Journal of Capital Markets Studies

Volume 1 Number 1 2017

Number 1
1 Editorial boards
2 Editorial
5 Statistical studies of financial reports and stock markets
Shyam Sunder
10 Understanding sustainability for socially responsible investing and reporting
Marianne Bradford, Julia B. Earp and Paul F. Williams
36 Leveling the playing field for less-sophisticated non-professional investors: does plain English matter?
Eileen Taylor and Jennifer Riley
58 The performance of US-based emerging market mutual funds
Halil Kiyimaz and Koray D. Simsek
74 Identifying the major reversals of the BIST-30 index by extreme outliers
Kim Ecol
89 The flash crash: a review
Ali N. Akansu

Access this journal online:
www.emeraldinsight.com/journal/jcms

ISSN 2514-4774
Journal of Capital Markets Studies (JCMS) is a high-quality research publication in economics and finance with a specific focus on capital markets. JCMS intends to explore the interface between academic research and practical application by disseminating new research findings alongside articles related to everyday professional practice.

JCMS is a scholarly refereed journal that invites empirical or policy-related manuscripts (research paper, literature review, case studies) as well as theoretical contributions from academics. The journal welcomes quality papers from subjects including, but not limited to:

- Capital Markets and Financial Institutions
- Financial Instruments and Financial Innovation
- Investment and Financial Decision Making
- Financial and Non-Financial Information
- Corporate Reporting & Investment Decision
- Capital Market Governance & Sustainability
- Financial Standards and Regulations
- Financial Risk Management
- Islamic Finance
- Behavioural Finance
- Capital Markets Technology & Cybersecurity

The Journal of Capital Markets Studies (JCMS) is a high-quality research publication in economics and finance with a specific focus on capital markets. JCMS intends to explore the interface between academic research and practical application by disseminating new research findings alongside articles related to everyday professional practice.

JCMS is a scholarly refereed journal that invites empirical or policy-related manuscripts (research paper, literature review, case studies) as well as theoretical contributions from academics. The journal welcomes quality papers from subjects including, but not limited to:

- Capital Markets and Financial Institutions
- Financial Instruments and Financial Innovation
- Investment and Financial Decision Making
- Financial and Non-Financial Information
- Corporate Reporting & Investment Decision
- Capital Market Governance & Sustainability
- Financial Standards and Regulations
- Financial Risk Management
- Islamic Finance
- Behavioural Finance
- Capital Markets Technology & Cybersecurity

JCMS is indexed and abstracted by:

- The British Library

Guidelines for authors can be found at:
www.emeraldgrouppublishing.com/services/publishing/jcms/authors.htm

JCMS is published by Emerald Publishing Limited. For more information about Emerald's regional offices please go to http://www.emeraldgrouppublishing.com/offices

Emerald Publishing Limited
Howard House, Wagon Lane, Bingley BD16 1WA, United Kingdom
Tel: +44 (0) 1274 777700; Fax: +44 (0) 1274 785201
Email: emerald@emeraldinsight.com

For more information about Emerald's regional offices please go to http://www.emeraldgrouppublishing.com/offices

Customer helpdesk:
Tel: +44 (0) 1274 785278; Fax: +44 (0) 1274 785201
Email: support@emeraldinsight.com

The Publisher and Editors cannot be held responsible for errors or any consequences arising from the use of information contained in this journal; the views and opinions expressed do not necessarily reflect those of the Publisher and Editors, neither does the publication of advertisements constitute any endorsement by the Publisher and Editors of the products advertised.

Emerald is a trading name of Emerald Publishing Limited
Printed by CPI Group (UK) Ltd, Croydon, CR0 4YY


ISO 19011:2002
Certificate Number 1985
ISO 14001
ADVISORY BOARD

Professor Reena Aggarwal
Georgetown University, USA

Professor Barry Eichengreen
University of California Berkeley, USA

Professor Umit Erol
Bahcesehir University, Turkey

Professor Mervyn E. King
Mervyn King Ec, South Africa

Professor Prem Sikka
University of Essex, UK

Professor Syham Sunder
Yale University School of Management, USA

Professor Paul Williams
North Carolina State University Raleigh, USA

EDITORIAL ADVISORY BOARD

Professor Ali Akansu
New Jersey Institute of Technology, USA

Dr Yuri Biondi
Cnrs - IRISSO (University Paris Dauphine PSL) and Labex ReFi, France

Prof. Dr Nihel Chabrak
United Arab Emirates University, UAE

Assoc. Professor Cem Demiroglu
Koc University, Turkey

Prof. Dr Colin Haslam
Queen Mary University, UK

Prof. Dr Jim Haslam
University of Sheffield, UK

Professor Burak Saltoglu
Bogazici University, Turkey

Assoc. Professor Koray Simsek
Rollins College, USA

Professor Kamil Yilmaz
Koc University, Turkey

Professor Burcin Yurtoglu
Otto Beisheim School of Management, Germany
Dear Colleagues

As the Founding Chief Editor, I am delighted to publish the very first issue of the Journal of Capital Markets Studies (JCMS), which contains six excellent papers. We launched this journal because we believe there is a need for a medium that covers the wide breadth of capital markets. Our aim is to publish high-quality research in the areas of economics and finance with a specific focus on capital markets.

The JCMS will be the official publication of the Turkish Capital Markets Association (TCMA). TCMA is the non-profit self-regulatory professional organization of brokerage firms, banks and asset management companies in Turkey. TCMA’s first and foremost objective is to contribute to the development of capital markets and expand the professional know-how in the industry. Other major goals of the Association include stimulating international dialogue and understanding of capital markets among academics, practitioners, investors and policymakers through various mediums. I would like to express my gratitude to TCMA chairman Mr Ilhami Koc, and all Board members who have wholeheartedly supported our vision from day one. I would also like thank the secretary general of TCMA, Mr Ilkay Arıkan and his team, who have been most instrumental in getting this very first issue ready in time and continuously helped along the rigorous publication process. Without their endless help and assistance, this journal could not have come to life.

In order to best serve our purposes, we intend to invite scholars around the globe to submit their research to the JCMS. We have brought together a reputable editorial team to establish a solid foundation for our journal. Barry Eichengreen (University of California Berkeley), Shyam Sunder (Yale University School of Management), Paul Williams (North Carolina State University Raleigh), Reena Aggarwal (Georgetown University), Prem Sikka (University of Essex) Umit Erol (Bahcesehir University) and Mervyn E. King (International Integrated Reporting Council) are some of the distinguished academics that agreed to serve as the advisory board members. We strive to diversify our advisory body in terms of experience, affiliation and geography. It is our belief that through our contributors’ respectable academic background and insight, the scope and impact of our journal will be further enhanced.

We propose to arm our readership with the latest research and commentary in all areas associated with the capital markets. We welcome empirical and policy-related manuscripts (articles, surveys and letters) as well as theoretical contributions from academics. The journal may contain a book reviews section.

Topics suitable for JCMS include, but are not limited to, financial institutions; financial instruments and financial innovation; financial decision making; valuation of financial assets; financial and non-financial information; corporate reporting and investment decision; capital market governance and sustainability; investor protection; financial literacy; financial standards and regulations; financial crime; market microstructure; financial risk management; financial engineering and derivatives; financial econometrics; Islamic finance; emerging markets studies; behavioral finance; international financial markets; international/regional...
financial centers; capital markets technology and cybersecurity. We believe there is a broad
global audience of academics and practitioners who will be interested in JCMS’ content.

The first issue consists exclusively of invited research and conceptual papers of
distinguished academics. We begin this issue with a piece of Shyam Sunder titled “Statistical
studies of financial reports and stock markets”. The author examines the
existence of co-variation between financial reporting method and stock market data and
argues that a change of accounting makes no difference to the price system. Shyam points
out the significance of continued research on this issue.

The next paper, “Understanding sustainability for socially responsible investing and
reporting” is co-authored by Paul Williams, Marianne Bradford and Julia Earp. The paper
explores the forms of sustainability activities used by US companies and the extent of
individuals’ understanding of the sustainability measures. It is argued that stakeholders’
comprehension of sustainability activities differs significantly from the generally accepted
structure of the Global Reporting Initiative. The differences in perspective render efforts on
SRI reporting a more complex issue.

Similar to Williams and his associates’ efforts, Jennifer Riley and Eilon Taylor, analyze
issues from the viewpoint of the investors in their paper titled “Leveling the playing field for
less-sophisticated non-professional investors: does plain English matter?” The paper explores
risk factors found in SEC 10Ks and evaluates how the format and wording of those factors
influence investor behavior. The experimental research provides further evidence that plain
English disclosures compensate for a lack of task-specific knowledge and increases credibility
perceptions of management among less-sophisticated non-professional investors.

The next piece titled as “The performance of US-based emerging market mutual funds”
co-authored by Halil Kiymaz and Koray Simsek concerns professional investors. The paper
examines the performance of US mutual funds that invest primarily in emerging market
equities and bonds. In this regard, this study contributes to the scarce literature on these
types of funds and provides a comprehensive performance assessment.

The following paper, Umit Erol’s “Identifying the major reversals of the BIST-30 Index
by extreme outliers” is an empirical attempt for the identification of the major reversals in
the Borsa Istanbul broad market index BIST-30. The author originally makes use of the rate
of change extremity as well as the use of wavelet decomposition for analysis.

The first issue concludes with “The Flash Crash: a review” written by Ali Akansu. The
paper provides an overview of what happened in the financial markets within a few
minutes on May 6, 2010 (the Flash Crash) and the collapse that followed with its historically
unmatched impact. Even though underlying reasons that triggered the Flash Crash are well
understood by traders, regulators and researchers, the author emphasizes that there are still
crucially significant outstanding issues requiring more sophisticated policies, procedures
and regulations to build more robust, fair and transparent financial markets.

We hope you enjoy the first issue of JCMS. Should you have any specific suggestions for
future releases, please feel free to contact us. We value your input. Our e-mail addresses are
provided below and JCMS’s website is available at: www.emeraldgrouppublishing.com/
services/publishing/jcms/index.htm

Güler Aras
Yildiz Technical University, Turkey

About the Editor-in-Chief
Güler Aras, PhD, CPA, is a Professor of Finance and Accounting at Yildiz Technical University, and
the Founding Director of Center for Finance, Governance and Sustainability (CFGS) at YTU. Professor
Aras is the Founding Chair of the "Integrated Reporting Network Turkey". She was a Visiting
Professor at Georgetown University McDonough School of Business and she is the former Dean of the
Faculty of Administrative and Economic Sciences and the former Dean of the Graduate School.
Her research focus is on financial economy and financial markets with particular emphasis on the relationship between sustainability, corporate governance, corporate social responsibility and corporate financial performance. She has published more than 25 books and has contributed over 250 articles in academic, business and professional journals. She also edited several book collections and conference proceedings. Her latest books, Sustainable Markets for Sustainable Business: A Global Perspective for Business and Financial Markets (2015) and co-authored books Transforming Governance (2016) and Corporate Behaviour and Sustainability: Doing Well by Being Good (2017) published by Routledge. She is the Editor in Chief of Journal of Capital Market Studies, Editor of Routledge Book Series, Finance, Governance and Sustainability: Challenges to Theory and Practice and the Editor of Routledge book series Corporate Social Responsibility; she has also served as an Editor of Social Responsibility Journal and Emerald Development of Governance and Responsibility book series. Professor Aras has spoken extensively at professional and academic conferences and has served as a Consultant to number of governmental and commercial organizations such as Minister of Development, Undersecretary of Treasury and Minister of Labour and Social Security Employment in Turkey. Professor Aras is working with business closely and she is acting as an independent board member of several institutions and independent board member of “Turkish Capital Market Association (TCMA)”. 
Abstract

Purpose – The purpose of this paper is to examine the usefulness of statistical studies of financial reports and stock market data for improving corporate financial reports.

Design/methodology/approach – Analytical writing.

Findings – It is often claimed that statistical studies of co-variation between financial and stock market data can help set better financial reporting policy. Such co-variation, even when it can be estimated, tells us little about which financial reports help to make better financial decisions. A case in support of such claims remains to be made.

Practical implications – The readers are advised to be extremely careful in drawing inferences from studies of co-variation between accounting and stock market data for financial reporting policy.

Social implications – Inference from accounting empirical studies to policy needs better rationale to avoid bad policy consequences.

Originality/value – This paper raises original questions about policy inferences from a large class of empirical research in accounting.

Keywords Efficient markets, Financial reporting policy, Statistical co-variation

It is possible to define better financial reporting in terms of the observable characteristics of the system, such as relevance, reliability, timeliness, etc. Such assessments, being dependent on context, are difficult to generalize; and it seems more attractive to assess financial reporting regimes on the basis of their observable consequences. Market phenomena being more readily observable than individual and organizational behavior, this line of reasoning leads one to the goal of financial reporting for better markets.

Financial reporting serves many constituencies, including markets for financial capital, labor, and products and services. There has been a strong but largely unexplained inclination to focus attention on markets for financial capital, leaving the other two classes of markets largely out of the accounting discourse. Even within the class of capital markets, most of the attention has been given to markets for common equity of publicly traded corporations.

This paper examines the inferences we may or may not be able to draw from statistical studies of financial reports and secondary markets for equity of public firms. The focus on equity markets should not be taken to suggest that the role of financial reports in other markets is unimportant. Nor does it imply that “better markets,” however defined, necessarily lead to higher welfare or a better society. Important as these questions are, they are outside the limited scope of this paper.
Better markets for common equity

It is useful to have a shared view of what we regard as better markets for common equity. Once we have mapped the linkages between alternative financial reporting regimes and the characteristics of equity markets, we may have the chance to improve markets by selecting financial reporting regime(s). The concepts of perfect and complete markets in economics suggest a promising starting point. A market is regarded as being perfect if (among other things) (1) all participants have the relevant information, (2) no participant has the power to influence the prices, (3) the cost of entering and exiting the market is zero, and (4) all participants have access to the relevant production technology. A market is complete if all possible resources, and state-contingent claims on them, are traded readily.

Perfect and complete markets are the idealized and simplified economic models of a complex world. Whether or not we can alter a market to be more perfect and complete, and whether or not such alteration is desirable, are difficult questions. When the mechanism to intervene in the market, with its own imperfections, is included in analysis, the results of such intervention are not necessarily closer to the goal of perfection. The desirability of intervention depends on the point of view selected for analysis, because few interventions can bring about Pareto improvements in the status quo.

The information condition for market perfection (mentioned as Item 1 above) gives rise to the possibility of a linkage between financial reporting and market results. However, financial markets operate in an environment of uncertainty, where perfection is not achievable, and the ideal is shifted from perfect to efficient markets. Informationally efficient markets are defined as those in which prices incorporate the aggregation of all information in possession of the market participants. In other words, prices in efficient markets are the same as they would be had every participant possessed the aggregation of all the information possessed by each of its individual participants. As Hayek (1945) pointed out, markets function as aggregators of bits and pieces of information in possession of the participants.

The market’s aggregation property, to the extent it is valid, tells us little about which pieces of information should be in the hands of which traders, and when traders should have the information for markets to be more efficient. Since much of the information about business concerns uncertain events of the future beyond anyone’s knowledge, making “all information” available to all the traders is not possible. Of the limited amount of information possessed by corporate managers, making it all available to the public has proprietary, regulatory, and competitive implications. Further, it is not clear what mechanism(s) can be used to induce managers to disclose information when they do not think that disclosure will serve their own personal interests.

Statistical analyses of linkages between financial reports and stock market data have sought to guide financial reporting policy. The argument is vaguely appealing: stock markets are efficient in the sense of reflecting all relevant information in the prices; we can therefore use guidance from stock markets to identify which information (and which information-generating accounting methods) markets find relevant to determination of prices; this can be achieved by identifying the accounting methods which produce data with greater statistical proximity (e.g. higher product moment or rank correlation) with stock prices; and such methods of accounting can be used to promote market efficiency. Since ideal markets in the presence of uncertainty are efficient, this method of arriving at accounting policy will promote market efficiency, and thus “better markets.” Each of these four arguments calls for some scrutiny.

Stock markets are efficient

What is the sense in which stock markets can be regarded as efficient, and what are the implications of the various available interpretations for financial reporting policy?
Interpretations range from “the price is right” to more or less tautological statements about efficiency (e.g., price is a function of the information system associated with the market, and each information system generates its own “efficient” market price). Under the first interpretation, price, being independent of the information system, becomes irrelevant to the determination of financial reporting policy. Under the second interpretation, the choice of reporting policy requires choosing among the multiple efficient equilibria (see Kanodia, 1980 on multiple efficient equilibria and Sunder, 1989 on the problems of using event studies from efficient markets to guide accounting policy).

Using accounting methods relevant to the determination of prices
This is a commonsense argument. The problem is, that under the first interpretation of efficiency, accounting methods do not matter. Under the second interpretation, all accounting methods (excluding trivial equivalents) are relevant.

Statistical proximity (e.g., correlation) as identifying criterion
If we use statistical proximity between accounting and stock market variables as the criterion for choosing among accounting alternatives, achieving greater proximity, indeed identity, is trivially simple: by using market variables as the basis of accounting. For example, if income were measured as the change in market capitalization of equity (adjusted for capital transactions and dividend payments), the accounting measures would be identical to market measures. Yet such a system, with perfect proximity between accounting and market variables, will provide no information to the market participants to assist them in making their investment decisions.

Such accounting methods promote efficient (“better”) markets
As mentioned above, accounting methods chosen on the basis of greater statistical proximity between accounting and stock market variables provide no assurance that they will help make the markets more efficient or better in some specified respect.

It has been popular to suggest that one should choose financial reporting practices on the basis of statistical co-variation between accounting and stock market data. I used this approach in my thesis (Sunder, 1973, 1975), and found it to be problematic. Yet, here is a quote from a recent paper:

This paper aims at determining the value relevance of financial reporting. […] This study aims at explaining likely impact of financial reporting by listed companies on the market prices of their shares. Our study reveals that the value relevance of published financial statements, per se, is negligible. […] The results of our investigation depict negligible value being added by cash-flow reporting (Vishnani and Shah, 2008, p. 84).

There are two problems with the argument. I shall not address the first: could the stock market be the sole or dominant basis for choosing corporate financial reporting? Rather, I focus on the second: does statistical co-variation between accounting and stock market data deliver on its promise of helping us choose better methods of financial reporting?

Consider the basic structure of the large body of research studies based on the premise that the criterion of statistical co-variation between accounting and stock market data can help us choose better financial reporting methods. The status quo Financial Reporting System A presumably causes Price System A, the causal direction being inherent in the search for information for better markets.
Once we have data on both sides of this arrow, we can plug that data into estimation algorithms on a computer and estimate $R(A)$, where $R(A)$ is the co-variation of accounting data on the left and prices on the right.

Financial Reporting System A $\rightarrow$ Price System A

$R(A)$

Suppose a policymaker is considering a hypothetical alternative, Financial Reporting System B. Presumably, Financial Reporting System B will cause Price System B to be generated.

Financial Reporting System B $\rightarrow$ Price System B

$R(B)$

It is conceivable that we could find some way of estimating what the accounting numbers will be under B, and then we would like to do a co-variation estimate, $R(B)$, which is the relationship between the stock prices under B and accounting data under B.

Financial Reporting System B $\rightarrow$ Price System B

What can we learn about the relative desirability of Method A vs Method B by comparing the co-variation of each method with the respective stock prices they generate – $R(A)$ and $R(B)$? Suppose that the statistical proximity of accounting numbers to stock prices is to be the criterion to be maximized. As mentioned earlier, it is trivial to achieve this accounting nirvana; all one has to do is to report the change in market capitalization of the firm as income, close the accounting department, and achieve a perfect “information” system. However, it would be accounting from the markets and not accounting for the markets.

But things are much worse than the failure of co-variation to guide us to better systems of reporting; in many cases, the co-variation $R(B)$ itself cannot be estimated. One might be able to estimate the accounting data under B which has not yet been implemented. Furthermore, one does not have the stock price data under the hypothetical financial reporting regime B, and therefore, one has no way of even estimating $R(B)$ so it can be compared to $R(A)$.

Financial Reporting System A $\rightarrow$ Price System A

$R(A)$

Financial Reporting System B $\rightarrow$ Price System B

$R(B)$

What can one do without data on Price System B? Here’s a trick, or sleight of hand – estimate the co-variation $R^*(B)$ of Financial Reporting System B with Price System A (Figure 1).

The published study of Vishnani and Shah (2008), as well as many other accounting studies, followed this route. What does the comparison of $R(A)$ with $R^*(B)$ tell us? It cannot tell us whether B is better or worse than A. In fact, their comparison is quite irrelevant, unless one believes that Price System B is the same as Price System A. Even then, a change

Figure 1.
Dealing with the absence of $P(B)$ by a sleight of hand
of accounting makes no difference to the price system. If accounting makes no difference to stock prices, why choose accounting on the basis of stock prices?

In summary, it is hard enough to derive the logical inference from the comparison of $R(A)$ and $R(B)$. The comparison of $R(A)$ to $R^*(B)$, which is the norm in a significant part of accounting literature, is not relevant to policy. Accordingly, the case for using the results of these popular studies of co-variation between accounting and stock market data for the purpose of setting financial reporting policy remains to be made.

References


About the author

Shyam Sunder is the James L. Frank Professor of Accounting, Economics and Finance at Yale School of Management. His research on financial markets and macroeconomics has appeared in over 200 articles in prestigious accounting, economics and finance journals. His inquiry (with D. Gode) into structural properties of markets using zero- or minimally intelligent agents was a seminal innovation whose applications have spread to computer science, engineering, finance, artificial intelligence, e-commerce, and other disciplines. His tenth book, Risky Curves: On the Empirical Failure of Expected Utility, co-authored with Dan Friedman, Mark Isaac, and Duncan James, appeared in 2014. Shyam Sunder can be contacted at: shyam.sunder@yale.edu; http://faculty.som.yale.edu/shyamsunder/

For instructions on how to order reprints of this article, please visit our website:
www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: permissions@emeraldinsight.com
Understanding sustainability for socially responsible investing and reporting

Marianne Bradford, Julia B. Earp and Paul F. Williams
Poole College of Management, North Carolina State University, Raleigh, North Carolina, USA

Abstract
Purpose – The purpose of this paper is to determine what types of sustainability activities companies are reporting and whether persons external to the companies understand how those reported activities correspond to the companies’ narratives about sustainability. That is to ascertain how people interpret the meaning of the activities included in the sustainability reports.

