of stocks for investing, social media activity to predict the current sentiment of the market, and past fraudulent transactions for fraud detection (Buchanan, 2019). The data can be structured (organized, classified data), unstructured (text, social media activity), or semi-structured (incorporating both structured and unstructured elements) (Hurwitz et al., 2013). However, most of today’s data is unstructured and/or semi-structured, leading to

greater difficulties in finding a way to use the data efficiently (Yaqoob et al., 2016).

Gandomi and Haider (2015) review the existing literature and define big data using the traditional three Vs (volume, variety, and velocity) along with three more recent Vs developed by International Business Machines (IBM) (veracity), Statistical Analysis System (SAS) (variability), and Oracle (value):

- **Volume:** the size of the data,
- **Variety:** the diversity of the data,
- **Velocity:** the speed at which the data is generated and processed,
- **Veracity:** the reliability of the data (for example, social media data),
- **Variability:** the variability in the velocity of data,
- **Value:** the quality and usefulness of the data.

Currently, many questions dominate big data scholarship and professional practice, including how data is collected (consumer privacy and ethical concerns), stored (environmental impacts), secured, as well as analyzed and used (Jain et al., 2016; Lucivero, 2020; Martin, 2015). This edited book explores new developments surrounding big data in finance, and aims to provide meaningful answers to the questions posed above. While a general use of big data has often been the subject of discussion, this book will take a more focused look at big data applications in the financial sector to differentiate itself from current offerings and close a large gap in the literature.

## 2 OVERVIEW OF CONTENT

This edited book critically analyzes new developments at the intersection of big data and finance and provides different perspectives on their impact on the financial sector and the way it operates. As illustrated above, the predictive capability of big data is still developing as increasingly complex algorithms are devised by institutions to extract the most out of available data. The chapters in this book explore the predictive ability of big data in the context of finance, using different model types such as supervised learning and unsupervised learning. In addition, several of the chapters discuss the leveraging of recent advancements made in deep learning in combination with big data to innovate the financial sector. Lastly, because

the use of big data in finance has many implications that go beyond their use in financial institutions, several chapters address the ethical, privacy, and environmental implications of these applications.

## 2.1 *Part I: Big Data in the Financial Markets*

The book's first section looks at the use of big data in the financial markets, where it can be employed to optimize investment performance in various ways.

The section begins with Chapter 2, *Alternative Data*, where the authors discuss the role alternative data sources (e.g., internet activity, web traffic logs, chat boards, and social media platforms) can play in financial investment decisions. **Grégoire and Jepson** provide a critical analysis of the value gained from using such data sources against the difficulties in using that data—including the high cost associated with collecting, validating, and maintaining the data. The authors discuss the role big data can play in making alternative data more affordable to use.

Chapter 3, *An Algorithmic Trading Strategy to Balance Profitability and Risk*, proposes an algorithmic trading strategy (AT) based on financial indicators. The primary financial indicator this strategy is based on is the Balanced Investment Indicator (BII) which is used to find a balance between profitability and risk for an investor. **Peña** argues that using this proposed financial indicator for portfolio construction in combination with big data can provide an effective algorithmic trading strategy.

Following this, Chapter 4, *High-Frequency Trading and Market Efficiency in the Moroccan Stock Market*, explores whether high-frequency trading (HFT) impacts market efficiency. **Ferrouhi and Bouabdellaoui** employ a case study approach using data from the Moroccan Stock Exchange to evaluate this hypothesis at different time frequencies.

Lastly, Chapter 5, *Ensemble Models using Symbolic Regression and Genetic Programming for Uncertainty Estimation in ESG and Alternative Investments*, presents an ensemble algorithm using symbolic regressions (SR) to estimate the environmental, social, and governance (ESG) scores of publicly traded private equities (PE) and sustainable exchange-traded funds (ETFs). **Venegas, Britez, and Gobet** demonstrate how using the SR algorithm can reduce uncertainty in this investment space.