Design/methodology/approach – From a sample of sustainability reports prepared by Global Reporting Initiative (GRI) guidelines, the authors identified the distinct activities reported. The authors prepared a survey comprised of these activities and asked a sample of people knowledgeable about business and investing to evaluate each activity on the extent to which they are relevant to sustainability performance. The responses were then factor analyzed to identify the most important dimensions of sustainability these persons employed to relate the activities to sustainability.

Findings – The dimensions employed by the subjects differed in some significant ways from those dimensions used to construct the GRI format. Subjects evaluated sustainability efforts as primarily efforts of being a good citizen with sustainability an end in itself rather than as constraint to be respected in achieving profitability goals.

Research limitations/implications – The study is a first attempt so results are preliminary, i.e. suggestive but not definitive. Though preliminary an intriguing implication is that closure on a sustainability reporting structure would be premature. More effort needs to be devoted to provide more clarity on the concept of corporate sustainability and what its implications are for corporate behavior.

Practical implications – Given the results that sustainability be regarded as a corporate end, what is the role of the corporation in society seems still to be disputatious. Sustainability may not be something achievable without changes in corporate law.

Originality/value – The study is an early attempt to assess the potential alternative narratives about corporate sustainability. Its value lies in providing insights into the age-old question of what should be the role of the corporation in a free society.

Keywords CSR, Sustainability, Global reporting initiative, Sustainability reporting, Stakeholder narratives

Paper type Research paper

Introduction
A Google search of “Socially Responsible Investing” will yield nearly 2,000,000 results, indicating the extent of interest by investors for investing in firms that are sustainable in the long run. Socially responsible investing has been a significant component of the investing landscape for at least the past 50 years (Rockness and Williams, 1988). As the persistence of problems, most notably environmental ones associated with climate change, have made a business response more-and-more mandatory, the terminology of corporate social responsibility has evolved. Today it is more common for firms to refer to their “sustainability” efforts in recognition that social responsibility is no longer a choice but is...
essential to sustaining a firm’s long-run viability. According to Morgan Stanley (2017, p. 1), “Sustainable investing is surging, having enjoyed a 135% increase in assets under management since 2012 to $8.72 trillion, and it’s still growing.” Now, even sustainability investing is in transition into a new approach now labeled environment, social, and governance (ESG), with firms being expected to provide detailed information about their accomplishments vis-à-vis the environment, the community and society, and the way in which they are governed (Mahn, 2016). Responsible investors recognize that sustainability encompasses these three areas, and are seeking to invest in companies that communicate clearly that they recognize this. Concomitant to the growth in sustainability investing is the development of systems of reporting information about firms’ sustainability performance. Unlike financial reports, which have been shaped by centuries of business and financial practices, reports on the results of sustainability performance have yet to reach the structured status of financial reports. In producing information about sustainability, there is still considerable ambiguity about what sustainability, or ESG, means when viewed as a comprehensive narrative about sustainability performance and, therefore, what represents pertinent information about whether a firm is performing in ways that achieve sustainability goals.

Sustainability is a simple enough concept when viewed as a dictionary definition, but may have different meanings because of connotations that depend upon the context in which the term is applied or the perspective of those employing the term. Ambiguity about the meaning of sustainability arises in particular when a company is reporting on its sustainability practices. While some companies focus only on environmental sustainability, others view sustainability as a “three legged stool,” including activities that are economically, environmentally, and socially responsible (Butler et al., 2011). One well-known definition of sustainability, which focuses on sustainable development, is “development that meets the needs of the present without compromising the ability of future generations to meet their own needs” (United Nations World Commission on Environment and Development 1987)[1][2].

Because many more investors are becoming sensitive to how the companies in which they invest perform sustainably, an increasing number of companies are investing time and effort into producing reports that describe their sustainability practices (Rahman and Akhter, 2010). Sustainability reporting has emerged as a common practice of twenty-first century business. Where once sustainability disclosure was the province of a few unusually green or community-oriented companies, today it is a best practice employed by companies worldwide (Ernst & Young, 2016). However, companies do not pursue similar objectives, nor do they use the same metrics to gauge their sustainability activities, making comparison among these reports difficult (White, 2005). The resurgence in environmental and corporate sustainability reporting (CSR) also suffers from its being merely a resurrection of the 1970s view, i.e. “[…] the disclosure appears to be viewed with little or no skepticism regarding its purpose” (Cho and Patten, 2013, p. 446).

There are various reporting guidelines companies can follow, but one of the most prominent ones comes from the Global Reporting Initiative (GRI) (2015), an Amsterdam-based nonprofit organization comprising business, government, social advocacy and other stakeholders[3]. The Global Reporting Initiative (GRI) maintains a “Sustainability Disclosure Database” that provides access to an abundance of sustainability information voluntarily reported by companies. Any company that has compiled a sustainability report is eligible to register it with the GRI. However, reporting on any particular activities is not mandatory, although the GRI suggests nearly 80 sustainability activities (indicators) in six different dimensions. These dimensions are labor and decent work, economic, environment, human rights, society, and product responsibility (Global Reporting Initiative (GRI), 2011[4]).

The mounting importance of sustainability and its inclusion in companies’ strategic initiatives may likely cause this type of information to be mandated in the near future and require assurance through a “unifying, global framework: that of integrated reporting”
According to an Ernst and Young study, 47 percent of respondents use third-party assurance to add credibility to information presented to external shareholders (Ernst & Young, 2011). Attestation via audit reduces the risk of misstatements associated with sustainability reporting and sends a message that reports are relevant, reliable, and free from bias.

This study investigates the question of what sustainability activities mean to external recipients of CSR compared to the meaning implied by the organization of information suggested by the GRI framework for reporting about sustainability. In other words, we are seeking to learn whether activities reported by companies that are related to sustainability comprise a more-or-less coherent narrative about corporate performance that, when considered by external recipients, matches the implied structure of the narrative provided by GRI guidelines. Our study considers the concept of sustainability broadly as including environmental, economic, and societal performance, which has been referred to as the “three legged stool” or “triple bottom line” (TBL) (Butler et al., 2011; Ernst & Young, 2011). The three-legged stool is consistent with an older concept of corporate social responsibility. This research is intended to provide preliminary insights into what stakeholders understand sustainability to mean, and to suggest sustainability activities and their interrelationships that need to be considered by corporations in their business processes when measuring and reporting them.

The remainder of this paper is organized into the following sections: a brief history of CSR reporting; a discussion of the narrative nature of corporate reports; stakeholder perceptions of sustainability actions; what sustainability information is reported by firms and how is it organized; what sustainability message is being received, i.e. how do external recipients of company reports organize the same information; concluding remarks.

Corporate sustainability and corporate reporting
Since the publication of the classic work by Berle and Means (1932), there has been continuous argument over what the role of the private corporation should be in a democratic society and, thus, what the responsibilities of its management should be[5]. The “classical” view of the role of the corporation is most succinctly stated by Friedman (1970), who claimed that the sole responsibility of the corporation is to singularly pursue profits for its shareholders. An alternative narrative about the role of the corporation grew in prominence after the Second World War with the scientization (Whitley, 1986) of the business disciplines and achieved its greatest prominence during the 1960s and 1970s. This alternative narrative about the corporation’s role has been labeled by Danley (1994) as “managerialism.” Managerialism is the notion that the corporate manager is a professional equipped with the tools of scientific management, who performs as an economic statesman, that is, he manages the corporation cognizant of the multiple effects corporate action has on many identifiable social groupings (stakeholders)[6]. Corporations have relationships with other than shareholders, and failure to act responsibly in those relationships will affect shareholders in both the short and long runs. Stakeholder theory is a theory of organizations developed initially by Freeman (1984) and is based on a managerialist narrative of the purposes of the corporation. Freeman’s (1984) definition of a stakeholder is “[...] any group or individual who can affect or is affected by the achievement of the organization’s objectives” (p. 46). Unlike the classical economic model of the organization as a receptacle of inputs from capital and labor providers to produce outputs for customers (with profit being the singular motive), stakeholder theory “[...] describes the corporation as a constellation of cooperative and competitive interests possessing intrinsic value” (Donaldson and Preston, 1995, p. 66). Thus, the stakeholder theory of organizations implies a role for management not as one solely focused on managing inputs of capital and labor, but one of managing “stakeholders,” i.e. the “[...] simultaneous attention to the legitimate interests of all appropriate
stakeholders” (Donaldson and Preston, 1995, p. 67) for the long-run survival of the firm. Unlike the standard economic view of organizations as mechanisms for turning inputs into outputs in an “economically” efficient way (the engineering approach (Sen 1988)), the stakeholder theory contains a normative element. It acknowledges that corporations are not simply nexuses of private contracts (Greenfield, 2006) nor is shareholder value the only responsibility of corporate managers (Stout, 2012). Corporations are responsible for the consequences of their conduct. Just as persons are responsible to others, the managerialist view is that corporations likewise have responsibilities to others.

A typical early expression of managerialism was provided by the Committee for Economic Development (CED) (1971):

> The modern professional manager also regards himself, not as an owner disposing of personal property as he sees fit, but as a trustee balancing the interests of many diverse participants and constituents in the enterprise, whose interests sometimes conflict with those of others (p. 22).

The rationale most frequently provided for this socially responsible management is “enlightened self-interest,” which is as follows:

> There is a broad recognition today that corporate self-interest is inexorably involved in the well-being of the society of which business is an integral part, and from which it draws the basic requirements needed for it to function at all – capital, labor, customers (sic). There is increasing understanding that the corporation is dependent on the goodwill of society, which can sustain or impair its existence through public pressures on government. And it has become clear that the additional resources and goodwill of society are not naturally forthcoming to corporations whenever needed, but must be worked for and developed (CED, 1971, p. 27).

During the 1970s, this narrative of the role of management went under the descriptor “corporate social responsibility” (Gambling, 1974; AICPA, 1977; Estes, 1972; Human Resources Network, 1975; Ramanathan, 1976; Chen, 1975, Bauer and Fenn, 1972). The ideas of social reporting and the social audit gained prominence in the early 1970s through the recognition that the modern business corporation, because of its size and power, and thus visibility, affected society, not just parties directly associated with the production process of the company (Danley, 1994). Many of the themes of CSR (e.g. environmental stewardship and employee welfare) have recently been incorporated under the rubric of “sustainability” to reflect a new urgency to the importance of these themes to the long-term success of the enterprise. The enlightened self-interest rationale for CSR made this case that acting in socially responsible ways came at no cost to profitability. Because of changing in societal values, doing right things would translate into making money. A similar but more urgent rationale is now being provided for sustainability (Laszio and Zhexembayeva, 2011; Esty and Simmons, 2011; Andrew and Cortese, 2013).

Corporate sustainability is a response to the environmental realities of climate change, species extinction, and resource decline. Contributing to the CSR movement of the 1970s was The Club of Rome report in 1972 (Meadows et al., 1972), which raised the alarm about the global system feedback loops that predicted dire environmental consequences should the trends continue. Sustainability has an even greater urgency since the 30-year follow-up study to the Club of Rome report (Meadows et al., 2004) indicates the environmental crisis described in the 1972 report has worsened rather than improved. Leading management scholars and consultants on corporate sustainability described the situation as follows:

> [...] never before have we seen the speed, extent, and magnitude of resource loss that we observe now. Whether it is soil, water, nutrition, a stable climate, or social equity as measured by the rich-poor gap, the list of declining resources in question is relevant for nearly the entire global economy, with no company left unaffected. And that, in turn, creates a fundamental change in how companies compete to create enduring value (Laszio and Zhexembayeva, 2011, pp. 9-10).
Reminiscent of the CED narrative from 1971, these authors see the remedy for declining global resources in a new paradigm still centered on business and corporate action, i.e.:

We are committed to sharing an exciting but largely invisible story of a shift in the conduct of business. In the new narrative (emphasis in original), the gloom and doom of declining resources is also the foundation for opportunity, an emerging paradigm of business that can be more sustainable and (emphasis in original) profitable (Laszlo and Zhexembayeva, 2011, p. 9).

This proposed new narrative is not all that new. It shares many features with the CSR one from 40 years ago. Then, CSR was alleged to translate into financial success; now sustainability will likewise translate into financial success.

Just as the social reporting/social audit movement encouraged the development of reporting schemas that could be audited to provide transparency on corporate social performance to various constituents (e.g. AICPA, 1977; Belkaoui, 1984), sustainability has led to the development of reporting schemas for providing transparency for corporations’ sustainability efforts. Indeed, the fostering of sustainability within the corporate world requires it, i.e., “Integrated business practices and reporting are key factors in fostering sustainability” (Brockett and Rezaee, 2012). Even though CSR/sustainability reporting has been around for over 40 years, there has yet to arise an integrated reporting system that corresponds to the narrative about being profitable by being socially responsible. In the next section, we will discuss the conceptual difficulties with creating such a system.

The problem of a structured narrative for sustainability reporting

Over its history, CSR/sustainability reporting has clung to a tripartite rubric of corporate performance. Though the elements of that performance have changed over time (operating in South Africa no longer carries the stigma it did before apartheid was eliminated), the reporting model is still ably described by the terminology “TBL.” The TBL emphasizes three basic elements of performance deemed most relevant for CSR/sustainability. Financial performance, the traditional bottom line, is one dimension reflecting the economic values attributable to corporate actions. Within the classical managerial narrative this is the only bottom line with which a corporation should concern itself (Danley, 1994). Fulfilling social and environmental values (Elkington, 2004) represent the remaining two dimensions of reporting.

TBL reporting is intended to provide information to the various stakeholders that will permit them to assess the performance of any corporation with respect to each of economic, social, and environmental performance. How the various “metrics” or “indicators” representing such a wide variety of actions are integrated by any stakeholder into a comprehensive and coherent assessment of overall performance is, as yet, not well understood. There is substantial research suggesting that various stakeholder assessments of corporate sustainability performance are, at best, only weakly correlated or even inversely correlated to the amount of TBL reporting that is provided by the companies themselves. This has led some academics to the conclusion that the substance of CSR is merely public relations (Milne and Patten, 2002; Laufer, 2003; Freedman and Patten, 2004; Milne et al., 2006; Deegan et al., 2006; Cho and Patten, 2007; Duchon and Drake, 2009).

Bayou et al. (2011) introduce McCumber’s (2005) concept of “situating” to characterize the nature of “truth” in accounting:

Accounting information systems (bookkeeping) basic foundational structure is essentially the chronological situating of the facts of the occurrences of specific events to be remembered over a timeline proceeding from the past to the present […]. Thus, the narrative of accounting is focused on responsibilities fulfilled or not fulfilled; it is a narrative that has historically been intended to provide the reliable memory about the important events that occurred in the past in order to determine what are the consequences up to now […] (Bayou et al., 2011, p. 118).
What TBL reporting purports to do is to broaden the narrative about consequences of corporate actions to include events that affect the long-term sustainability of a firm, which broadens the responsibilities and strategic horizons of corporate management. The complexity of managing these broadened responsibilities can be illustrated via a simplified model adapted from Williams (2010, p. 27) and Bayou et al. (2011, p. 121), which is developed below.

The first of the TBLs representing the economic dimension is a highly structured narrative (or codified discourse (Llewellyn and Milne, 2007)) about the role of the corporation in society that corresponds to the classical model. It takes the form of the familiar accounting equation:

\[
\text{Net income} = \text{Revenues} - \text{Expenses}
\]

It is structured because the terms of the expression are defined by an elaborate regulatory bureaucracy consisting of standard-setters (e.g. FASB, IASB), governmental regulatory bodies (e.g. SEC, IRS, HMRC), and legislative bodies (e.g. US Congress, UK Parliament). In addition, centuries of commercial practices and business cultures influence the bureaucratic apparatus that specifies what activities constitute those that lead to “revenues” and what activities lead to “expenses.” Bayou et al. (2011) provide an expanded version of this expression that illustrates the fundamental value judgments underlying this representation of corporate economic success and how this expression implicitly contains the activities represented by the other two of the TBLs, but at the same time suppress their fulfillment, thusly:

\[
\text{Gross income of shareholders} = \text{Revenues} - (\text{Payments to: labor, suppliers, creditors, governments}) - (\text{Net externalities})
\]

The payments to labor, suppliers, creditors, and governments are part of the gross incomes of those stakeholders and, within the classical model of corporate purpose being the maximization of shareholder gross income, this implies that management actions must be simplistically geared toward privileging the short-term income of shareholders over those of other stakeholders. The last expression, “net externalities,” is presumed to be taken care of by the rules of the economic game (Friedman, 1970) comprised of market forces, laws that make certain actions illegal, and social/business customs. Net externalities represent the positive and negative effects of corporate actions that make up the social and environmental values that comprise the remaining two bottom lines of TBL that are not easily represented in financial terms[8]. The GRI labels for the components that make up net externalities are currently environment, human rights, labor practices and decent work, product responsibility, and society.

TBL reporting presumes a managerial, rather than a classical, model of the corporation since it advocates for the corporation taking ownership of its social and environmental externalities and managing them so as to internalize the negative ones and create more of the positive ones. Rearranging the income equation once again we get a representation of corporate responsibility implied by TBL, i.e.[9]:

\[
\text{Revenues} = \text{Gross income of shareholders} + \text{Gross incomes of other stakeholders} + \text{Net positive externalities}
\]

Corporate resources that flow from the customer (a primary stakeholder) are available for providing benefits, either financial or otherwise, to all of the other stakeholders with none occupying the principal place in the objective function[10]. The resources the corporation has available may be used for a multiplicity of outcomes affecting various stakeholders...
none of which necessarily takes priority over the others, which makes the management task more complex since multiple objectives exist and must be traded off. An obvious problem with such tradeoffs is that social and environmental effects represent the as yet unstructured portion of the managerial narrative about CSR/sustainability. Were we able to achieve what Gambling (1974) proposed as the ideal measurement scheme of monetizing social and environmental outcomes, then profit would be a comprehensive indicator of sustainable corporate performance. However, such a prospect is very unlikely (Jensen, 1977) since these outcomes involve incommensurate values.

Corporate environmental and social performance reporting focuses on describing activities and their outcomes deemed to lead to positive net externalities. Various groups (e.g. GRI, Sustainability Accounting Standards Board) have proposed reporting formats or rubrics for identifying and organizing information about externality affecting activities. That is, there are more efforts being made to create a structure for corporate social and environmental performance reporting to match the structured nature of economic performance. Recent efforts by some of the world's largest corporations (e.g. Walmart, Coca Cola) to become more sustainable are emphasizing what Dauvergne and Lister (2013) have described as “eco-business.” The sustainability narrative of eco-business is one of being profitable by doing good and is substantively about maintaining control over and access to vital natural resources (e.g. water and “eco-efficiency”), which Dauvergne and Lister (2013, p. 58) describe as conserving energy, managing material usage through reductions and recycling, reducing toxins, and reducing water usage. The focus is on cost reduction and is not a dramatic departure from the way business has described social responsibility/sustainability for nearly half a century – eco-business pays off through lower costs and more secure supply chains. The social and environmental values pursued most energetically seem to be those that demonstrably contribute to achieving long-run economic success. This view of “sustainable development” has been subjected to rather extensive critiques (e.g. Gray, 2010; Milne et al., 2006).

Stakeholder perceptions of sustainability activities
As noted previously, the most prevalent attempt to date to provide a structure to the narrative about the net externalities component of corporate sustainability is that of the GRI. The GRI provides dimensions under which activities are classified that allegedly contribute to fulfilling social and environmental values. Companies are encouraged to report their sustainability activities under the various dimensions in order to provide consistency and comparability with respect to assessing social and environmental values fulfillment. But whether the GRI actually represents a coherent system of reporting depends on whether the activities reported by various companies are interpreted by the various stakeholders in the same manner as implied by the GRI framework. To be a structured narrative like that for financial reporting, the activities reported by companies should imply the achievement of sustainability ends across recipients of the companies' reports. The activities reported should situate the company as a sustainable or socially responsible company just as an income statement permits situating the company as a profitable or unprofitable one. Such reports would facilitate social responsibility investing as well as giving assurance to other stakeholders that the firm is a trustworthy steward of those things stakeholders value. So the narrative about corporate sustainability should convey a reasonably common message among investors and other external users of corporate reports. Activities reported should imply companies are pursuing the same objectives across stakeholder groups.

There is extensive research concerning the contents of CSR. One criticism of CSR is the extent to which what is reported and how it is reported is still largely at the discretion of management. The research reported in this paper is an initial effort to understand
whether “sustainability” under the GRI framework represents a shared understanding of sustainability about corporate performance received by external parties. We are making a preliminary attempt to address a significant question recently raised by Milne (2013, p. 143):

At the heart of assessing corporate “sustainability” reporting are fundamental differences about what corporate reporting for sustainability means (emphasis in original) and, implicitly within these differences, what purposes it serves (or might serve), and whose interests are (or might be) served by it. What is to be sustained?

To achieve our overall research objective, we performed a number of preliminary steps. These steps comprise a process of distillation to arrive at a set of activities that can be confidently classified as representing a particular GRI dimension. These activities were then employed in a factor analysis of external users’ perceptions of the dimensionality of these activities. The result is a set of sustainability dimensions from the perspective of those viewing the firm from the outside.

“Sustainability” reported: what actions do corporations report?

To ascertain what actions companies include in their GRI reports, we employed content analysis, which is a grounded theory approach (Glaser and Strauss, 1967) for making valid inferences from text-based data (Krippendorff, 1980). A sample of companies was drawn from the GRI Sustainability Disclosure Database (GRI, 2011) of USA sustainability reports. We used a diverse sample in our analysis, as well as a purposeful sampling strategy to capture heterogeneity (Maxwell, 1996). The sample included companies in industries that are known to be “embracers” and those known to be “cautious adopters” of sustainability practices as well as those that fall in between (Haanaes et al., 2011). We determined that selecting 15 companies would include the diverse types of industries identified by Haanaes et al. (2011). To determine the companies in each industry, we randomly selected companies listed in the Sustainability Disclosure Database until we had one from each of our 15 predefined industries. If more than one company was selected from a given industry, we only included the first in our selection of 15. Our study is based on the activities that companies report under sustainability dimensions so the particular companies are not of interest. The companies’ purpose for us is simply to provide a diverse set of sustainability activities that are actually reported. Since each of the approximately 40-100 page reports describes many activities, our resulting sample of activities ended up being quite large. Table I presents the 15 companies randomly chosen for our sample.

The GRI reports of the 15 companies were thoroughly read and analyzed by the authors. The focus of our analysis was the identification of activities engaged in by the companies that pertained to achieving economic, environmental, or social values. “Activities” was interpreted broadly to include policies, programs, or specific actions that the companies deemed reportable either quantitatively or qualitatively as significant for telling their sustainability stories[12]. In other words, the activities conveyed meaning to stakeholders about the companies’ sustainability performances. We next categorized each of the activities

| Abbott | Dresser Rand | LG |
| Alcoa | ExxonMobil | Mohawk |
| AT&T | Fifth Third Bank | Toshiba |
| Caterpillar | Haworth | Volkswagen |
| Coca-Cola Bottling Co. | Intel | Walmart |

Table 1. Company sustainability reports analyzed
as belonging to one of the GRI dimensions, which classify activities based on general guidelines[13]:

1. Economic – activities that describe how a company affects the economic conditions of its stakeholders and systems, locally, nationally, and globally.
2. Environment – activities that affect both living and non-living systems of nature.
3. Human rights – activities that indicate how a company considers human rights in its daily operations, specifically with regard to suppliers and contractors.
4. Society – activities that specify how a company affects its communities and the risks that result from interactions with other social entities.
5. Product responsibility – activities that describe how a company’s products and services directly affect the well-being of customers.
6. Labor – activities that are recognized and defined by a variety of universal labor standards set forth by the United Nations and others (GRI, 2015).

The categorization process consisted of several steps that involved six people (the researchers and graduate research assistants), who enabled cross-checking to minimize classification ambiguity. First, GRI reports were read for the purpose of identifying reported “activities” (as defined above). Next, one set of researchers categorized these activities as belonging to one of the six GRI dimensions. Then, those categorizations were verified independently by a second set of researchers. Thus, if a reported activity was related to sustainability, but did not specifically refer to a GRI dimension or for which the two groups of researchers could not reach consensus, we made note of the activity for later categorization. We identified two such additional, non-GRI categories upon which companies were reporting, which we labeled “governance” and “risk and responsibility” (categories combined in Table II)[14]. If an activity in a report was not related to any dimension of sustainability, then it was not categorized. For example, an activity found in Intel’s report we deemed not related to sustainability is: “We sell our products primarily to original equipment manufacturers (OEMs) and original design manufacturers (ODMs).” Statements like these were not included in any of our categorizations of sustainability activities. Table II provides some examples of GRI performance indicators (activities) found in our sample.

The initial identification and categorization process resulted in 476 sustainability activities across the 15 sustainability reports. These data enabled us to ascertain what activities within GRI dimensions US companies are emphasizing.

Our content analysis revealed a number of sustainability activities within each dimension (Table III). Our research indicates that the following dimensions (in order of

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Example sustainability activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Economic</td>
<td>Financial implications and other risks and opportunities for the organization’s activities due to climate change</td>
</tr>
<tr>
<td>Environment</td>
<td>Percentage of materials used that are recycled input materials</td>
</tr>
<tr>
<td>Human rights</td>
<td>Total number of incidents of discrimination and corrective actions taken</td>
</tr>
<tr>
<td>Labor practices and decent work</td>
<td>Return to work and retention rates after parental leave, by gender</td>
</tr>
<tr>
<td>Product responsibility</td>
<td>Total number of incidents of non-compliance with regulations and voluntary codes concerning health and safety impacts of products and services during their life cycle, by type of outcomes</td>
</tr>
<tr>
<td>Society</td>
<td>Percentage of operations with implemented local community engagement, impact assessments, and development programs</td>
</tr>
</tbody>
</table>
number of activities) are being reported on the most: environment, labor, economic, and society. For the companies analyzed, activities pertaining to environmental far outnumbered any other type of sustainability effort. This is hardly surprising given that the term “sustainability” emerged from ecology and the study of natural systems connoting the ability to perpetuate existence in the natural environment. The other most numerously reported activities are in dimensions that are reflective of corporate social responsibility more than sustainability, *per se,* and are dimensions that have persisted since the CSR of the 1970s (Rockness and Williams, 1988).

Of the 476 activities reported, there is a range among the companies of 36 in the total number reported: Dresser Rand reported just 13, while Intel reported 49 (Intel, Coca Cola, LG Electronics, Volkswagen, Toshiba and Caterpillar reported the most activities for sustainability). This reflects the still discretionary, unstructured, and variable nature of sustainability reporting. Unlike reporting of economic activities, highly structured by custom and the regulatory apparati that exist, what and how social and environmental activities are reported (qualitatively or quantitatively) is at the discretion of management. Lacking the structure of financial reporting, there is neither consensus on what is germane to sustainability’s aim nor consensus on how the results of common sustainability initiatives should be reported. Furthermore, there is no information that provides links to the financial reporting bottom line. The two “sets” of information are not integrated.

The 476 activities are raw data that reflect only the unique way a particular company described those activities. We noted that although descriptions of activities varied, what is described is often the same. To use a financial reporting analogy a company might report credit sales as “accounts receivable” while another might do so as “due from customers.” So for the 476 separate sustainability activities, we reconciled dissimilar descriptions of similar activities from multiple companies into single activities following a careful item development process suggested by Nunnally (1978). For example, one report stated that “the company continually improved its safety practices, processes and performance,” while another company stated that they “performed daily safety walkthrough in facilities to identify and correct any problems.” We identified these two descriptions as being of the same sustainability activity of improving employee safety, which came under the dimension...
of “labor practices and decent work.” We judged from their descriptions that they are similar enough that we could combine them into a single activity. Through this reconciliation process we identified that there are 145 unique sustainability activities embedded in the 476 descriptions contained in the 15 reports. As a means to validate these categorizations, we engaged the services of an expert judge, an internationally known expert in sustainability reporting, who assessed the content validity of the list. Edits were made based on this expert’s feedback. This ensured we had adequately tapped into the content area. The list of unique activities was rephrased to become the initial survey items we utilized to address our research objective.

The results of the reconciliation analysis are presented in Table IV. The four most prominent dimensions in terms of unique activities reflect the same dimensions that have persisted over time as central to CSR. The largest number of unique sustainability activities reported belongs to the society dimension followed by economic, labor, and environment. Although the reports tend to emphasize the environmental dimension, interestingly, there were fewer unique activities in this dimension indicating there is more common recognition across companies about what constitutes relevant responsible environmental behavior[15]. The society dimension had significantly more activities associated with corporate sustainability indicating companies believe they can support sustainability in more ways when considering society than they do when considering the environment. Thus, there is less agreement among companies about what represents appropriate activities for corporations to fulfill their responsibilities as members of society than as environmental stewards. This is not particularly surprising since what constitutes good environmental stewardship can be discerned via natural science research, i.e. there is some scientific basis to deciding good and bad environmental stewardship. What constitutes a good society is much less amenable to scientific direction and depends more on moral discernments, which the “merchant caste” (Priestland, 2013) that epitomizes corporate values may not accept.

The differences in number between the most unique activities (society) and the least (human rights) are likely attributable to the interactions of diversity of the industries, the importance given to each dimension by companies, and the consensus about common activities clearly identified with each dimension. “Society” has the most unique activities because all companies want to be seen as good for society, yet each company has many idiosyncrasies including industry, diversity of societies in which it operates, management’s political/moral philosophy about what a good society means, etc. On the other hand, not all companies imagine they have a proactive role to play in improving the state of human rights in the world or that their activities can have any significant effect on the issue of human rights around the world[16].

We then reviewed all activities in each dimension to see if there were any commonalities among types of activities reported. The following list shows results of our analysis. For each dimension, the most commonly mentioned activities are provided along with a

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Number of unique activities reported</th>
</tr>
</thead>
<tbody>
<tr>
<td>Society</td>
<td>35</td>
</tr>
<tr>
<td>Economic</td>
<td>27</td>
</tr>
<tr>
<td>Labor</td>
<td>25</td>
</tr>
<tr>
<td>Environment</td>
<td>23</td>
</tr>
<tr>
<td>Risk and compliance</td>
<td>14</td>
</tr>
<tr>
<td>Product responsibility</td>
<td>13</td>
</tr>
<tr>
<td>Human rights</td>
<td>8</td>
</tr>
<tr>
<td>Total sustainability ideas analyzed</td>
<td>145</td>
</tr>
</tbody>
</table>

Table IV. Content analysis of sustainability activities after reconciliation.
representative example of how an activity was reported. For example, within the “Labor” dimension, activities contributing to employee health were the most prevalent. Activities creating career opportunities and developing corporate culture were also reported across the 15 companies. Main activities in each dimension and specific company examples represents the types of activities associated with each dimension considered most frequently by the companies for reporting on success in each dimension. Thus, we have a view of the narrative structure of corporate activities provided by the GRI and the companies themselves.

Main activities in each dimension and specific company examples are as follows.

Society dimension:

1. Educational opportunities:
   - Haworth: “Provided educational programs for underprivileged children, tuition reimbursement, and scholarships.”

2. Employee volunteering:
   - Intel “Engaging the employees in meaningful volunteer experiences positively impacts their satisfaction and pride, and help us attract and retain talented people.”

3. Supporting charitable causes:
   - Caterpillar “Give generously [...] supporting educational and environmental causes, health and human services, culture and art and civic and community foundation.”

4. Women and minorities:
   - Walmart “Helped nearly 8,000 women gain access to meaningful employment and financial security in the USA.”

Economic dimension:

5. Supporting businesses:
   - Intel: “Intel’s non-US sites significantly impact the economies of countries around the world, e.g., in Costa Rica jobs created and local industries supported.”

6. Supporting/Seeking sustainable ideas:
   - Coke: “Looking for new innovation and ideas and learn from the sustainability successes from colleagues in Norway and Sweden.”

Product responsibility dimension:

7. Environmentally friendly (fair) products:
   - LG: “Awarded sustainable product certifications and several other “green awards” for our products.”

8. Recyclable materials:
   - Haworth: “Choose products based upon their recyclability or recycled content, have PVC-free products, and carbon neutral products.”

Environment dimension:

9. Environmental policies:
   - Toshiba: “Environmental management is guided by 3 key concepts: greening of process, greening or products, and greening by technology.”
Renewable energy sources:
- Abbott: “Use wind-solar hybrid technology to run our plants.”

Goal setting and measurement:
- Intel: “Reducing PFC emissions 45% in absolute terms and over 80% on a per chip basis.”

Human rights dimension:

Policies and training:
- Toshiba: “Provided human rights education for new recruits and employees at the at the time of promotion and management positions; also had lectures on harassment.”

Diversity:
- Fifth third: “Have a diversity council that regularly meet with senior management to enact diversity and discuss initiatives.”

Labor dimension:

Enhanced healthcare:
- Intel: “Created a portfolio of health benefit plans and wellness programs designed to encourage employees to evaluate, improve, and maintain their health and the health of their families.”

Career development:
- Abbott: “Launched a rigorous training program to improve quality assurance practices at operational sites.”

Corporate culture:
- Mohawk: “Believe in open, direct and constructive dialogue between workers and management to reach mutually agreeable solutions in workplace issues.”

Risk dimension:

Business risk assessment:
- Mohawk: “Audit their overseas suppliers before signing any contractual agreements.”

Employee corruption risk:
- Volkswagen: “Have an international hotline of lawyers to answer questions from all employees.”

Sustainability message received: report users perceptions of sustainability activities
From the 145 unique activities we developed the instrument used in our factor analysis of stakeholder perceptions of sustainability activities. Our survey development approach follows the methodology of Churchill (1979) and Straub (1989). Once we distilled the 476 descriptions of sustainability activities to 145 descriptions of unique activities, our next step was to engage a focus group to evaluate the 145 items for clarity and to validate our categorizations. This group consisted of 28 MBA students familiar with corporate sustainability and TBL reporting. Each of the 28 people independently read the 145 statements and categorized each one as belonging to a single GRI sustainability
dimension, as belonging to more than one GRI dimension, or not applicable to any of the GRI sustainability dimensions. The result of this process was an array of classifications of 145 items into GRI dimensions. In each cell of the array was the number of focus group members who placed an item in a particular dimension. Given there were 28 evaluators, any item count of 14 or above in any GRI dimension meant that at least 50 percent of the evaluators agreed that the item belonged there. That is, a tally of $\geq 14$ indicated a plurality agreed an activity belonged in only one particular GRI dimension.

Applying a majority decision rule left 71 unique activities classified uniquely in one GRI dimension to include in our preliminary survey. The survey instrument consisted of questions in the form of probe statements that required Likert-scale responses, 1 = not at all important to 5 = very important, in order to answer the following question: “How important do you think these activities are to an organization achieving corporate sustainability?”

A list of the 71 activities included in the survey is reproduced in the Appendix.

We then administered a pilot of the survey to 25 respondents. The result of the pilot indicated no additional revisions were required in the wording of the statements. We then administered the survey to an additional $n = 94$ respondents. Demographic data from our respondents show that we have good variability with regard to economic characteristics. Our respondents’ income levels range from less than $10,000 to over $150,000 with the median range being between $70,000 and $79,999[17]. Based on exploratory factor analysis and reliability heuristics (Nunnally, 1978), 45 activities emerged as the most important. We did a comparison between our content analysis data and our survey data. Specifically, we compared the stakeholder views of the activities to how the companies reported these same activities. This told us if companies are reporting the same information that stakeholders view as most important and whether companies’ sustainability narratives are consistent with the stakeholders’ sustainability narratives.

**Factor analysis**

Results of the factor analysis and scree plot are very interesting, especially when compared to our findings from the content analysis of sustainability reports. Six factors emerged from the data (Table V) and the exploratory factor analysis revealed 45 of our 71 survey items to be the most important and reliable in terms of survey methodology. The breakdown of these 45 survey items is shown in Table V.

After reviewing the items in the six factors that emerged from our analysis, we were able to name the construct they described, which is different than the GRI framework, the most recognized reporting framework used by companies and the one we used to categorize the activities from our content analysis (Table VI). A new factor of “risk and compliance” emerged among the activities. Another factor that emerged included items that were in various categories in the GRI, but held together during exploratory factor analysis into a factor we identify as “community building.” These community building activities were in several of the dimensions of the GRI, but what is important is that our sample of respondents viewed community activities as more important than societal actions that would not directly affect them. The GRI includes one dimension for “Labor and Decent Work” with various suggested activities, but our sample companies reported other labor-related activities. Exploratory factor analysis suggests that our sample viewed “labor” as two dimensions rather than one which we named “employment opportunities” and “employee information/education,” with more respondents being concerned about societal labor activities that give employees opportunities. What is interesting is that Factors 7-10, which did not have high factor loadings (i.e. factor loadings $< 0.50$), include activities that are primarily in the GRI dimensions of “product responsibility” and “economic.” So, what our preliminary data revealed is that US companies filing GRI reports are emphasizing economic and product responsibility
<table>
<thead>
<tr>
<th>Factor</th>
<th>Item</th>
<th>GRI dimension</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Measure and disclose climate change strategies (e.g. carbon disclosure project)</td>
<td>Environmental</td>
</tr>
<tr>
<td></td>
<td>Work with locals to reforest land that improves rainwater infiltration and reduces erosion</td>
<td>Society</td>
</tr>
<tr>
<td></td>
<td>Deploy technology that captures and reclaims greenhouse gas emissions ordinarily vented into the atmosphere</td>
<td>Environmental</td>
</tr>
<tr>
<td></td>
<td>Increase recycling rate</td>
<td>Environmental</td>
</tr>
<tr>
<td></td>
<td>Incorporate green design standards and building concepts into the construction of facilities</td>
<td>Environmental</td>
</tr>
<tr>
<td></td>
<td>Utilize wind power in plants</td>
<td>Environmental</td>
</tr>
<tr>
<td></td>
<td>Save tons of paper and plastic through packaging and shipment redesigns, thus reducing landfill waste</td>
<td>Environmental</td>
</tr>
<tr>
<td></td>
<td>Use innovations that utilize an agricultural input to reduce independence on petroleum</td>
<td>Economic</td>
</tr>
<tr>
<td></td>
<td>Establish a take back program in order to recycle used products and comply with the European Union regulations</td>
<td>Product</td>
</tr>
<tr>
<td></td>
<td>Minimize climate impact by reducing emissions, increasing efficiency, and changing the method of sourcing and using energy</td>
<td>Responsibility</td>
</tr>
<tr>
<td></td>
<td>Commit to purchasing renewable energy credits</td>
<td>Environmental</td>
</tr>
<tr>
<td></td>
<td>Utilize solar power in facilities</td>
<td>Environmental</td>
</tr>
<tr>
<td></td>
<td>Strive to make products “green” and eco-friendly</td>
<td>Product</td>
</tr>
<tr>
<td></td>
<td>Develop effective processes to prohibit the uncontrolled release of pollutants (e.g. wastewater, sulfur dioxide, nitrous oxide)</td>
<td>Environmental</td>
</tr>
<tr>
<td></td>
<td>Work to sustain biodiversity (conserve the flora and fauna) in the areas of operation</td>
<td>Environmental</td>
</tr>
<tr>
<td></td>
<td>Reduce power consumption for operating activities (e.g. air conditioning)</td>
<td>Environmental</td>
</tr>
<tr>
<td></td>
<td>Create products or services using hybrid technology</td>
<td>Environmental</td>
</tr>
<tr>
<td></td>
<td>Conserve water consumption in company facilities worldwide</td>
<td>Environmental</td>
</tr>
<tr>
<td></td>
<td>Implement a rigorous training program to improve quality assurance practices at operational sites for employees and contractors</td>
<td>Labor</td>
</tr>
<tr>
<td></td>
<td>Corporate business groups monitor performance (including training, management tone, risk assessment, and more) on a quarterly basis and send the results to the Ethics and Compliance Program Office</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>Have a process that informs senior management and the board about misconduct issues including periodic reports of overall misconduct statistics as well as communication of details about key investigations in progress and those completed</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>Provide educational programs for employees to mitigate risks</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>Complete a comprehensive risk assessment review with a recognized industry standard setter</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>Assess strategic business risks annually and disclose in the Form 10-K</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>Offer free and reduced priced medicines, nutrition and diabetes products to US-based patients with financial difficulties</td>
<td>Human rights</td>
</tr>
<tr>
<td></td>
<td>Donate to charitable causes that support organizations across the globe</td>
<td>Society</td>
</tr>
<tr>
<td></td>
<td>Have as a company mission to work with all people to assess their goals and then put resources and services in place to help them build better tomorrows</td>
<td>Society</td>
</tr>
<tr>
<td></td>
<td>Have a healthy living strategy that invests in community-based sports programs</td>
<td>Labor</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table V.
Factor results – labels, activities loading, and GRI where companies classified the activity
per the GRI guidelines, but our study respondents did not think these activities were important in their view of a corporation’s sustainability.

The factors that emerged from our respondents also closely resemble the principles of the United Nations Global Compact that should be implemented within any organization’s scope and mission (UNGP, 2016). This is particularly the case in the correspondence between the respondents’ factor “human rights/equality/fairness” and UNGC’s principles on human rights and the correspondence between “risk and compliance” and UNGC’s principle of anti-corruption. The factors that emerged from our respondents closely, if not more closely, resemble the UNGC principles than the GRI dimensions.

Because of their inherent dissimilarities, the sustainability content analyses data and the survey data could not be quantitatively compared. We, therefore, applied a qualitative
comparative approach to address the question “Is there alignment between what companies report regarding sustainability and what stakeholders view as important?”

As previously noted, researchers coded statements into categories (see Table III). The final data set of coded statements enabled us to count the frequencies of each category. An implication we can draw from analyzing all the texts is that the more often a type of statement is made, the more emphasis, or importance, the organization is placing upon it. Since the activity serves as a signal to the stakeholder, it should relay information that is of importance to them. The category with the most frequently made statements was environment. The most important activities for the respondent sample are recycling, reducing climate impact, and being “green” – all types of activities now associated directly with sustainability. It is reasonable, therefore, to interpret that statements in these reports are placing highest importance on informing stakeholders about environmental activities. It is important to note that during the reconciliation process, we discovered that some reports mentioned a given activity more than once within that report.

The dimension with the next highest number of activities revealed across the 15 reports was labor with 78. Although labor was second in importance for the sample companies, the subjects did not consider it as important nor did they see labor as distinctly a single dimension as did the companies. Labor-related activities were least important to the respondents and were seen to be in two distinct dimensions. One factor related to the company providing a setting for individuals to develop and progress; the other factor is related to informing employees so that they are better prepared to manage their own risks as employees.

The third highest category mentioned in the reports was economic. However, it did not emerge as an identifiable dimension for the respondents. The economic aspect of performance for the respondents pertained to community building, i.e., economic development rather than economic growth per se for the company. The fourth most significant dimension for the respondents is related to compliance and mitigating risks. Respondents perceive it to be very important that companies “behave themselves” and not be cavalier about externalities and other risk shifting behaviors.

What receivers regard as important clearly reflects their different priorities, i.e., a difference in the perceived corporate role in society. In contrast to the areas emphasized in the sustainability reports, the survey data showed that stakeholders are most concerned with (from Table VI): human rights/equality/fairness; environmental stewardship; and employment opportunities. The information emphasized in the reports is not what our respondents feel is most important to achieving sustainability as they perceive it.

**Summary and conclusions**

The results of our analyses of external users’ assessments of activities associated with sustainability, indicates some significant differences between how companies categorize and relate their sustainability activities and how stakeholders perceive those activities. There is consonance between companies and stakeholders in terms of both the importance and the nature of environmental sustainability activities. The environment is important and is defined by activities that preserve natural systems. Beyond this category, however, respondents use different dimensions to classify activities than the GRI structure and place different emphases on them. Beyond the environmental dimension, stakeholders see corporate sustainability in dimensions associated with being a “good citizen,” i.e., community building; compliant behavior avoiding externalizing risks; fairness and respect for human rights; and providing employment that encourages employee development. From a stakeholder perspective, corporate sustainability may be an issue of corporate “virtue” rather than one conceived as merely a constrained economic value maximization problem.

Our research does not reveal anything about what activities companies should report, since we utilized as our treatments only those activities that companies actually report.
What our research indicates is that the narrative that coherently ties sustainability activities together is different for stakeholders than it is for companies. We do find that given the set of reported activities, readers classify the activities differently than does the GRI, and the relative importance of activities does not coincide with frequency of appearance in the reports. Our results add credence to Milne’s (2013, p. 143) contention “[…] that acting alone, voluntarily and on the basis of economic motives, most businesses seem incapable of addressing the fundamental issues of sustainability.” This is so because the narrative of sustainability acted out by companies is not always consonant with what those outside the company want to know about its actions. Information systems, such as GRI reporting, are the architecture of the narrative that ties a company’s disparate activities together into a coherent story about a “good” company, a “mediocre” company, or a “bad” company. Indeed, there are likely numerous general narratives about sustainability depending on which particular stakeholder is constructing the narrative[18]. What our results do indicates that the GRI guidelines, largely reflecting a corporate management view of sustainability, would be a premature foreclosure of our understanding about what the potentially disparate narratives about sustainability are, and how sustainability activities are interconnected.

Our results also relate to understanding the moral significance of sustainability reporting. Scanlon (2008) considers the problem of “double effect,” the problem of whether the moral assessment of an action really depends on the actor’s intentions. Scanlon argues that intentions do not enter into the assessment since permissibility depends only on principles as used in a critical sense (rather than a deliberative sense) to determine whether an act is permissible. So whether an act is permissible does not depend on whether the actor committed the action by employing the relevant principles to deliberate upon whether the action was permissible. However, assessing the meaning of an action does depend on the actor’s reasons for acting[19].

The subjects of this study indicate that the sustainability actions communicated to them are indicative of a company that is a good citizen, not merely a company being an economic actor doing the right things re sustainability, but also as a moral actor being a good citizen. What our subjects’ narrative about corporate sustainability implies are stakeholders have a moral relationship to companies, as well as whatever economic ones they might have. This moral relationship is the deciding factor for investors engaged in socially responsible investing. According to Scanlon morality is a normative relationship “[…] that specifies attitudes and expectations that we should have whenever certain conditions are fulfilled” (Scanlon, 2008, p. 139)[20].

The expectation that a collective agent, like a business corporation, should possess the necessary attitudes or feelings that make a moral relationship between stakeholders and itself as indicated in the anthropomorphic way that our subjects viewed corporate sustainability actions seems misplaced. Sustainability reported via the GRI structure entertains only that such right actions were performed; the reasons for performing them are not important. However, for the subjects in our study the narrative constructed out of those actions seemed to align them in a way that the reasons for doing those actions did matter, i.e. they are done to be a good citizen. Meaning of the actions was important to the subjects and not simply those actions permissibility as actions contributing to corporate sustainability.

We conclude that in order for sustainability reporting to correspond to the way it is perceived by those outside the firm, including socially responsible investors, the information in the reports should regularly also correspond to information managers utilize to make decisions about the business. Just as financial reports are representative of the same kinds of information that enter into the decisions that managers make, i.e., financial outcomes are ends to be achieved, so do our subjects seem to believe that sustainability information should be about ends as well. Sustainability reporting should result in sustainability thinking and action on the part of management.
1. This definition sounds innocuous enough, but is radical in its implications. It presumes a systems perspective because it speaks of generations of humans (populations) and it focuses on needs, not wants. Given that a substantial proportion of the current generation lacks fulfillment of basic needs, the definition encompasses alterations not just in systems of production but also of distribution.

2. The latest scholarly term that may replace “sustainability” is the concept of “resilience,” which is the ability to plan for and survive adverse environmental events (Carlson, 2013).

3. Major providers of guidance for sustainability reporting are the GRI, the Carbon Disclosure Project, Organization for Economic Cooperation and Development (OECD), The United Nations Global Compact, and the International Organization for Standardization. The reports should feature disclosures around multiple issues as mentioned previously including environmental, economic, and society.

4. For 2014, 4,169 companies from around the world, including 557 from North America, listed reports with the GRI, 79 percent of which followed the GRI guidelines. The majority of all reports were “self-declared,” or not checked by any third party, even though the GRI “recommends” that companies seek external assurance. As of September 22, 2015, 4,974 out of 28,334 reports (18 percent) in the database were non-GRI-based reports.

5. CSR or sustainability is intimately related to the much more fundamental question of what should be the role of the corporation in society, which is ultimately a matter of morality, politics, and law (Danley, 1994).

6. Recently, some Harvard Business School graduates have initiated a movement to resurrect the idea of managers as professionals. The movement aims to persuade business schools to take The MBA Oath (Anderson and Escher, 2010). The oath consists of eight promises the manager vows to keep, one of which has direct bearing on the subject of this paper: “I will strive to create sustainable economic, social, and environmental prosperity worldwide (Anderson and Escher, 2010, p. xvi).”

7. For example, KPMG (2011) utilizes the logic of business success to advocate for corporate responsibility reporting: “[…] clearly, CSR reporting is now an essential requirement for any company hoping to be seen as a responsible corporate citizen. Innovation and learning, in particular, has consistently ranked highly as a driver for corporate social responsibility reporting over the past decade. This is indicative of the large number of companies that see CR as a means (emphasis added) to drive greater innovation through their business and products in order to create a discernible competitive advantage in the market” (p. 7).

8. Gambling (1974) proposed that societal accounting for corporations be focused on economic measurements of these externalities so that they could be explicitly integrated into the bottom line, thus making a TBL unnecessary. “Net externalities” also reveals the “system” nature of sustainability, i.e., the extent to which a single firm’s actions are inextricably linked to a wider system connecting companies, people, and the natural environment. This realization was emphatically noted by Milne (2013, p. 145) “[…] sustainability only really makes sense at a systems level, and perhaps more pertinently at a planetary systems level.”

9. We propose this model only as a conceptual device for situating social and environmental values within the context of the traditional structured economic narrative. We accept Gray’s (2013, p. 459) assertion that sustainability reporting “[…] would not start with the somewhat bizarre and tortured foundations of conventional financial accounting.”

10. There is a paradox in this expression that seems irresolvable at the firm level. The paradox lies in the conflicting values implied on either side of the equation. On the right hand side, the production side, the value is on efficiency, while on the left hand side, the consumption side, the corporation, in order to grow, must implicitly advocate less efficiency on the consumption side. This bipolarity exists for every individual: as producers we should be as efficient as possible, but as customers we cannot apply the same logic since economic growth depends on ever increasing consumption (70 percent of US GDP depends on consumption of consumer goods).
11. One of the leading sustainability reporting scholars has expressed pessimism about the prospects for a coherent system of representing corporate performance: “It would seem, that on the available evidence at least, we are unable to derive any generalisable inference about the value of social disclosure as a signal of social performance” (Gray, 2006, p. 78). In addition, there is the problem of power. Shareholders have power exercised through securities markets to leverage corporations to act in certain ways based on the information they receive via financial reports. No such equivalent power exists to leverage corporate behavior based on sustainability reporting. Government environmental and social regulations act as constraints, but do not provide the power necessary to make environmental and social performance the primary aims of the corporation.

12. We did not include any activities that were deemed to contribute to “unsustainability.” Companies tend not to report negative events (Boiral, 2013) so there is a bias in reports toward only those activities deemed to cast the firm in a favorable light. Given the nature of our research objective, this bias is what we want since we are focused on activities that are deemed to result in a corporation being responsible and sustainable by doing the right things.

13. The GRI is still a rather crude system for organizing “good actions” that confirm a company is a responsible, sustainable company.

14. Since the many scandals like Enron and WorldCom, corporate governance has emerged as an important consideration in assessing corporate conduct along with the new discipline commonly referred to as “enterprise risk management.”

15. This may also reflect the fact that there are many large, well-financed and organized NGOs who have provided direction on what the issues on the environmental agenda should be. According to Dauvergne and Lister (2013), many of these groups have moved from antagonism toward corporations to a more conciliatory strategy of working with them to encourage more environmentally responsible behavior by educating them on the benefits.

16. Ironically, there exists a document that specifies 30 explicit human rights. “The Universal Declaration of Human Rights” was adopted by the UN General Assembly on December 10, 1948. Few of the universal human rights are mentioned as actionable by the companies in our sample. The only rights implied by the activities described are health (worker safety), nondiscrimination (diversity), and fair treatment of workers. The activities taken under the rubric of human rights were paternalistic and means-centered. No companies acknowledged an employee right for desirable work and to join a trade union, a right to social security, a right to rest and leisure, or a right to an adequate living standard.

17. The resulting data set had four missing values. In order to perform statistical analysis, we replaced the missing values with the item mean, an approach known as mean imputation, which is a common approach to handling missing data.

18. Our results lend empirical support to Owen et al.’s (1997, p. 180) observation that “[…] the social account opens up new categories of meaning for the organization and its stakeholders and moves us further from the idea of an organization as purely an economic entity […]”

19. Scanlon (2008) explains that he equates his reference to the meaning of an action as like Kant’s notion of the moral worth of an action: Kant says that for an action to have moral worth it is not enough that it be “in accord with duty” (i.e. permissible); it must also be “done from duty. So what Kant calls moral worth is like what I am calling meaning, in its dependence on the reasons for which an action is performed” (p. 101).

20. According to Scanlon (2008, p. 141), blameworthiness is attributable to individuals who commit actions that permit others “[…] to have attitudes toward them different from those that constitute the default moral relationship.”

References


Belkaoui, A. (1984), Socio-Economic Accounting, Quorum Books, Westport, CN.


Freeman, R.E. (1984), Strategic Management: A Stakeholder Approach, Pitman, Boston, MA.
McCumber, J. (2005), Reshaping Reason, Indiana University Press, Bloomington, IN.


Further reading


Appendix. Corporate sustainability actions*

(1) Have a sustainability-oriented growth strategy based upon appealing to customers in all parts of the world.
(2) Support programs that reduce starvation.
(3) Give employees career development opportunities to expand their skills by rotational, temporary, or sabbatical coverage assignments.
(4) Add comprehensive information in labeling all of products.
(5) Report and disclose serious accidents related to products.
(6) Have a process that informs senior management and the board about misconduct issues including periodic reports of overall misconduct statistics as well as communication of details about key investigations in progress and those completed.
(7) Have as a company mission to work with all people to assess their goals and then put resources and services in place to help them build better tomorrows.
(8) Use life cycle analysis to identify and reduce carbon footprint of products.
(9) Offer a global promotion and reward system to attract top candidates and respond to increasing global mobility.
(10) Offer opportunities for people within the company to develop ideas through competitions and programs they develop.
(11) Support and obey laws that prohibit discrimination everywhere in places the company does business.
(12) Conserve water consumption in company facilities worldwide.
(13) Develop effective processes to prohibit the uncontrolled release of pollutants (e.g. wastewater, sulfur dioxide, and nitrous oxide).
(14) Save tons of paper and plastic through packaging and shipment redesigns, thus reducing landfill waste.
(15) Invest worldwide in remanufacturing joint ventures with other organizations to expand sustainability outreach or capabilities.
(16) Establish a take back program in order to recycle used products and comply with the European Union regulations.
(17) Strive to become the number one manufacturer in our industry, both economically and sustainably in the next five years.
(18) Mentor professionals, support startups and build R&D opportunities in developing countries.
(19) Provide educational programs for employees to mitigate risks.
(20) Sell Fair Trade products.
(21) Reduce pesticide use.
(22) Publish white papers examining the economic impact of broadband internet connectivity in developing nations, strategic information and communication technology spending and e-learning environments.
(23) Actively investigate regulatory and safety-related issues when necessary to promptly remove products from all avenues.
(24) Obtain third-party attestation of sustainability measures.
(25) Realize throughput improvement, cycle time reduction and improvement in on-time delivery because of manufacturing improvements at our facility.
(26) Commit to purchasing renewable energy credits.
(27) Strive to become an employer of choice and measure progress through positive employee response rates in surveys.
(28) Have a high-quality, comprehensive medical examination and advice session available to all employees.
(29) Provide employees with information on how to protect and secure personal data.
(30) Assess strategic business risks annually and disclose in the Form 10-K.
(31) Deploy technology that captures and reclaims greenhouse gas emissions ordinarily vented into the atmosphere.
(33) Actively promote gender equality and diversity in the workplace.
(34) Use responsible sales and marketing to carry clear information to the consumer.
(35) Corporate business groups monitor performance (including training, management tone, risk assessment, and more) on a quarterly basis and send the results to the Ethics and Compliance Program Office.
(36) Consider it a business imperative that the corporate presence contributes to the social and economic development of local communities and host nations.
(37) Minimize climate impact by reducing emissions, increasing efficiency, and changing the method of sourcing and using energy.
(38) Create products or services using hybrid technology.
(39) Ensure all customers are treated fairly and consistently.
(40) Have a transparent pay structure from the bottom level to the top level.
(41) Commit to source key products from sustainable sources.
(42) Acknowledge and respect the diversity that exists among social customs and cultural traditions in the countries of operation.
(43) Commit to continuing our success in global markets and create more opportunities for global investors to benefit from long-term growth plans.
(44) Utilize solar power in facilities.
(45) Offer leadership training to identify future business leaders in the company.
(46) Work with regulatory authorities to promote the eradication of counterfeit items and to protect brands.
(47) Increase recycling rate.
(48) Provide human rights education for new recruits and employees at the at the time of promotion and management positions.
(49) Implement a rigorous training program to improve quality assurance practices at operational sites for employees and contractors.
(50) Incorporate green design standards and building concepts into the construction of facilities.
(51) Prohibit the use of child, forced or involuntary labor.
(52) Perform self-audits to ensure compliance with company-wide standards.
(53) Create a development corporation to invest in low-income housing, community revitalization, and minority efforts.
Work with locals to reforest land that improves rainwater infiltration and reduces erosion.

Pursue business excellence and profit in a climate of free enterprise, free trade, and unencumbered competition.

Strive to make products “green” and eco-friendly.

Seek to build relationships with women and minority-owned businesses.

Offer free and reduced priced medicines, nutrition and diabetes products to US-based patients with financial difficulties.

Utilize wind power in plants.

Promote policies that reduce or eliminate trade and investment barriers.

Measure and disclose climate change strategies (e.g. carbon disclosure project).

Provide an adaptation program when sending employees overseas.

Routinely audit suppliers to ensure they meet expectations for quality and social responsibility standards.

Donate to charitable causes that support organizations across the globe.

Reduce power consumption for operating activities (e.g. air conditioning).

Have a healthy living strategy that invests in community-based sports programs.

Take sustainability issues into consideration when negotiating mergers and acquisitions.

Work to sustain biodiversity (conserve the flora and fauna) in the areas of operation.

Complete a comprehensive risk assessment review with a recognized industry standard setter.

Collaborate with government, local vendors, and policy makers to support educational initiatives and activities.

Use innovations that utilize an agricultural input to reduce independence on petroleum.

*A copy of the complete survey instrument is available from the corresponding author.

Corresponding author
Paul F. Williams can be contacted at: pfwms@mcsu.edu
Leveling the playing field for less-sophisticated non-professional investors

Does plain English matter?

Eileen Taylor
Poole College of Management, North Carolina State University, Raleigh, North Carolina, USA, and
Jennifer Riley
University of Nebraska – Omaha, College of Business Administration, Omaha, Nebraska, USA

Abstract

Purpose – The purpose of this paper is to explore how non-professional investors (NPIs) with varying levels of financial sophistication interpret and perceive corporate disclosures and management credibility, specifically risk factors, when those disclosures are presented in readable and less-readable formats.

Design/methodology/approach – The paper uses an online experiment to test hypotheses related to the effects of financial sophistication (measured) and readability (manipulated) on NPIs’ equity valuations and perceptions of management credibility (competence and trustworthiness).

Findings – Increased readability appears to counteract less-sophisticated NPIs’ conservatism in equity valuations, such that they are not statistically significantly different from more-sophisticated NPIs’ equity valuations. Further, less-sophisticated NPIs judge management as less competent when disclosures are less readable, while more-sophisticated NPIs judge management as more competent when disclosures are less readable.

Research limitations/implications – The paper has important implications for the SEC’s regulations related to plain English requirements for risk factor and other corporate disclosures. Financial sophistication varies among NPIs, and readability appears to influence these individuals in different ways.

Practical implications – The SEC’s Concept Release (April 13, 2016) acknowledges the need to update and improve risk factor disclosure regulations. This study provides evidence that contributes to those decisions.

Originality/value – The paper extends the research on processing fluency, by examining readability of disclosures with a consistent tone (negative). The NPIs surveyed are directly representative of the population of interest for risk factor disclosure regulations.

Keywords Risk factor, Processing fluency, Corporate disclosures, Regulation S-K, Non-professional investor

Paper type Research paper

Introduction

With the increase in online trading and wider stock market access for non-professional investors (NPIs), in the USA, and especially globally, it is of interest to examine how current reporting standards and practices may differentially influence subsets of this group. In the USA, the SEC has offered guidance on ways to improve disclosures by making them more timely and understandable; i.e., investors must be able to read and understand corporate disclosures in order to make informed decisions and manage their risk (Linsley and Shrives, 2005). Research finds that NPI underperformance in the market...
largely arises from their inability to understand financial information (e.g. Barber and Odean, 2000; Odean, 1999). Coram (2010) further proposed that a lack of task-specific knowledge by NPIs results in conservatism in equity valuation and investment behavior.

In 1998, the SEC helped to increase the understandability of financial information by passing the Plain English Rule 421(d), accompanied by the “Plain English Handbook” (Securities and Exchange Commission, 1998), which provides guidance to registrants in writing certain disclosures using plain English principles. Such principles include short sentences; definite, concrete, and everyday words; active voice; tables or bullet point lists; no legal jargon and highly technical business terms; and no multiple negatives (Securities and Exchange Commission, 1998).

The rationale for increasing readability through plain English seems clear and logical. Plain English disclosures should improve readability by presenting information akin to individuals’ normal processing style (Securities and Exchange Commission, 1998). This “processing fluency” represents the ease with which individuals process external information (Winkielman et al., 2003). Such disciplines as psychology, communications, and marketing have studied fluency (for a review see Alter and Oppenheimer, 2009), and consistently find that fluency influences judgments over and above content (Schwarz et al., 1991). Disclosure information presented in plain English should be more readable and more fluent, and thus be easier for NPIs to process.

Research indicates, however, that financial information has become too disfluent and complex for NPIs to process (Miller, 2010; Li, 2008; Courtis, 1995, 1998, 2004; Linsley and Lawrence, 2007; Jones and Shoemaker, 1994). Additionally, registrants may use lower readability to hide information or confuse investors. Some studies find evidence that managers use less-readable disclosures to obfuscate negative and declining performance in communications to investors such as the MD&A (Courtis, 2004; Subramanian et al., 1993; Li, 2008; Rutherford, 2003; Jones and Shoemaker, 1994). Other managers may go so far as to decrease readability to deceive investors into believing that performance is better than the numbers show (Lee et al., 2013).

While it seems logical that plain English disclosures would be more useful to investors than less-readable or other complex language disclosures, this is far from proven. Questions remain about whether plain English is always better for communicating information; particularly in communicating corporate information to investors, researchers have questioned the effects of requiring plain English on investors’ understanding, perceptions, and decision making (Cui, 2016; Tan et al., 2014, 2015; Rennekamp, 2012). Overall, these studies suggest that we need more research to understand the varied effects of readability on understanding, perceptions, and decision making, and that the relationship between readability and investor behavior is more complex than previously assumed.

This paper explores the effects of readability in the context of risk disclosures, which became part of the required SEC registrants’ disclosures since 2005. This regulation requires registrants to disclose risks “[...] that may adversely affect the issuer’s business, operations, industry, or financial position, or its future financial performance” (Securities and Exchange Commission, 2004, 2005). Because the SEC intended this disclosure to help all investors, including NPIs, it falls under the above plain English principles. This study is especially timely in light of the SEC’s Concept Release on April 15, 2016 that seeks public comment on suggested changes to business and financial disclosures required by regulation S-K, and presents the results of “a comprehensive evaluation of the Commission’s disclosure requirements recommended in the staff’s report on Review of Disclosure Requirements in Regulation S-K, which was mandated by Section 108 of the Jumpstart Our Business Startups Act (JOBS’ Act)” (Securities and Exchange Commission, 2016, p. 9).

Using survey responses from 359 NPIs, we examine how readability and task-specific knowledge (i.e. financial sophistication) affect judgments about equity valuation, ease of understanding, and management credibility (i.e. competence and trustworthiness).
Overall, findings indicate that investors with low financial sophistication (measured using a composite index similar to Tan et al., 2014) consistently provided lower valuation estimates than those with high financial sophistication, regardless of readability. However, all investors are influenced by readability and provide higher valuation estimates when readability is high. No evidence finds plain English significantly increases less-sophisticated NPIs’ understanding over more-sophisticated NPIs. Although understanding does increase for all investors under plain English, it is more pronounced in the more-sophisticated NPI group. Financial sophistication and readability interact to influence perceptions of management credibility. Low-readability results in decreased perceptions of credibility among less financially sophisticated investors and increased perceptions among more financially sophisticated investors. This finding implies that less financially sophisticated investors may be skeptical of managers who use difficult-to-read language, but financially sophisticated investors may actually reward managers who use complex language.

Importantly, we find an apparent similarity of decisions and perceptions between less-sophisticated NPIs presented with plain English disclosures and high sophisticated NPIs presented with less-readable disclosures. On most measures, both groups exhibited nearly identical behaviors, implying that increasing the readability of disclosures does help compensate for a lack of task-specific knowledge, and “levels the playing field” between less- and more-sophisticated NPIs. Of course, corporations will not produce disclosures in two different formats, and since plain English did appear to impact more-sophisticated NPIs as well, more research is needed to determine whether increasing readability is beneficial to the NPI population overall.

Similarly, to the studies by Rennekamp (2012) and Cui (2016), we conclude that readability is one part of a complex formula for investor judgments, and that plain English requirements are not a panacea for improving corporate disclosure usefulness. Readability differences have inconsistent effects across a variety of disclosures studied in prior research, and across the NPI population. Focusing on plain English may divert attention from other challenges to the SEC’s goal to provide useful information to the investing public. The results contribute to the growing body of research on the style and presentation of disclosures, and the subsequent effect on investors’ behavior. Unlike the study by Miller (2010), which examined the entire 10-K, this study isolates risk disclosures, thereby limiting the potential influence of other information and information overload. This study adds to Rennekamp’s (2012) findings that are based on a single press release regarding earnings, by exploring the effects of sophistication on non-financial disclosure judgments. It also complements Tan et al.’s (2014) findings by holding language sentiment constant. The findings indicate that readability affects NPIs regardless of management’s choice of positive or neutral language. Further, this study contributes to this literature because our subject pool comprises members of NPIs’ education and support association. Thus, the conclusions are based upon individual respondents who are the target of the SEC’s initiatives on plain English disclosures.

These results should be of interest to public company registrants, the SEC, and other investor protection agencies (e.g. Financial Industry Regulatory Authority), NPIs and their supporting organizations (e.g. BetterInvesting), and standard-setters. Further, because US investors have access to many resources, such as analyst reports, financial media, etc., it may be even more important for regulatory bodies in countries other than the USA to be aware of our findings. Overall, this area will benefit from additional research, and theory development and testing, to understand and predict how written language influences investment decisions.

Background and hypotheses

Task-specific knowledge (sophistication)

In concert with the substantial increase of NPIs in the market (Coram, 2010), much research has examined their use of financial information to make investment decisions. Such research
has indicated that the internet-fueled increase in access to valuable financial information has not seen improved investment behavior. NPIs systematically misinterpret the financial information available (Odean, 1999), such that they underperform the market by 3.7 percent annually (Barber and Odean, 2000). This research indicates that a lack of financial knowledge and an inability to understand financial information contribute to NPIs’ investment underperformance. In keeping with prior literature, we use the term sophistication to refer to the level of investor’s task-specific knowledge, namely, their familiarity with investing and financial analysis.

Coram (2010) proposed that a lack of sophistication results in conservatism in equity valuation and investment behavior because the likelihood of unintended effects (i.e. conservatism) of cognitive impairment is negatively related to such knowledge (Smith and Kida, 1991). When compared to professional investors who are sophisticated, NPIs make more conservative judgments when reading positive non-financial disclosures (Coram, 2010).

Thus, sophistication impacts NPIs’ investment behavior vs professional investors; it may also influence behavior within the NPI group. NPIs are not a homogenous group (Pinsker and Wheeler, 2009), and some NPIs are likely to be more financially sophisticated than are others. Sophistication is a contributing factor to investment behavior because it enables an investor to understand and appropriately use the value-relevant information (Victoravich, 2010; Frederickson and Miller, 2004). However, because it is not possible, nor fair to exclude NPIs from the capital markets based on their lack of sophistication, it is important to determine if any interventions can compensate for this shortcoming. Readability may be a solution.

Readability
Encouraging readability in financial reporting has been a consistent theme in the SEC’s ongoing effort to assist “least-sophisticated investors” (Securities and Exchange Commission, 1998). In discussing the usefulness of financial reporting, “Probably the most familiar theme is plain language. The swamp of legalese found in many annual reports and mutual fund prospectuses can frustrate even the most experienced investor” (Securities and Exchange Commission, 2007). Research supports the sentiment that financial reports have become too complex for small investors to understand and process. Jones and Shoemaker (1994) concluded that the reading level of nearly all corporate annual reports is difficult or very difficult, requiring an undergraduate or graduate degree for comprehension. Several subsequent readability studies (Courtis, 1995, 1998, 2004; Linsley and Lawrence, 2007; Smith and Taffler, 1995) agree that annual reports are “inaccessible to a large proportion of private lay shareholders” (Jones and Shoemaker, 1994). Most recently, Li (2008) labeled 10-Ks as “unreadable” and found that readability has significantly decreased from 1993-2004 (p. 226). Miller (2010) supported Li’s (2008) conclusions. Using the FOG Readability Index (Gunning, 1952), Li reported a mean readability level for a sample of 2006 10-Ks as 19.4, much greater than the level considered difficult for most people to read, which ranges from 14 to 18, and certainly greater than the ideal, which ranges from 12 to 14. Miller (2010) also generated a multidimensional measure of readability using the SEC’s plain English principles. This measure indicates a mean readability level of 20.795 for the same sample and year, again representing an unreadable level for most.

We also find negative effects of complexity and readability on investor decisions that the archival literature documented. For example, You and Zhang (2009) found investors under-react to 10-K information, which is greater for firms with more complex reports. Miller (2010) found that more complex 10-Ks result in lower trading activity by, and less consensus among, small investors, concluding that complex disclosures may be too costly for NPIs to process.
Plain English disclosures should improve readability by presenting information consistent with individuals’ normal processing style (Securities and Exchange Commission, 1998). This “processing fluency” represents the ease with which individuals can process external information (Winkielman et al., 2003). Disciplines such as psychology, communications, and marketing have studied fluency (for a review see e.g. Alter and Oppenheimer, 2009), which is shown to impact judgments beyond content (Schwarz et al., 1991). Disclosure information presented in plain English should be more readable, or more fluent, and thus should be easier for less-sophisticated NPIs to process.

However, despite regulators, legislators, and professional groups’ contentions on plain English’s positive effects, research evidence is inconsistent. Some studies suggest that plain English results in greater comprehension of bank contracts (Campbell, 1999), insurance policies (Blazzard and Hasenauer, 1998), and mutual fund prospectuses (Johnson, 2004). Others, however, caution that plain English initiatives may not always result in the intended outcome. Bernstein (2006) reported that some earlier SEC attempts to make 10-Ks more readable to average investors caused companies to simply eliminate disclosures. Jacoby et al. (1982) found that the Federal Trade Commission’s attempts to improve communications by using plain English actually resulted in greater confusion in a wide range of audiences.

Recent research in accounting raises even more questions about the effects of plain English disclosures. Remnekamp (2012) found small investors susceptible to overreaction, responding more positively to good news and more negatively to bad news when press releases are more readable. This finding is based on a homogenous pool of NPIs, and is therefore unclear if the effect holds across all NPIs (i.e. those with both high and low sophistication).

Recent readability and sophistication studies find both variables affecting investor decisions and behavior. Tan et al. (2014) showed that readability effects may reverse when considered jointly with other variables such as positive or negative language sentiment. Tan et al. (2015) found that inconsistencies in benchmarks affect the response of investors to strategically readable management disclosures. Cui (2016) also examined inconsistencies, and found that sophistication and readability interact.

While these studies examine readability and NPI’s sophistication levels in the context of some inconsistent message and/or tone, and use either MBA students (Cui, 2016; Tan et al., 2014, 2015) or Amazon MTurk participants, the present study differs. We examine readability and sophistication in a consistent message (presentation of risk factors, a negative disclosure), and the participants are all active non-professional US nationwide investors who have earned varying educational degrees ranging from high school to PhDs and MDs. Thus, this adds generalizability to Tan et al.’s (2014, 2015) and Cui’s (2016) findings. Tan et al. (2014) noted that their sample may be unrepresentative of the continuum of investor sophistication, and suggested future research on this topic. These points do not criticize Tan et al. (2014, 2015) or Cui (2016), but rather point to divergence to indicate the complementary aspects of this work.

Based on prior research, a more conservative bias should prevail for less-sophisticated NPIs, if they cannot fully process a risk disclosure. Increasing the readability of disclosures should increase understanding and reduce conservatism. H1a tests whether less-sophisticated NPIs are more conservative than more-sophisticated NPIs. H1b tests the effect of readability on conservatism. If readability of risk factors reduces conservatism (increases equity valuations) for less-sophisticated NPIs, then the SEC’s focus on readability is justified. Formally:

H1a. Less-sophisticated NPIs will estimate lower equity valuations than more-sophisticated NPIs.

H1b. Less-sophisticated NPIs will estimate lower equity valuations when a disclosure is less readable than when it is more readable.
If support suggests that less-sophisticated investors are influenced by readability, before whole-heartedly recommending plain English across the board, we must also explore readability’s effects on more-sophisticated investors. Does readability similarly affect both high- and low-sophistication NPIs? High-sophistication NPIs should understand disclosures regardless of readability, which should result in little difference in their equity valuations. When considered with the above hypotheses, this would indicate an ordinal interaction: the more-sophisticated investors’ valuations forming a straight line between readability conditions and the less-sophisticated forming a diagonal line increasing from low to high readability (approaching, but remaining below the more-sophisticated investors). However, an important factor affecting the use of disclosures is the perception of message or source credibility (Pinsk er and Wheeler, 2009; Mercer, 2004). NPIs’ perceptions of credibility may be influenced by assurance (Pinsk er and Wheeler, 2009), labeling (Koonce et al., 2005), and other factors (Mercer, 2004), including readability. Lower source credibility may negatively influence equity valuations, introducing complexity into the judgment.

On this relationship, Rennekamp (2012) found a positive effect of readability on management credibility indirectly through increased processing fluency. But again this result relies on studying NPIs as a homogenous group. Tan et al. (2014) reported that more-sophisticated NPIs may view less-readable disclosures as management’s attempt to obfuscate information, and that they see through management’s strategic intentions when a stimulus is difficult to read by punishing management with lower equity valuations. Supplemental analysis indicates that this effect arises partially from lower perceptions of credibility when positive language is difficult to read. This conforms to psychology research that perceived bias from a source reduces the message’s credibility (Birnbaum and Stegner, 1979). The management obfuscation hypothesis proposes that managers have an incentive to make annual reports difficult to read to impair investors from incorporating adverse information into stock prices (Bloomfield, 2002). Thus, more-sophisticated NPIs may be more skeptical of complex wording regardless of whether the message is positive or adverse. They may then react by lowering their opinion of managers who use difficult-to-read language, thus resulting in a lower equity valuation. The effect of readability on equity valuations for this more-sophisticated group of NPIs is proposed in $H_2$ below:

$H_2$. More-sophisticated NPIs will estimate lower equity valuations when a disclosure is less readable than when it is more readable.

While $H_{1a}$ and $H_2$ predict lower equity valuations for all NPIs using less-readable disclosures, we propose there are different reasons at the different investor sophistication levels. To test directly these proposals, we examine whether the hypothesized effects arise from less-sophisticated NPIs’ increased understanding of plain English disclosures or more-sophisticated NPIs’ reduced perceptions of management credibility using less-readable disclosures. Prior research implies that it may be both. Plain English should compensate for the lack of task-specific knowledge (sophistication) expected to contribute to less-sophisticated NPIs’ conservatism[1], leading to $H_3$:

$H_3$. Less-sophisticated NPIs will perceive lower understanding when disclosure readability is lower than when readability is higher.

At the same time, low readability should decrease perceptions of management credibility, but only for more-sophisticated NPIs. Stated formally:

$H_4$. More-sophisticated NPIs will perceive lower management credibility when disclosure readability is lower than when readability is higher.
**Method**

**Participants**

Study participants consisted of 359 NPIs. Participants completed an online survey distributed by BetterInvesting Inc., an organization that provides educational webinars, web-based mutual funds and stock tools, in-person learning events, and similar activities for NPIs. The sample contains 207 (58 percent) males and 152 (42 percent) females. Average years investing is 23.38 years, average trades per year is 10.23, and average age is 65.03. Although this average age appears high, it is common for research samples of NPIs (Coram, 2010; Elliott et al., 2008). Of the 44,000 people who received the e-mail survey invitation, 1,279 people started the survey, for an initial response rate of 3 percent. We eliminated incomplete attempts, resulting in 359 complete, usable responses. Full demographic information appears in Table I.

**Design**

A $2 \times 2$ (disclosure readability by investor sophistication level) between-participants design tests the hypotheses. Participants are randomly assigned to either more- or less-readable treatments. Investor sophistication level is measured through demographic information. Tan et al. (2014) created a sophistication index using the number of accounting and finance courses, and the frequency of reading annual reports and earnings releases to determine delineation of sophistication of participants. Following similar logic, the sophistication index in the present study uses the presence of an accounting or finance degree and the frequency of buying and selling stock over the last year.

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Range</td>
</tr>
<tr>
<td>Number of years investing</td>
<td>358</td>
<td>23.38 (13.09)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0-62 years</td>
</tr>
<tr>
<td>Average no. of trades per year</td>
<td>359</td>
<td>10.23 (14.25)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0-120</td>
</tr>
<tr>
<td>Average age</td>
<td>359</td>
<td>65.03 (10.06)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>25-90</td>
</tr>
</tbody>
</table>

Gender = 58% Male (207/359); 42% Female (152/359)

Certified public accountant or certified financial analyst = 4% (16/359)

<table>
<thead>
<tr>
<th>Highest degree earned</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>High school</td>
<td>5</td>
</tr>
<tr>
<td>Some college/associate’s degree</td>
<td>42</td>
</tr>
<tr>
<td>Bachelor’s degree</td>
<td>141</td>
</tr>
<tr>
<td>Graduate degree</td>
<td>171</td>
</tr>
<tr>
<td>Total</td>
<td>359</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Field of highest degree</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accounting and finance</td>
<td>36</td>
</tr>
<tr>
<td>Other business</td>
<td>64</td>
</tr>
<tr>
<td>Healthcare</td>
<td>50</td>
</tr>
<tr>
<td>Engineering</td>
<td>39</td>
</tr>
<tr>
<td>Education</td>
<td>36</td>
</tr>
<tr>
<td>Physical sciences</td>
<td>31</td>
</tr>
<tr>
<td>Other college degree</td>
<td>98</td>
</tr>
<tr>
<td>High school diploma</td>
<td>5</td>
</tr>
<tr>
<td>Total</td>
<td>359</td>
</tr>
</tbody>
</table>

Table I. Demographic data
Task

Participants received identical financial statements and background information for a fictional company (see Appendix 1), followed by the disclosure that is the subject of the readability manipulation (see Appendix 2). The experiment investigates the readability and sophistication issue in the context of one specific type of required disclosure: risk factors. Regulation S-K, implemented by the SEC in 2005, requires that companies list and discuss “[…] the most significant factors that may adversely affect the issuer’s business, operations, industry, financial position, or its future financial performance” within their 10-K filing (Securities and Exchange Commission, 2004, 2005). The regulation represents the latest in the SEC’s ongoing goal of providing full and fair risk disclosure: a focus that began in earnest in 1994 after massive derivative losses surprised investors of companies such as Proctor & Gamble (Linsmeier and Pearson, 1997). We examine risk disclosure readability because the SEC specifically required companies to use plain English for these disclosures. Related research uses earnings’ press releases (e.g. Tan et al., 2014, 2015; Rennekamp, 2012) that have no plain English requirements, and additionally, are quantitative. Thus, this study complements this research by examining a qualitative disclosure that the SEC currently mandates be written in plain English.

To develop the disclosure, we reviewed risk factors from the 10-Ks of a sample of 90 Fortune 500 companies. A review of item 1a disclosures (2004-2008) shows that companies report a range of risks (Blaskovich et al., 2012). Some are simple and common, while others are complex and company specific. Examples of common risks include lost revenues from competition, regulation, and unforeseen natural disasters. Less commonly mentioned risks include inability to produce accurate financial reports due to control weaknesses, forced repayment due to debt covenant violations, and risk of declining financial performance due to changes in accounting rules. The frequency of other risks falls somewhere between such as risk of litigation due to computer hacking or security lapses.

To simulate reality, disclosures in three diverse areas were created: sales, data security, and control over financial reporting[2]. Investor reactions to risk are partially based on familiarity (Song and Schwarz, 2009), warranting a diverse set of factors along this continuum. The first risk factor describes a common risk of sales revenue declines, due to a competitive environment and potentially negative effects of changing consumer preferences. The second risk factor discusses the moderately common risk of data security breaches, related to the increasing practice of collecting and storing sensitive client data. The third risk factor addresses the uncommon risk that an internal control weakness may compromise the reliability of the financial reports. We varied risk presentation order to test and control for potential order effects.

To manipulate readability, we chose the text of a risk factor disclosure from an actual 10-K, which served as the low-readability condition[3]. The SEC’s “Plain English Handbook” (Securities and Exchange Commission, 1998) served as a guide for revision of each risk factor into a high-readability condition, using readability scores from MS Word.

An important consideration is the difference in length that is inherent in a plain English disclosure, which by definition, calls for short sentences and active voice. Since length is related to disfluency (Miller, 2010), prior studies hold length constant across readability (Rennekamp, 2012). However, length was allowed to occur naturally in the plain English disclosures. If firms adhere to the SEC mandate, current disclosures would most likely be shorter and more concise. Thus, these disclosures more closely represent what investors would see in a plain English disclosure. Although the impact of report length is a consideration in investor behavior, we must note that Miller’s findings are based on the length of a full 10-K. This study’s experimental materials are a small subset of information, and the effect of the difference in overall length between the two readability conditions is small. To investigate the effect of length, participants were asked to report their perception of the
length of the disclosures. Although the perception of length between the groups differs significantly, it does not correlate significantly with their equity valuation or their perceptions of credibility (not tabulated). Accordingly, length differences are not believed to affect the hypothesized results.

Two questions assessed the language manipulation’s effectiveness. Differences between both readability conditions are significant in the expected direction for the following questions: the risk factors were easy to read ($t = -5.040, p < 0.000$, two-tailed), and written in plain English ($t = -5.206, p < 0.000$, two-tailed). These results hold when the questions are examined for investor sophistication level, as both less- and more-sophisticated participants successfully indicate more-readable vs less-readable language. Accordingly, the conclusion is that the readability manipulation was successful.

After reviewing the information, participants provided their expected equity valuation and perceptions of management credibility (discussed in more detail in the following section). Finally, standard demographic data were collected.

**Results**

**Sample**

Following the index procedure discussed previously, the sophistication scores of the sampled NPIs range from $-2.56$ to $5.32$. Investors were split into sophistication groups based on the median, resulting in 169 (190) less-sophisticated (more-sophisticated) participants. The mean sophistication scores for more and less sophistication are $-0.66$ and $0.86$, respectively. Less (more) sophisticated investors buy and sell stock an average of $7.97$ ($12.56$) times per year. No significant differences arose between the two groups for mean age or years investing. Less- vs more-sophisticated investors’ mean age is $65.77$ and $64.28$ years, and mean years investing is $23.35$ and $24.01$ years, respectively.

**Equity valuations**

The first two hypotheses test whether the equity valuations of NPIs differ based on their sophistication level; and whether readability influences equity valuations for less-sophisticated investors. These hypotheses were tested by examining the stock price estimates of less-vs more-sophisticated participants who were provided with more-readable vs less-readable risk disclosures. Descriptive statistics appear in Panel A of Table II.

Our two-way ANOVA helped us to determine the effects of sophistication and readability on equity valuations. Results from Panel B of Table II, indicate that the interaction is not significant ($F = 0.75, p = 0.39$). A statistically significant main effect difference in stock price estimates arose for all investors because of sophistication level ($F = 3.89, p = 0.02$), and for all investors because of readability ($F = 5.28, p = 0.05$). As shown in Figure 1, less-sophisticated investors estimate lower stock prices than more-sophisticated investors, and all investors estimate lower equity values when disclosures are less readable. Because of unequal sample sizes, the main effects of sophistication and readability using the estimated marginal means are used for interpretation. The marginal means for equity value are $25.46$ for less-sophisticated NPIs and $25.92$ for more-sophisticated NPIs, a statistically significant difference of $-0.47$ ($p = 0.05$). This supports $H1a$, which predicted that less-sophisticated investors would estimate lower equity prices than more-sophisticated investors. Turning to readability, the marginal means for equity value are $25.42$ for less-readable disclosures and $25.96$ for more-readable disclosures, a statistically significant difference of $-0.54$ ($p = 0.02$). $H1b$ and $H2$ predict lower equity prices when disclosures appear in less-readable vs more-readable language.

These results support the overall expectation that a lack of task-specific knowledge contributes to less-sophisticated NPIs’ conservatism. It also supports the proposal that disclosure readability influences NPIs’ judgments, the SEC’s main target for
plain English requirements. Of critical note in the results is the similarity between less-sophisticated NPIs using plain English disclosures and more-sophisticated NPIs using less-readable disclosures. A t-test performed between these two groups finds no significant difference between their mean equity valuations ($25.83 vs $25.75, respectively, \( t = 0.27, \ p = 0.79 \), not tabulated). This suggests that an increase in
readability does appear to compensate for task-specific knowledge in less-sophisticated investors to the point of creating a “level playing field” with more-sophisticated investors. Considering that the current state of readability in financial information is quite low, this provides support for the SECs continuing effort to encourage plain English.

Understanding and management credibility
The final two hypotheses investigate whether the difference in equity valuations relates to increased understanding (for less-sophisticated NPIs) and decreased management credibility perceptions (for more-sophisticated NPIs). Specifically, less-sophisticated NPIs are expected to report greater understanding when readability is high vs low. More-sophisticated investors are expected to be more skeptical of management when readability is low vs high, as they are more likely to recognize management obfuscation. To measure understanding, we asked participants how difficult it was for them to understand the disclosures. Descriptive statistics of the dependent variable for ease of understanding by treatment condition are presented in Table III, Panel A, and ANOVA results in Panel B. The interaction between sophistication and readability on understanding is not significant ($F = 0.56, p = 0.45$). A significant main effect of readability on all investors ($F = 7.56, p = 0.01$) is found, indicating that NPIs viewed the plain English disclosure as easier to understand, regardless of sophistication level. $H3$ is not supported, as increased readability results in increased ease of understanding for NPIs overall, but not solely for less-sophisticated NPIs. When each sophistication level is examined separately, the more-sophisticated NPIs perceive a difference in understanding from readability. $t$-Tests on the two levels indicate that while not statistically significant for less-sophisticated NPIs ($t$-statistic $= 1.55, p = 0.12$), the difference in ease of understanding is significant for more-sophisticated NPIs ($t$-statistic $= 2.26, p = 0.03$). The more-sophisticated NPIs drive the overall significant effect of readability on ease of understanding. This unexpected result is considered in the “Discussion” section of the paper (Figure 2).

Panel A: descriptive statistics: mean (standard deviation); range

<table>
<thead>
<tr>
<th>Financial sophistication</th>
<th>Readability</th>
<th>Low ($n = 169$)</th>
<th>High ($n = 190$)</th>
<th>Overall ($n = 359$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less financially sophisticated ($n = 182$)</td>
<td>5.17 (1.26)</td>
<td>5.44 (1.11)</td>
<td>5.31 (1.18)</td>
<td></td>
</tr>
<tr>
<td>More financially sophisticated ($n = 177$)</td>
<td>4.88 (1.58)</td>
<td>5.36 (1.18)</td>
<td>5.13 (1.40)</td>
<td></td>
</tr>
<tr>
<td>Overall ($n = 359$)</td>
<td>5.02 (1.43)</td>
<td>5.40 (1.14)</td>
<td>5.22 (1.30)</td>
<td></td>
</tr>
</tbody>
</table>

Panel B: ANOVA tests of between-subjects effects

<table>
<thead>
<tr>
<th>Type III SS</th>
<th>df</th>
<th>MSE</th>
<th>$F$</th>
<th>$p$-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Financial sophistication</td>
<td>2.97</td>
<td>1</td>
<td>2.97</td>
<td>1.79</td>
</tr>
<tr>
<td>Readability</td>
<td>12.52</td>
<td>1</td>
<td>12.52</td>
<td>7.56</td>
</tr>
<tr>
<td>Sophistication $\times$ readability</td>
<td>0.90</td>
<td>1</td>
<td>0.90</td>
<td>0.56</td>
</tr>
<tr>
<td>Error</td>
<td>355</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Panel C: $t$-Tests of ease of understanding

<table>
<thead>
<tr>
<th>Mean difference of low vs high readability</th>
<th>$t$-Statistic</th>
<th>$p$-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less sophisticated</td>
<td>0.27</td>
<td>1.55</td>
</tr>
<tr>
<td>More sophisticated</td>
<td>0.48</td>
<td>2.26</td>
</tr>
</tbody>
</table>

Note: DV = disclosure was easy to understand (1 = strongly disagree, 7 = strongly agree)
A six-item questionnaire assessed perceptions of credibility, which we adapted (presented in Appendix 3) from the studies by McCroskey (1966) and Leathers (1992), and validated in the study by Mercer (2005). Following Mercer, a composite measure of management credibility is calculated, based on trustworthiness and competence, which we evaluated individually below. Descriptive statistics of the composite measure of credibility by treatment condition appear in Table IV, Panel A, and ANOVA results appear in Panel B. A significant interaction arises between sophistication and readability on management credibility ($F = 4.76$, $p = 0.03$). Simple main effects analysis indicates a statistically significant difference in management credibility between less-sophisticated NPIs in the two readability conditions. These results appear in Panel D of Table IV. When readability is low, less-sophisticated NPIs' perceptions of management credibility are significantly lower than when using more-readable disclosures (3.88 vs 4.05, $p = 0.04$). However, more-sophisticated NPIs' perceptions of management are not significantly affected by readability (4.04 low readability vs 3.97 high readability, $p = 0.32$). Also, when readability is low, less-sophisticated NPIs' perceptions of management are significantly lower than more-sophisticated NPIs (3.88 vs 4.04, $p = 0.05$). Taken together, these results imply that it is less-sophisticated NPIs who are more skeptical of management who use low-readability disclosures, not more-sophisticated NPIs. Thus, $H4$ is not supported, and rather, is the opposite of expected for sophistication level (Figure 3).

Per Mercer (2005), the credibility scale contains two factors: competence (the first three questions) and trustworthiness (the last three questions). We examine both components separately. Tables V and VI present descriptive statistics (Panel A) and ANOVA results (Panel B) for respective measures of competence and trustworthiness. While no significant differences in results on the trustworthiness measure are found, results do support an interactive effect of readability and education on competence. As presented in Panel B of Table V, the interaction is significant ($p = 0.04$). Simple effects tests (Panel C) indicate a significant difference in mean competence perceptions by less- vs more-sophisticated NPIs when readability is low (3.68 vs 3.92, $p = 0.04$). Additionally, the lowest perception of
Panel A: descriptive statistics: mean (standard deviation); range

<table>
<thead>
<tr>
<th>Financial sophistication</th>
<th>Readability</th>
<th>Low (n=169)</th>
<th>High (n=190)</th>
<th>Overall (n=359)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less financially sophisticated</td>
<td></td>
<td>3.88 (0.53)</td>
<td>4.05 (0.48)</td>
<td>3.97 (0.51)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.0-5.2</td>
<td>2.3-5.3</td>
<td></td>
</tr>
<tr>
<td>More financially sophisticated</td>
<td></td>
<td>4.04 (0.50)</td>
<td>3.97 (0.57)</td>
<td>4.0 (0.54)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.8-5.5</td>
<td>2.5-5.2</td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td>3.96 (0.52)</td>
<td>4.01 (0.53)</td>
<td>3.98 (0.52)</td>
</tr>
</tbody>
</table>

Panel B: ANOVA tests of between-subjects effects

<table>
<thead>
<tr>
<th>Factor</th>
<th>Type III SS</th>
<th>df</th>
<th>MSE</th>
<th>F</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Financial sophistication</td>
<td>1.1</td>
<td>1</td>
<td>1.17</td>
<td>0.43</td>
<td>0.51</td>
</tr>
<tr>
<td>Readability</td>
<td>0.16</td>
<td>1</td>
<td>0.16</td>
<td>0.61</td>
<td>0.44</td>
</tr>
<tr>
<td>Sophistication × readability</td>
<td>1.29</td>
<td>1</td>
<td>1.29</td>
<td>4.76</td>
<td>0.03</td>
</tr>
<tr>
<td>Error</td>
<td>96.13</td>
<td>355</td>
<td></td>
<td>0.27</td>
<td></td>
</tr>
</tbody>
</table>

Panel C: simple main effects tests

<table>
<thead>
<tr>
<th>Factor</th>
<th>df</th>
<th>MSE</th>
<th>F</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Readability within low sophistication</td>
<td>1</td>
<td>1.20</td>
<td>4.44</td>
<td>0.04</td>
</tr>
<tr>
<td>Readability within high sophistication</td>
<td>1</td>
<td>0.26</td>
<td>0.97</td>
<td>0.32</td>
</tr>
<tr>
<td>Financial sophistication within low readability</td>
<td>1</td>
<td>1.03</td>
<td>3.81</td>
<td>0.05</td>
</tr>
<tr>
<td>Financial sophistication within high readability</td>
<td>1</td>
<td>0.34</td>
<td>1.26</td>
<td>0.27</td>
</tr>
</tbody>
</table>

Table IV.
Non-professional investors’ perceptions of management credibility

Note: DV = composite measure of credibility

Figure 3.
Management credibility by readability and sophistication level

management competence is found when less-sophisticated NPIs experience low-readability disclosures (mean = 3.68) and the highest perception is noted when less-sophisticated NPIs experience high-readability disclosures (mean = 3.97). This difference is significant (p = 0.01). When readability is high, less-sophisticated investors will perceive higher
competence of management than when readability is low, while more-sophisticated investors will perceive lower competence of management when readability is high than when low. These results may result from expectations gaps between more- and less-sophisticated investors when it comes to management competence.

Plain English as a compensating factor
The primary results above indicate the similarity between the equity valuations of less-sophisticated NPIs using plain English disclosures and more-sophisticated NPIs using
less-readable disclosures. Increased readability may compensate for less-sophisticated investors’ lack of task-specific knowledge to help create a “level playing field” with more-sophisticated investors. With the exception of ease of understanding, the same results are found for the other variables examined. The means of less-sophisticated NPIs using high-readability disclosures and more-sophisticated NPIs using low-readability disclosures are nearly identical on overall management credibility (4.05 vs 4.04, respectively), management competence (3.97 vs 3.92, respectively), and management trustworthiness (4.17 vs 4.16, respectively). These results mirror the findings presented earlier on equity valuations, bolstering the proposition that plain English pulls less-sophisticated NPIs to a similar standing as more-sophisticated NPIs.

Discussion and conclusion
Research indicates that financial reports have become too complex for NPIs to grasp (Miller, 2010), requiring at least a graduate degree to understand. A lack of task-specific knowledge contributes to less-sophisticated investors’ conservatism (Coram, 2010). The SEC, which wants to level the playing field for NPIs, has responded by requiring that risk disclosures in the 10-K be written using plain English principles. Some research questions the seemingly intuitive belief that plain English communications are “better,” and can create a level playing field for NPIs, particularly among varying levels of NPI sophistication. This study provides experimental evidence on this issue. This study hypothesizes that less- and more-sophisticated NPIs’ equity assessments, ease of understanding, and perceptions of management credibility differ when risk disclosures are presented in plain English vs less-readable language.

The significant main effect for both sophistication and readability indicates that less-sophisticated NPIs provide lower equity valuations, and that low-readability disclosures result in lower equity valuations. But, a noteworthy result is that less-sophisticated NPIs report significantly higher valuation estimates when disclosures are more readable, to the extent that they make judgments similar to those of more-sophisticated NPIs with less-readable disclosures. This implies that sophistication level is, in fact, an important consideration in the readability of financial reports, and that plain English affects less-sophisticated investors in such a way as to level the playing field – a primary objective of the SEC’s mandate. Three of the four perception measures follow the same pattern. The means of less-sophisticated NPIs using high-readability disclosures and more-sophisticated NPIs using low-readability disclosures are nearly identical on composite management credibility, management competence, and management trustworthiness. Taken together, these results provide further evidence that plain English disclosures compensate for a lack of task-specific knowledge, and increase credibility perceptions of management among less-sophisticated NPIs.

Readability also affects the perceived fluency of disclosures. Surprisingly, this effect is significant for more-sophisticated NPIs only, which is opposite of our expectation. Perhaps less-sophisticated investors do not consciously perceive a difference in understandability afforded by plain English disclosures, or they are less likely to admit a lack of understanding. The lowest perception of fluency was noted by the more-sophisticated NPIs with low-readability disclosures. This may be because these investors are experienced in reading financial information and are thus more likely to recognize the complexity of current financial reporting. They may be more cognizant of the growing chorus of complaints regarding the lack of readability present in financial reporting, and identify “legalese” as such. These explanations are speculative, and encourage additional research to identify the factors that contributed to this result.

In another unexpected result, less-sophisticated NPIs perceive lower management credibility for less-readable disclosures. The opposite effect occurs for more-sophisticated NPIs.
Prior research finds more-sophisticated NPIs should be more sensitive to management obfuscation and to punish management for providing low-readability disclosures. More-sophisticated NPIs, who expect low-readability disclosures, may question the sincerity and authenticity of plain English disclosures. Some recent research suggests that difficult-to-process stimuli can actually increase perceived source credibility because the recipient ascribes greater intelligence and favorability to someone who can use big words and complex linguistics (e.g., Alter et al., 2007). If more-sophisticated NPIs are accustomed to seeing disclosures written in “legalese,” they may question the intelligence of the source and reduce their perceived credibility. This supports Tan et al.’s (2014) suggestion that there are boundaries to expected effects on NPI behavior. They found that NPIs with greater sophistication may punish management for strategically disfluent communications, but evidence presented in the present paper indicates that this may not hold if there is not a perceived strategic intent to obfuscate.

For regulators, the question becomes whether the readability of financial disclosures improves investor judgments, and for which investors. Within the recent SEC concept release, “One commenter suggested that current disclosure is too complicated for the everyday person to read and that it should be less duplicative and more straightforward. Another commenter noted the diversity of the investor community and that the Commission’s mandate is to protect all investors” (Securities and Exchange Commission, 2016, p. 48). These comments indicate the importance of acknowledging investors with varying sophistication levels, and the concept release call for comment on the very issue of whom the disclosures should serve. The SEC has long pushed plain English reporting, to provide a more level playing field for NPIs. Our study indicates that it does help level the playing field for the least-sophisticated NPIs, but it also influences the more-sophisticated NPIs. Because reports will be either plain English or not, the implications and consequences must be examined in detail. More research should help us understand and predict the effects and limitations of disclosure presentation style. Is the effect on perceptions of credibility a positive outcome? Does increased readability reduce conservatism for less-sophisticated investors, and is that outcome truly beneficial to the investor in the long term, or does such conservatism act as a protection for overconfidence? Finally, should the SEC worry about this type of investor at all, or assume that efficient markets will effectively constrain and guide them? The SEC calls attention to this in the concept release by stating, “To the extent some investors rely on market prices to efficiently incorporate all public information, rather than relying on disclosures directly, it could be argued that disclosures should be tailored to those users most likely to actively follow a registrant, transact in the registrant’s securities and set the market price” (Securities and Exchange Commission, 2016, p. 58). Future research should examine these questions to guide regulators’ attempts to provide useful information and improve capital market efficiency.

As for all experimental research, these findings face several limitations. The experimental materials are limited, and NPIs would have access to much more information for making equity valuation decisions. We also make no claims of the accuracy of equity assessments or management credibility, and thus offer no normative solutions. Future research that includes additional disclosures and financial information offers the potential to further examine the effect of information overload and readability on NPIs, in conjunction. Lastly, the respondents likely expended less effort on a hypothetical case than they would in an actual investment situation. A controlled experiment involving compensation for investment choices may increase participants’ engagement. Despite these limitations, the results, obtained from a relatively large sample of active NPIs, provide important information for regulators as well as opportunities for future research.
Notes
1. More-sophisticated NPI's understanding should not be sensitive to readability.
2. Of the 90 companies reviewed, 86 included a factor regarding the competitive environment, 56 included a factor on sensitive client data, and only 15 included a factor on internal control weaknesses.
3. Our own review of risk disclosures from a random sample of 100 companies from the Fortune 500 companies for the period 2005-2008 finds that they are primarily written in what is often called legalese, and lack the readability intended by the SEC mandate. We calculated the Flesch Reading Ease (Flesch, 1948) score for the risk disclosure section of this random sample. The average grade level necessary to comprehend the 10-Ks is 17.0, which is representative of post-graduate study.
4. The differences in mean age and mean years investing are not statistically significant.

References


Song, H. and Schwarz, N. (2009), “If it’s difficult to pronounce, it must be risky”, Psychological Science, Vol. 20 No. 2, pp. 135-142.


Further reading


Appendix 1. Information provided to all conditions

Please review the information for Genyx, a fictitious US company. Assume that you do not own stock in the company yet, but are considering whether to invest in it.

Genyx information appears below in the following order:

- Short summary about the company – provided by management.
- Three basic audited financial statements – provided by management.
- Three company risk factors – provided by management.
- Analyst estimates of earnings per share (EPS) and stock price (publicly available information).

Note: Genyx’s CPA firms have provided unqualified (clean) opinions on the financial statements and on internal control. The opinion letters are not included in these materials to save space.
Summary

Genyx Incorporated was founded in 2003 to provide human resource and payroll services to technology companies. The company grew successfully and went public in 2005 – it is currently traded on the New York Stock Exchange and the share price since going public has ranged from $14.00-$28.18.

Appendix 2. Risk factor disclosure in low vs high readability

Risk factor 1 – less readable

We must continue to differentiate our services from those of our competitors; we operate in an evolving industry that requires us to anticipate changes in client preferences and deliver services that demonstrate value to our clients. We operate in a highly competitive environment and in an industry that is subject to significant ongoing changes from market pressures brought about by client demands,
as well as business consolidations, strategic alliances, legislative reform, and marketing practices. These factors require us to differentiate our services by anticipating changes in client preferences and delivering services that demonstrate value to our clients. Failure to anticipate changes in client preferences and deliver services that demonstrate value to our clients can affect our ability to retain or grow our client base, which can adversely affect our results of operations.

Flesch Reading Ease 24.2.
Flesch-Kincaid Grade Level 16.4.

Risk factor 1 – more readable (plain English)
Our sales and profits depend on our competitors and clients’ behavior. We operate in a constantly changing industry where many different events occur that might hurt our sales and profits. We need to be aware of market trends and respond quickly to our competitors and clients’ behaviors. If we do not keep up with the competition, we could lose clients, sales, and profits.

Flesch Reading Ease 59.2.
Flesch-Kincaid Grade Level 8.9.

Risk factor 2 – less readable
We would be adversely affected if we fail to adequately protect client payroll and other sensitive information. We maintain large amounts of client payroll and other sensitive data about our clients in the ordinary course of our business. Our business therefore depends substantially on our clients’ willingness to entrust us with their human resource related and other sensitive information. Events that negatively affect that trust, including failing to maintain appropriate safeguards to keep sensitive information secure, whether as a result of our action or inaction or that of one of our vendors, could expose us to litigation and other proceedings, fines and/or penalties, any of which could adversely affect our business, operating results or financial condition.

Flesch Reading Ease 21.6.
Flesch-Kincaid Grade Level 17.5.

Risk factor 2 – more readable (plain English)
As part of our daily operations, we receive and retain large amounts of personal information about our clients and their employees. Our clients provide this information to us because they believe we will keep it confidential. If we do not keep their information confidential, we will lose their trust and our reputation might suffer. Additionally, clients could sue us and we could be fined and assessed penalties.

Flesch Reading Ease 53.4.
Flesch-Kincaid Grade Level 9.9.

Risk factor 3 – less readable
The Company’s internal controls may not be sufficient to ensure timely and reliable financial information. As reported under Item 9a of this Form 10-K, the Company’s management completed its assessment of the effectiveness of the Company’s internal control over financial reporting as of December 31, 2009 and based on that assessment, concluded that the Company maintained effective internal control over financial reporting as of December 31, 2009. The Company’s auditor has issued an attestation report on management’s assessment that expresses unqualified opinions on management’s assessment and on the effectiveness of the Company’s internal control over financial reporting. Still, however, the Company’s growth continues to place stress on its internal controls, and there can be no assurance that the Company’s control procedures will continue to be adequate. The effectiveness of the Company’s controls and procedures may be limited by a variety of risks, including, among other things, faulty human judgment, simple errors, omissions and mistakes, collusion of two or more people or inappropriate management override of procedures. If the Company fails to have effective internal controls and procedures for financial reporting in place, it could be unable to provide timely and reliable financial information.

Flesch Reading Ease 2.6.
Flesch-Kincaid Grade Level 20.9.
Risk factor 3 – more readable (plain English)

In order to produce timely and reliable financial reports, we need effective internal controls. We completed a formal assessment of our internal control system and found it to be effective. Our auditors agree. If our existing controls weaken, we may be unable to prepare timely and reliable financial information. This could happen because of our growth, human errors and mistakes, collusion, or management override.

Flesch Reading Ease 35.2.
Flesch-Kincaid Grade Level 11.5.

Appendix 3. Management Credibility Questionnaire (competence and trustworthiness)

(1) I believe that Genyx management is very competent at providing risk factor disclosures.
(2) I believe that Genyx management has little knowledge of the factors involved in providing useful risk factor disclosures.
(3) I believe that few people are as qualified as Genyx management to provide useful risk factor disclosures about Genyx.
(4) I believe that Genyx management is very trustworthy.
(5) I believe that Genyx management is very honest.
(6) I believe that Genyx management may not be truthful in their risk factor disclosures.

Scale = 1 (strongly disagree) to 7 (strongly agreed).

Corresponding author
Jennifer Riley can be contacted at: jenriley@unomaha.edu
The performance of US-based emerging market mutual funds

Halil Kiymaz and Koray D. Simsek
Rollins College – Crummer Graduate School of Business,
Winter Park, Florida, USA

Abstract
Purpose – The purpose of this paper is to examine the performance of US mutual funds that invest primarily in emerging market equities and bonds.

Design/methodology/approach – The study adopts the Morningstar classification of mutual funds and uses the Lipper US Mutual Fund Database through FactSet to obtain monthly returns and various metrics for emerging market equity and bond mutual funds covering the period from January 2000 to May 2017. Several descriptive statistics for these funds are reported as well as various risk-adjusted performance measures. Alphas are computed for different sub-periods using different factor models to mitigate potential biases.

Findings – The results show that diversified emerging market funds generate some significant alphas for their investors during the study period. Emerging market bond funds, on the other hand, do not provide any significant positive alphas; mostly alphas are negative. An analysis of sub-period performance suggests that these funds do not consistently provide excess returns, showing great variations from one period to another.

Originality/value – The emerging market funds provide US investors with an alternative source of exposure for their portfolios. Emerging markets differ from developed markets on a wide range of market and economic characteristics, including size, liquidity, and regulation. This study contributes to the scarce literature on these types of funds and provides a comprehensive performance assessment against various benchmarks during a period that encompasses significant bear and bull markets across the world.

Keywords Emerging markets, Performance evaluation, Mutual funds, Diversified funds, Bond funds

Paper type Research paper

1. Introduction
The growth of emerging market economies has been remarkable over the past 35 years. Their share of the global GDP had risen from 21 percent in 1980 to 36 percent by 2014 (International Monetary Fund, 2016, p. 63). Coupled with this pace, capital flows to emerging markets have boomed. During the 2000-2013 period, these markets received cumulative capital inflows of roughly $10 trillion (Plantier, 2015). About 20 percent of this amount is attributable to portfolio flows, defined as the purchases of emerging market stocks and bonds by non-residents of these countries. Plantier (2015) notes that emerging market stocks and bonds held by foreign investors increased from $1.5 trillion in 2005 to roughly $3.5 trillion by 2013, also reflecting the investment returns.

Because of increased interest in investing in emerging markets, investment companies have created various emerging market mutual funds to meet the needs of investors. The popularity of these investment funds comes from their ability to provide diversification benefits as well as their impressive return performance in the long term. Since operating and economic environment in emerging markets are different from those of developed markets,
investing in emerging markets can provide opportunities for fund managers to obtain excess returns or alpha. On the other hand, the volatility in these markets is higher so the additional risk should be included in the analysis.

The question of whether active fund managers can produce positive alphas has received attention from both practitioners and academicians. The proponents in manager’s ability to generate positive alphas believe alphas represent disequilibrium returns that can exist in complex financial markets. For example, Jarrow (2010) argues that persistent and frequent arbitrage opportunities are much rarer, even in complex markets and therefore positive alphas are more fantasy than fact. Roll (1992) shows such portfolios are suboptimal and risky because they do not belong to the mean-variance frontier. Alexander and Baptista (2010) propose a method to lessen this sub-optimality that involves the objective of selecting a portfolio from the set of portfolios that have minimum tracking error variance. As persistent and frequent arbitrage opportunities are much rarer, even in complex markets, Jarrow (2010) argues that positive alphas are more fantasy than fact and unobservable factors can create false positive alphas.

This study provides an analysis of US-based diversified emerging market funds and emerging market bond funds during January 2000 and May 2017 period. Morningstar© labels mutual funds that primarily invest in emerging market equities as diversified emerging markets. Emerging markets bond mutual funds, on the other hand, predominantly invest in hard-currency-denominated bonds issued by entities in emerging market countries. These two categories of funds provide an alternative to US investors who want to expose their portfolios to these markets. They offer an inexpensive and convenient way of obtaining high returns from emerging markets while diversifying risk. The number of funds has increased sharply during the last two decades. For example, more than 80 percent of the mutual funds in our sample were established after 2000.

To assess the performance of the mutual funds, this study examines the fund managers’ efforts for searching alphas in their portfolios. Erb et al. (1999, 2000) and Harvey (1995) argue that evaluating the performance of emerging capital markets is difficult as market conditions influence the return characteristics of the emerging market significantly. So, it is important to have a data set that would cover various market conditions. The period under study covers significant ups and downs in the financial markets. These include technology bubbles of the USA through 2003, one of the greatest expansions in US markets during 2003 and 2007, and finally global financial crisis of 2007-2009. The sample includes 222 diversified emerging market and 78 emerging market bond funds. The findings show that diversified emerging market fund managers experience limited success in their search for alphas during 2000 and 2017 while emerging market bond fund managers fail to realize positive alphas in general. Most funds do not provide statistically significant alphas.

To the best of our knowledge, this is the first study to comprehensively analyze the performance of both diversified and bond emerging market mutual funds in the USA. The remainder of the paper is organized as follows. A detailed review of the literature is given in the next section. Section 3 describes the data and details of the empirical methodology. Findings of the data analysis are presented and discussed in Section 4 and conclusions are provided in Section 5.

2. Literature review

Academic research on emerging markets has become very popular since various developing economy data sources became available in the mid-1990s. Kearney (2012) provides an overall review of this booming literature, while Atilgan et al. (2015) present a detailed review of the empirical studies on emerging equity markets.

Studies on performance of emerging market equity and bond funds report mixed results. Among earlier studies, Cumby and Glen (1990) examine the performance of a sample of
15 US-based, internationally diversified mutual funds for the 1982-1988 period. Findings suggest that these funds did not over-perform a broad international equity index over the sample period. There is, however, evidence the funds outperformed the US index. Eun et al. (1991) report similar results while analyzing 19 international mutual funds during the 1977-1986 period. The findings show that these funds have allowed US investors to diversify risk internationally. While most funds studied outperformed S&P 500 Index during the study period, they underperformed the MSCI World Index.

Chen and Jang (1994) examine the performance of 15 US-based international mutual funds for the 1980 and 1989 period to identify managers’ selection and timing abilities. Findings show that most of the internationally diversified mutual funds outperformed the domestic stock market index in both selectivity and timing. On the other hand, when these funds were evaluated against World Market Index, there is little evidence of stock selection ability for those fund managers. Findings further show that most international fund managers have more macro-forecasting skills than micro-forecasting skills.

Droms and Walker (1994) find that alphas for international equity funds are not significantly different from 0. Also, they show that investment returns are not related to load status, asset size, expense ratios, and turnover rates. The analysis finds no reward for paying a load fee when investing in mutual funds.

Kao et al. (1998) examine the selectivity and market timing ability of international mutual fund managers using 97 funds during the 1989-1993 period. Findings show that international fund managers are poor market-timers. Managers of certain funds, including those of Pacific, Foreign, and World funds, have good selectivity performance. Further, there is a negative correlation between the selection and timing ability of international fund managers and managers of European funds show weaker performance than those managing other groups.

Lin (2006) examines the performances of Japanese broad-market equity managers. Findings show that these managers outperform index returns during the 1981-2004 period. The findings further suggest that these managers underweight large cap stocks and financials and take less market risk.

Gottesman and Morey (2007) examine the performances of diversified emerging market mutual funds by using 54, 83, and 74 funds for the years 1997, 2000, and 2002, respectively. They examine various fund characteristics including expense ratio, portfolio turnover, and manager tenure on fund performance. Findings show that only expense ratio influences the fund performance, lower expense ratio funds are associated with higher fund performance.

Latif and Kazemi (2007) use a stochastic model to examine US-based international mutual funds during the 1990-2003 period. Sample funds are classified based on regions (such as Europe, Pacific, and World). Results show that global equity markets are well integrated. Fund managers cannot consistently earn excess returns above a buy and hold strategy in US equity market.

Michelson et al. (2008) examine the benefits of investing in emerging markets mutual funds during the 1999-2005 period. The authors show the emerging market funds outperform the MSCI Index and the S&P 500 Index but underperform the emerging market index. They further report a negative relation between emerging market fund returns and turnover, and a positive relation between fund returns and size. Lin et al. (2009) argue that having a global view adds flexibility to asset allocation process as fund managers can shift their investments between US and non-US stocks. With skilled managers, a higher alpha can be achieved without adding more risk.

Huij and Post (2011) are the first to document performance persistence for emerging market equity funds by utilizing the approach pioneered by Carhart (1997). They also compare them to other US mutual funds and find that winner funds in emerging markets generate significant returns net of fees and contribute more to the winner-loser return differential. They find that momentum effects in emerging markets can explain only of this
return spread. Their findings suggest a general pattern of over-performance of emerging market equity funds as compared with US equity funds.

Basu and Huang-Jones (2015) also look at diversified emerging market equity funds from 2000 through 2010. They exclude bond funds but include those that are domiciled outside of the USA. With respect to alpha, their findings are similar to ours and show that most of these funds underperform their benchmarks. They detect performance persistence as in the study of Huij and Post (2011). However, they show that poor performers drive this. Finally, employing a non-linear model, they find no evidence of market timing ability among these funds.

In addition to these studies that investigate the emerging market equity funds, there are a few others exploring the performance of bond funds. Among these, Gallo et al. (1997) examine the monthly returns of 22 US-based international bond mutual funds from 1988 to 1994. They report these funds perform better than the Salomon Brothers Non-US Dollar World Government Bond Index. The excess returns measured with the multi-index models are similar to those measured with the single-index model. The authors find that portfolio consisting of all funds outperform the multi-index benchmark while five of the funds outperform the benchmark individually. When comparing the results of the two models, the authors find the multi-index model is better at explaining returns.

Polwitoon and Tawatunthachai (2008) analyze the 50 emerging market bond funds during the 1996-2005 period and report that these funds outperform both domestic bonds and global bonds funds. Authors argue that these bonds provide international diversification benefits to both US and international bond and equity portfolios. The findings further show that exchange rate risk does not explain the differences in portfolio performance. Country-specific and liquidity risks explain a large portion of variation in performances of these funds.

Overall, the literature on the performances of international equity and bond funds reports mixed results with most studies outlining the benefit of international diversification benefits. This paper provides a comprehensive analysis of both emerging equity and bond fund performance.

3. Data and methodology

3.1 Data

Monthly returns from January 2000 through May 2017 for diversified emerging market funds and emerging market bond funds are obtained from the Lipper US Database on FactSet. The fund classifications are determined based on Morningstar categories with the same names. The initial diversified emerging market fund sample includes 787 funds. We remove funds with more than one share classes and those with fewer than 12 months of data[1]. The net sample of diversified emerging market funds consists of 222 funds. Applying a similar screening procedure to an initial sample of 304 emerging market bond funds resulted in a net sample of 78 bond funds. These steps are also presented in Table I.

<table>
<thead>
<tr>
<th>Diversified emerging market funds</th>
<th>% of total</th>
<th>Emerging markets bond funds</th>
<th>% of total</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial sample of funds</td>
<td>787</td>
<td>72</td>
<td>304</td>
<td>28</td>
</tr>
<tr>
<td>Less: multiple share classes</td>
<td>555</td>
<td>71</td>
<td>224</td>
<td>29</td>
</tr>
<tr>
<td>Less: fewer than 12 observations</td>
<td>10</td>
<td>83</td>
<td>2</td>
<td>17</td>
</tr>
<tr>
<td>Final sample</td>
<td>222</td>
<td>74</td>
<td>78</td>
<td>26</td>
</tr>
</tbody>
</table>

Notes: The table explains how filters are used to obtain the final sample consisting of 222 diversified emerging market funds and 78 emerging markets bond funds. Study period spans from January 2000 to May 2017.
Table II outlines various fund characteristics and portfolio holdings of both diversified emerging market funds and emerging markets bond fund categories. It should be noted that the mean expense ratio for diversified emerging funds (1.22 percent) is higher compared to bond funds (0.86 percent). These values are lower than those in the study of Basu and Huang-Jones (2015), confirming a general trend toward lower cost funds in general. Average diversified equity funds ($1,759) have higher net assets under management (in millions) as compared with bond funds ($703). Given that holdings of non-residents in emerging bond and equity markets are roughly equal (Plantier, 2015), it appears that equity investors have a higher tendency to invest through mutual funds compared to bond investors. Other characteristics of interest include turnover, front load, management tenure, Morningstar overall star ratings, portfolio holdings, and various equity- and bond-specific holding characteristics.

Panel A in Table III provides the descriptive statistics for diversified emerging market funds. The average annualized geometric return for 222 funds during the study period is 4.87 percent, with the best-performing fund at 38.32 percent and worst performer at -4.68 percent. The panel further provides minimum, maximum, and quartiles distribution. The average standard deviation is 17.43 percent for this group. The data distribution is negatively skewed, indicating that the mean is less than the median. This means these funds tend to earn extreme negative returns. The data have positive kurtosis, suggesting a peaked distribution. This would show the portfolio’s returns cluster closer to the mean value than they would if they were normally distributed and are also fatter tailed.

Panel B in Table III reports the descriptive statistics of emerging market bond funds. With respect to the annualized geometric returns, the average value for 78 bond funds is 5.64 percent yearly with the best- and worst-performing funds yielding 13.42 percent and -4.68 percent, respectively. The average standard deviation for this group is 11.92 percent. The data distribution is negatively skewed, indicating that the mean is less than the median. This means these funds tend to earn extreme negative returns. The data have positive kurtosis, suggesting a peaked distribution. This would show the portfolio’s returns cluster closer to the mean value than they would if they were normally distributed and are also fatter tailed.

<table>
<thead>
<tr>
<th>Fund characteristics</th>
<th>Mean</th>
<th>Median</th>
<th>SD</th>
<th>Mean</th>
<th>Median</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turnover (%)</td>
<td>65.17</td>
<td>51.00</td>
<td>56.22</td>
<td>129.00</td>
<td>86.00</td>
<td>167.76</td>
</tr>
<tr>
<td>Expense ratio (%)</td>
<td>1.22</td>
<td>1.25</td>
<td>0.40</td>
<td>0.86</td>
<td>0.90</td>
<td>0.27</td>
</tr>
<tr>
<td>Front load (%)</td>
<td>0.31</td>
<td>0.00</td>
<td>1.27</td>
<td>0.28</td>
<td>0.00</td>
<td>1.04</td>
</tr>
<tr>
<td>Net assets ($ millions)</td>
<td>1,832.9</td>
<td>231.50</td>
<td>6,712.7</td>
<td>720.23</td>
<td>113.00</td>
<td>1,431.2</td>
</tr>
<tr>
<td>Manager tenure</td>
<td>4.56</td>
<td>4.00</td>
<td>3.07</td>
<td>4.70</td>
<td>4.00</td>
<td>3.42</td>
</tr>
<tr>
<td>Morningstar overall star rating</td>
<td>3.23</td>
<td>3.00</td>
<td>1.06</td>
<td>3.11</td>
<td>3.00</td>
<td>1.19</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Portfolio holdings</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Equity holdings (%)</td>
<td>91.04</td>
<td>94.77</td>
<td>13.01</td>
<td>0.10</td>
<td>0.00</td>
<td>0.28</td>
</tr>
<tr>
<td>Bond holdings (%)</td>
<td>1.74</td>
<td>0.00</td>
<td>7.20</td>
<td>88.42</td>
<td>92.25</td>
<td>14.10</td>
</tr>
<tr>
<td>Cash holdings (%)</td>
<td>4.63</td>
<td>2.55</td>
<td>9.09</td>
<td>9.18</td>
<td>4.95</td>
<td>14.49</td>
</tr>
<tr>
<td>Other holdings (%)</td>
<td>2.59</td>
<td>1.87</td>
<td>2.53</td>
<td>2.30</td>
<td>1.72</td>
<td>2.61</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Average holding characteristics</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Price/Earnings ratio</td>
<td>13.35</td>
<td>12.53</td>
<td>3.25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Price/Book ratio</td>
<td>2.12</td>
<td>1.87</td>
<td>0.87</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Market capitalization ($ millions)</td>
<td>17,653</td>
<td>17,982</td>
<td>10,912</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table II: Summary fund characteristics and portfolio holdings. The table provides summary statistics (mean, median, and standard deviation) for various fund characteristics, portfolio holdings, and average holding characteristics specific to each type of funds. Study period spans from January 2000 to May 2017. These statistics are computed as of June 30, 2017, among funds with available data.
and −3.91 percent returns, respectively. The average standard deviation is 7.68 percent. These bond fund returns are even more negatively skewed and leptokurtic compared to the diversified equity mutual funds, implying the presence of extreme losses and heavier tails.

In addition to the historical data on mutual funds, we obtain monthly changes on various total return indices from FactSet to use as benchmarks and to derive the emerging market counterparts for the research factors of Fama and French (1993) and Carhart (1997). Finally, as a proxy for the risk-free rate, we download one-month US Treasury Bill returns from the data library of Kenneth R. French (accessible at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html).

### 3.2 Methodology

As a preliminary step in evaluating emerging market mutual funds against their benchmarks, we compute several risk-adjusted performance measures. First, we calculate the Sharpe ratio as the funds’ average excess return above the risk-free rate divided by the standard deviation of fund returns. More specifically, we use the standard formulation as follows:

\[
\text{Sharpe ratio} = \frac{\bar{R}_t - R_f}{\sigma_t}
\]

where \(\bar{R}_t\) is the annualized average return of the fund, \(R_f\) is the annualized risk-free rate (30-day US Treasury Bill rate), and \(\sigma_t\) is the annualized standard deviation of fund returns. As this calculation does not depend on a proxy for the market benchmark, it may be appropriate for examining the performance of emerging market funds. A higher Sharpe ratio implies outperformance of the fund on a risk-adjusted basis.

Next, we compute the Treynor ratio, which differs from the Sharpe ratio since it uses the beta of the fund with respect to its benchmark as the risk measure instead of the standard deviation of returns (Treynor and Black, 1973). Hence, excess return per unit of systematic
risk is measured by this ratio, which can be formulated as:

\[ \text{Treynor ratio} = \frac{R_i - R_f}{\beta_i} \]  

(2)

where \( \beta_i \) is the beta of the fund with respect to its benchmark.

Other risk-adjusted return measures we look at are the information ratio, Sortino ratio, and the M-squared. Information ratio is defined as the average annualized excess return of the fund above the benchmark divided by its tracking error. Tracking error, which is a statistical similarity measure, is computed as the annualized standard deviation of excess returns. A tracking error of 0 means that fund returns are identical to benchmark returns. Therefore, information ratio measures the performance of a fund as a multiple of its similarity with its benchmark. Sortino ratio focuses on the downside risk with respect to a minimum acceptable rate of return (MAR). In this paper, we assume that MAR is the annualized risk-free rate and therefore the numerator of this measure is identical to Sharpe and Treynor ratios. The denominator, on the other hand, is a semi-deviation type of measure where negative deviations from MAR are considered. Finally, M-squared is a transformed version of the unit-free Sharpe ratio. By multiplying the Sharpe ratio with the annualized benchmark return and adding the risk-free rate, one obtains this measure in units of percent.

In the main part of the empirical analysis, we estimate alphas of emerging market mutual funds using three different factor models. First, we run the capital asset pricing model (CAPM) to obtain the standard alpha in the spirit of Jensen (1968, 1969). This so-called Jensen’s \( \alpha \) is the difference between a fund’s average rate of return and its expected position on the security market line given its market risk level. If a mutual has a positive (negative) Jensen’s \( \alpha \), then it outperforms (underperforms) what the CAPM would predict. We run the following regression:

\[ R_i - R_f = \alpha_i + \beta_i (R_{mt} - R_f) + \epsilon_{it} \]  

(3)

where \( R_i \) is the return on fund \( i \) in month \( t \), \( R_f \) is the one-month risk-free rate of return in month \( t \), \( \alpha_i \) is the Jensen’s \( \alpha \) of fund \( i \), \( \beta_i \) is fund’s systematic risk, and \( R_{mt} \) is the return on benchmark index in month \( t \). As noted earlier, we use 30-day US Treasury Bill rates to proxy risk-free rate. Benchmark market index for diversified emerging market funds is chosen as Morgan Stanley Capital International Emerging Markets Investable Index (MSCI EM IMI). We similarly use Bloomberg Barclays Emerging Markets USD Aggregate as the benchmark index for emerging market bond funds. Jensen’s \( \alpha \), the intercept of this regression, is a measure of the risk-adjusted incremental return obtained by the fund manager. A statistically significant positive (negative) alpha indicates superior (inferior) investment performance for each fund. The \( t \)-test is used to examine the statistical significance of these alphas.

Second, we apply the Fama-French three-factor model of Fama and French (1993) to estimate the fund alphas. This model accounts for the size and value factors and performs better than CAPM in most empirical studies:

\[ R_i - R_f = \alpha_i + \beta_i (R_{mt} - R_f) + \delta_i + \phi_i HML_i + \epsilon_{it} \]  

(4)

where \( SMB_i \) and \( HML_i \) are the size and value factors in month \( t \). MSCI has a large cap and a small cap emerging market index that accounts for roughly 70 and 15 percent, respectively, of the market capitalization of these markets. MSCI EM Large Cap index returns are subtracted from MSCI EM Small Cap index returns to derive SMB factor returns. MSCI also splits the MSCI EM IMI Index as Growth and Value. We subtract the monthly returns of the Growth index from those of the Value index to generate the HML factor for emerging markets.

Finally, we estimate the Carhart (1997) four-factor model, which adds the momentum factor to the Fama-French three-factor model. This model adds one more term to the
regression equation:

\[ R_{it} - R_{ft} = \alpha_i + \beta_i (R_{m, t} - R_{ft}) + \delta_i SMB_t + \phi_i HML_t + \gamma_i MOM_t + \epsilon_{it} \]  

where \( MOM_t \) is the momentum factor in month \( t \). We use the monthly changes in MSCI EM Momentum total return index as a proxy for this factor.

The market conditions influence the return characteristics and performance of emerging market securities dramatically, as pointed out by Erb et al. (1999, 2000). As a robustness check, we divide the entire period into four sub-periods, two involving recessions and significant bear markets, and the other two involving economic expansions and bull markets. We repeat the analyses above for these sub-periods and report our findings. It should also be noted that the analyses explained above are carried out for all funds individually as well as two equal-weighted portfolios, one representing each category of funds.

4. Empirical findings

4.1 Performance of diversified emerging market funds against benchmark indices

Panel A in Table IV reports risk-adjusted return measures for diversified emerging market funds. The mean Sharpe ratio, Jensen’s \( \alpha \), and Treynor ratio are 0.21, 0.03, and 7.58, respectively. The Sharpe ratio ranges from 2.64 to –0.49 while Jensen’s \( \alpha \) has maximum and minimum values of 0.80 and –0.61, respectively. When we compare the performance of emerging market equity funds with various market indices, we find mixed results. Specifically, on average, these funds underperform S&P500 as well as both emerging market indices (MSCI and S&P/IFC) on a risk-adjusted basis. As their average Sharpe ratio (0.21) is lower compared to those of S&P500 (0.22), S&P/IFC (0.28), and MSCI Emerging Market (0.23). More than a quarter (56 out of 222) of these funds have better Sharpe ratios than these indices. Furthermore, the equal-weighted portfolio of these funds achieves a Sharpe ratio of 0.25, placing it only behind the S&P/IFC index.

Table V reports the sub-period analysis of diversified emerging market funds. In the extended US bear market period that runs from March 2000 to March 2003, the average Sharpe ratio for diversified emerging market funds is –0.83 while Sharpe ratios of S&P500, S&P/IFC, and MSCI Emerging Market indices are –0.90, –0.76, and –0.90, respectively. So, these funds seem to be providing US investors a better risk-adjusted performance and beating one of their benchmarks on average. Similar patterns are reported during the significant bear market of August 2007-February 2009, which is sometimes labeled as the Great Recession. Although we observe negative Sharpe ratio for funds as well as all benchmarks, it is better for the average fund as well as the equal-weighted portfolio of funds, compared to the S&P 500. Even though these funds fail to beat their emerging market benchmark indices on a risk-adjusted basis, they can provide US investors with better risk diversification.

During the first of the two expansion periods under consideration, emerging market equity funds outperform all the indices. Specifically, in the April 2003 and July 2007 period, the average Sharpe ratio for diversified emerging market funds was 2.66, while S&P500, MSCI EM IMI, and S&P/IFC have Sharpe ratios of 1.47, 2.51, and 2.62, respectively. We see mixed but mostly negative results during the March 2009-May 2017 period, which is the ongoing long US bull market period. During this era of unusually low volatility, S&P 500 posts a Sharpe ratio of 1.39 and emerging market funds or indices cannot quite match this performance. For example, Sharpe ratio for the diversified emerging market group is 0.17 compared to those of S&P500 (0.20), MSCI EM (0.44), and S&P/IFC (0.28). Only 7 out of 222 funds in our sample can outperform S&P 500 on a risk-adjusted basis.
In sum, diversified emerging market funds appear to present a robust risk-adjusted performance for US investors except for the recent extended US bull market. Even though they have mixed results against their own benchmark indices, these funds outperform S&P 500 during our study period, both on average and as a portfolio.

4.2 Performance of emerging market bond funds against benchmark indices
Panel B in Table IV reports risk-adjusted performance of emerging market bond funds. The average Jensen’s $\alpha$ has a value of $-0.13$ percent, suggesting most funds failed to provide positive alpha for their investors during the study period. The equal-weighted portfolio, on the other hand, has a Jensen’s $\alpha$ of 0.03, although this is statistically insignificant. The average Sharpe ratio is 0.56 with a minimum value of $-0.65$ and a maximum value of 3.07. The emerging market bond funds underperformed both emerging and US bond indices as the Sharpe ratios for Bloomberg Barclays EM USD Aggregate and US Aggregate bond indices are 0.82 and 1.05, respectively. Only 5 of these 78 funds can outperform both indices on a risk-adjusted basis. The equal-weighted portfolio of emerging market bond funds has a Sharpe ratio of 0.82, matching the benchmark but still below the US Aggregate bond index. For the most part, these results are in contrast with the emerging market equity funds in Panel A, which are more likely to outperform their developed market benchmark. It appears that the emerging market bond funds are falling short of their diversification potential for investors in these markets.
Table VI presents a sub-period analysis of emerging market bond fund performance. Average Jensen’s α of individual funds is positive in first two of the four sub-periods under consideration, ranging from 0.10 to 0.33. Worst performance comes during the Great Recession period of 2007-2009 with −0.36. We also provide the average Sharpe ratios for sample funds as well as benchmark indices for each sub-period. Sharpe ratio for diversified emerging market funds during the first sub-period (March 2000 to March 2003) is 0.90 while Sharpe ratios of Bloomberg Barclays EM USD Aggregate and US Aggregate bond indices are 0.57 and 1.94 during the same period. This is a very interesting sub-period as all 12 funds outperform the emerging market benchmark but underperform the US benchmark under consideration. During the second sub-period in which US interest rates have consistently increased, we see a different pattern as emerging bond funds outperform both EM and US bond indices on average and on an equal-weighted portfolio basis. The story reverses during the Great Recession (August 2007-February 2009), the emerging market

### Table V. Diversified emerging market funds: sub-period analysis

<table>
<thead>
<tr>
<th>Panel A: March 2000-March 2003 (n = 39)</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>−14.41</td>
<td>−0.83</td>
<td>−20.38</td>
<td>0.04</td>
</tr>
<tr>
<td>Minimum</td>
<td>−23.41</td>
<td>−1.65</td>
<td>−54.26</td>
<td>−1.28</td>
</tr>
<tr>
<td>Maximum</td>
<td>−2.14</td>
<td>−0.26</td>
<td>−6.50</td>
<td>0.78</td>
</tr>
<tr>
<td>Equal-weighted portfolio</td>
<td>−14.61</td>
<td>−0.87</td>
<td>−20.13</td>
<td>0.05</td>
</tr>
<tr>
<td>MSCI EM IMI</td>
<td>−17.18</td>
<td>−0.90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S&amp;P/IFC Investable</td>
<td>−14.19</td>
<td>−0.76</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S&amp;P 500</td>
<td>−13.10</td>
<td>−0.90</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Panel B: April 2003-July 2007 (n = 57)**

| Mean                                   | 40.55 | 2.66 | 42.49 | 0.10 | −0.50 |
| Minimum                                | 17.99 | 1.49 | 22.78 | −0.52 | −6.46 |
| Maximum                                | 63.43 | 7.15 | 117.03 | 1.49 | 3.79 |
| Equal-weighted portfolio               | 41.16 | 2.50 | 40.08 | 0.04 | −0.51 |
| MSCI EM IMI                            | 42.49 | 2.51 |  |  |  |
| S&P/IFC Investable                     | 44.36 | 2.62 |  |  |  |
| S&P 500                                | 15.32 | 1.47 |  |  |  |

**Panel C: August 2007-February 2009 (n = 70)**

| Mean                                   | −42.21 | −1.33 | −47.75 | −0.46 | −0.51 |
| Minimum                                | −73.05 | −2.57 | −97.07 | −3.31 | −7.03 |
| Maximum                                | −28.06 | −1.03 | −37.39 | 0.46 | 1.01 |
| Equal-weighted portfolio               | −40.16 | −1.28 | −44.72 | −0.41 | −0.30 |
| MSCI EM IMI                            | −38.86 | −1.18 |  |  |  |
| S&P/IFC Investable                     | −38.84 | −1.18 |  |  |  |
| S&P 500                                | −33.46 | −1.79 |  |  |  |

**Panel D: March 2009-May 2017 (n = 222)**

| Mean                                   | 7.09 | 0.42 | 12.44 | 0.04 | −0.98 |
| Minimum                                | −4.68 | −0.57 | −6.73 | −0.55 | −4.90 |
| Maximum                                | 38.32 | 2.81 | 986.58 | 0.80 | 11.71 |
| Equal-weighted portfolio               | 12.06 | 0.63 | 12.85 | 0.05 | −0.04 |
| MSCI EM IMI                            | 12.17 | 0.60 |  |  |  |
| S&P/IFC Investable                     | 12.64 | 0.63 |  |  |  |
| S&P 500                                | 17.98 | 1.39 |  |  |  |

**Notes:** The table provides various performance measures computed for diversified emerging market funds during four distinct sub-periods. The mean, minimum, and maximum are calculated across all funds in this category. Statistics for three benchmarks as well as an equal-weighted portfolio of these funds are also reported for each sub-period.
bond funds underperform both benchmark indices. While the equal-weighted portfolio of emerging market bond funds experiences a Sharpe ratio of −0.66, Bloomberg Barclays EM USD Aggregate and US Aggregate bond indices had Sharpe ratios of −0.44 and 0.71, respectively. During this period of global quantitative easing, it appears that emerging bond funds could not manage to earn positive dollar returns due to unfavorable exchange rate movements. Finally, during the last sub-period, the emerging market bond funds continued to underperform both indices although this was a much better period for the emerging market bond index.

Overall, we note mixed results in emerging market bond fund performance. As with the emerging market equity funds, these funds cannot consistently beat their benchmarks. On the other hand, their dismal performance compared to the US bond index will deter the investors in these markets to utilize these funds for bond investing in emerging markets.
4.3 Performance assessment of emerging market mutual funds with factor models

As explained in Section 3, we measure alphas using three different factor models: CAPM, Fama-French, and Carhart. We first estimate time-series regressions in Equations (3)-(5) for all individual funds in our sample and tabulate the results in Table VII. In Panel A, 118 of 222 diversified emerging market funds (53 percent) have positive CAPM alphas, but only 11 of them are statistically significant (5 percent). In total, 47 percent of these funds experience negative alpha (only 3 percent being statistically significant). On the other hand, the majority of emerging market bond funds do not provide positive alphas. For example, only 21 of 78 funds have positive alpha, and 5 of these are statistically significant at 5 percent level. Most bond funds have negative alphas (57 of 78 funds), and about 29 percent are statistically significant. Overall, the findings show that fund managers’ search for alpha in emerging markets is not successful. Only 5 percent of diversified emerging market funds can provide statistically significant positive excess returns to investors. The findings for bond funds are slightly better as 6 percent of funds experience statistically significant alphas for their investors.

As the number of factors increase (going from CAPM to Fama-French to Carhart), an interesting pattern emerges. While the percentage of funds with positive (and statistically significant) alphas in the diversified emerging market category decreases, the same figure increases for the emerging market bond funds. This pattern can also be detected by looking at the histograms of factor model alphas in Figure 1. The charts on the left correspond to the diversified emerging market funds, and the distribution becomes more skewed to the left from top to bottom (i.e. as the number of factors in the model increases). Since the factors are effectively derived from emerging market equity index returns, it is not surprising to see that additional factors can drive away the alphas, the excess return unexplained by the factors. The charts on the right that correspond to the bond funds become more symmetric (or less skewed to the left) by the addition of multiple factors. In both cases, average adjusted $R^2$ values (not reported here) increase by approximately 0.03, implying that

<table>
<thead>
<tr>
<th>Number of funds with</th>
<th>Diversified emerging market funds ($n = 222$)</th>
<th>Emerging market bond funds ($n = 78$)</th>
<th>Total ($n = 300$)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panels A, B, C</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive alpha</td>
<td>118 (53.2%)</td>
<td>21 (26.9%)</td>
<td>139 (46.3%)</td>
</tr>
<tr>
<td>Positive and significant alpha</td>
<td>11 (5.0%)</td>
<td>5 (6.4%)</td>
<td>16 (5.3%)</td>
</tr>
<tr>
<td>Negative alpha</td>
<td>104 (46.8%)</td>
<td>57 (73.1%)</td>
<td>161 (53.7%)</td>
</tr>
<tr>
<td>Negative and significant alpha</td>
<td>6 (2.7%)</td>
<td>23 (29.5%)</td>
<td>29 (9.7%)</td>
</tr>
<tr>
<td>Average adjusted $R^2$</td>
<td>0.8764</td>
<td>0.8354</td>
<td></td>
</tr>
<tr>
<td>Positive alpha</td>
<td>110 (49.5%)</td>
<td>24 (30.8%)</td>
<td>134 (44.7%)</td>
</tr>
<tr>
<td>Positive and significant alpha</td>
<td>8 (3.6%)</td>
<td>3 (3.8%)</td>
<td>11 (3.7%)</td>
</tr>
<tr>
<td>Negative alpha</td>
<td>112 (50.5%)</td>
<td>54 (69.2%)</td>
<td>166 (55.3%)</td>
</tr>
<tr>
<td>Negative and significant alpha</td>
<td>9 (4.1%)</td>
<td>18 (23.1%)</td>
<td>27 (9.0%)</td>
</tr>
<tr>
<td>Average adjusted $R^2$</td>
<td>0.9009</td>
<td>0.8487</td>
<td></td>
</tr>
<tr>
<td>Positive alpha</td>
<td>100 (45.0%)</td>
<td>28 (35.9%)</td>
<td>128 (42.7%)</td>
</tr>
<tr>
<td>Positive and significant alpha</td>
<td>7 (3.2%)</td>
<td>6 (7.7%)</td>
<td>13 (4.3%)</td>
</tr>
<tr>
<td>Negative alpha</td>
<td>122 (55.0%)</td>
<td>50 (64.1%)</td>
<td>172 (57.3%)</td>
</tr>
<tr>
<td>Negative and significant alpha</td>
<td>11 (5.0%)</td>
<td>15 (19.2%)</td>
<td>26 (8.7%)</td>
</tr>
<tr>
<td>Average adjusted $R^2$</td>
<td>0.9060</td>
<td>0.8661</td>
<td></td>
</tr>
</tbody>
</table>

**Notes:** The table presents the number of individual funds with alphas that are positive, negative, and statistically significantly positive or negative at 5 percent level. Panels A, B, and C report alphas based on CAPM, Fama-French three-factor model, and Carhart four-factor model, respectively. Percentages out of column total are given in parentheses.

Table VII. Alphas for individual emerging market mutual funds
Carhart model works slightly better than CAPM in explaining monthly variation in emerging market mutual fund returns.

Our final analysis is devoted to the performance of the equal-weighted portfolios constructed within the two categories of emerging market mutual funds. Table VIII reports that alphas are insignificant for both portfolios under each of the three models. CAPM works so well for the diversified emerging market portfolio that adjusted $R^2$ barely changes by adding multiple factors. Still, the momentum factor is statistically significant, effectively reducing market beta. For the bond portfolio, CAPM works remarkably well, although with a slightly lower adjusted $R^2$ compared to the other portfolio. What is different is the weak significance of the size factor, which is further weakened by the addition of the momentum factor. In any case, the economic significance of these equity-based factors should be taken with a grain of salt.

Figure 1. Distribution of alphas for individual emerging market mutual funds

Notes: The left half of this figure displays the histograms of estimated alphas for individual diversified emerging market funds using the capital asset pricing model (a), Fama-French three-factor model (b), and Carhart four-factor model (c). The right half of this figure displays the histograms of estimated alphas for individual emerging market bond funds using the capital asset pricing model (d), Fama-French three-factor model (e), and Carhart four-factor model (f).
Overall, apart from a handful minority within both categories, these emerging market mutual funds resemble index funds with a tiny momentum factor tilt. It appears to be a difficult task to justify the fee structure, although the expense ratios have been shrinking over the past decade.

5. Conclusions
This study provides an analysis of US-based diversified emerging market equity and bond funds during January 2000 and May 2017. We examine the fund managers’ efforts for searching alphas in their portfolios. The emerging market funds provide US investors with an alternative factor to expose their portfolios. These markets differ from developed markets on a wide range of market and economic characteristics, including size, liquidity, and regulation. The sample includes 222 diversified emerging market and 78 emerging market bond funds. The findings show that the diversified fund managers achieve limited success in their search for alphas during 2000 and 2017 while emerging market bond fund managers fail to realize positive alphas in general. Most funds do not provide statistically significant alphas. While 5 percent of diversified equity funds provide statistically significant positive alphas to their investors, about 6 percent of emerging market bond funds provide statistically significant positive alphas.

The study further provides evidence on the sub-period performance of these funds. These sub-periods cover significant ups and downs in the financial markets, including technology bubble of the USA, two of the greatest expansions in US markets during 2003-2007 and 2009-present, and finally financial crises of 2007 and 2009. The sub-period results suggest that while diversified emerging market funds experience negative returns during two of the four sub-periods, bond funds experience significant negative returns during the 2007-2009 period. The study concludes that emerging market fund managers are unable to provide positive alphas to their clients on a consistent basis.

These findings have important implications for investors. It appears that investors in developed markets such as the USA can benefit significantly from including the diversified emerging market funds in their portfolios. Unless there is an extended bull market such as the one that has been continuing through 2017, these funds can generate higher risk-adjusted returns compared to the developed market benchmark. On the other hand,
these funds do not appear to be generating alphas above the emerging market risk factors under consideration. This implies that investors can realize the same risk-adjusted performance by investing in emerging market index funds or ETFs. These results are less valid for emerging market bond funds as they seem to have difficulty matching the risk-adjusted performance of developed market bond benchmarks. Investors are most likely aware of this situation. As Plantier (2015) reports, foreign investors allocate equally between emerging market equities and bonds. However, they are approximately half as likely to prefer regulated funds (such as mutual funds) to access the bond markets.

Note
1. We only include funds with at least 12 months of returns during the study period. As with most mutual fund studies, the mutual fund return data are subject to survivorship bias. Elton et al. (2001) show that survivorship bias of Morningstar is minimal. Grinblatt and Titman (1989) also conclude that the survivorship bias is negligible for a sample that includes surviving and non-surviving funds in their study.

References


**Corresponding author**

Koray D. Simsek can be contacted at: ksimsek@rollins.edu
Identifying the major reversals of the BIST-30 index by extreme outliers

Ümit Erol
University of Bahçeşehir, Istanbul, Turkey

Abstract

Purpose – The purpose of this paper is to show that major reversals of an index (specifically BIST-30 index) can be detected uniquely on the date of reversal by checking the extreme outliers in the rate of change series using daily closing prices.

Design/methodology/approach – The extreme outliers are determined by checking if either the rate of change series or the volatility of the rate of change series displays more than two standard deviations on the date of reversal. Furthermore, wavelet analysis is also utilized for this purpose by checking the extreme outlier characteristics of the A1 (approximation level 1) and D3 (detail level 3) wavelet components.

Findings – Paper investigates ten major reversals of BIST-30 index during a five year period. It conclusively shows that all these major reversals are characterized by extreme outliers mentioned above. The paper also checks if these major reversals are unique in the sense of being observed only on the date of reversal but not before. The empirical results confirm the uniqueness. The paper also demonstrates empirically the fact that extreme outliers are associated only with major reversals but not minor ones.

Practical implications – The results are important for fund managers for whom the timely identification of the initial phase of a major bullish or bearish trend is crucial. Such timely identification of the major reversals is also important for the hedging applications since a major issue in the practical implementation of the stock index futures as a hedging instrument is the correct timing of derivatives positions.

Originality/value – To the best of the author’s knowledge; this is the first study dealing with the issue of major reversal identification. This is evidently so for the BIST-30 index and the use of extreme outliers for this purpose is also a novelty in the sense that neither the use of rate of change extremity nor the use of wavelet decomposition for this purpose was addressed before in the international literature.

Keywords BIST-30, Outliers, Wavelet approximation and wavelet details, Volatility, Daubechies

Paper type Research paper

I. Introduction

This paper is about the timely identification of the major reversals of the BIST-30 index which is the leading stock index of the Borsa İstanbul. A major reversal is a change in the direction of the stock prices (indices) from a bullish trend to a bearish trend (or vice versa) which is followed by a substantial rise or fall in the prices. Though it is not easy to differentiate a minor reversal from a major reversal in the financial markets; we may define, as a rule of thumb, a major reversal as a change in the direction of the prices (indices) which is followed by at least a 10 percent increase or decrease in the subsequent days. The importance of early detection of such reversals for the fund managers is evident.

The identification of major reversal points is a major issue in technical analysis with a plethora of different tools (bullish/bearish patterns, moving averages, etc.) used for this purpose. The validity of technical analysis, however, is quite a debatable issue with an inevitable subjective element. Moreover, some of the benchmarks used to identify the
change in the primary trend such as 100-day or 200-day moving averages are quite late in signaling the change.

The academic literature on reasons for propelling the major rallies or crashes is relatively meager but still includes some interesting attempts for the explanation of the aberrant behavior in the financial markets. According to the orthodox view which assumes efficiency of markets; a crash or a major reversal is the result of the announcement of a dramatic and unexpected piece of information. Earlier work which focused on a rational expectations model of bubbles and crashes such as Blanchard (1979) and Blanchard and Watson (1982) admits that the observed prices may deviate significantly from the fundamental prices for extended periods of time but these deviations are still explained in terms of rational expectations. An alternative approach ascribes these deviations to the behavioral and psychological mechanisms such as herding by arguing that investors are not fully rational but rather are characterized by bound rationality (Thaler, 1993; Shefrin, 2000). There is also increasing awareness of the fact that financial data sets often display empirical evidence of the regime shifts (Van Norden and Schaller, 1993).

An interesting viewpoint for the explanation of the financial market crashes is adopted by Sornette (2003) who borrows his ideas heavily from the behavior of complex systems in physics which are known to display certain critical points (reversal points). Such systems with a large number of mutually interacting parts have the ability to self-organize their internal structures leading to emergent properties over time. These systems are characterized by coherent large-scale collective behaviors resulting from the non-linear interactions among its constituents. Sornette argues that big rallies are caused by local self-reinforcing imitation between the traders which leads to bubbles. This process leading to the bubbles, however, is also characterized by an increasing potential of instability due to the amplification of the non-linear feedback chains eventually culminating in critical points with an eventual regime shift (crash). Furthermore, he also argues that the timing of a crash in advance is possible by using certain log-periodic precursory signals (Sornette, 2003).

The search for certain mathematical or statistical signals accompanying the market crashes and major reversals is not limited by Sornette’s log-periodic signals. Extreme value theory (Embrechts et al., 1997; Bassi et al., 1998) employs peak over threshold distributions in the form of Generalized Pareto distributions for this purpose. Lillo and Mantegna (2000) shows that the shape of the return distributions display significant distortions (both in tails and center) in the extreme trading days compared to normal days. L’vov et al. (2001) argues that extreme events in the financial markets is not simply a magnified version of usual normal scale fluctuations but rather the overall return distribution is made of two different populations of the body and of the tail which have different physical properties and different scaling. Most significantly, turbulences exhibit outliers.

The gist of this paper shares the essential idea in the literature cited above and argues that the major reversals in a financial market are characterized by certain unique statistical footprints. The paper, however, does not try to explain the reasons leading to a rally or crash though I personally believe that herding plays some role in these processes with some dose of irrationality. What the paper aims to accomplish is to identify a set of fairly reliable statistical tools which signal that the market is changing its mood from a bullish rally to a major bearish tone or vice versa. It implicitly assumes that such a change is characterized by emotional excesses which statistically lead to extreme price moves on the day of reversal. In other words; major reversal days are characterized by outliers. These outliers can easily be determined by employing fairly simple tools. We observe a more than two standard deviation in the daily price rate of change (daily return) in a major reversal day which we do not see in earlier days. We may also occasionally see a sudden volatility burst in such days. However, these outliers may be more disguised in some cases whose identification may require a more sophisticated approach such as searching these
outliers in the wavelet components of the series. The paper is about the test of this proposition in the context of the BIST-30 index and is motivated by an earlier research of us (Oztürk et al., 2016) which demonstrated that the use of extreme volatility estimators (which incorporates day’s extreme prices) substantially improves the daily BIST-30 volatility forecasts.

The rest of the paper follows the usual format. Section 2 deals with data and methodology. Section 3 presents the empirical evidence and Section 4 concludes.

II. Data and methodology

The data are the daily closing prices of the BIST-30 index from January 4, 2010 to June 15, 2016 with a total of 1,627 observations derived from Bloomberg. The major emphasis is to identify the major reversals (from bullish to bearish and vice versa) in the latter part of that period and then to check if these reversal days are characterized by statistical outliers.

The price series itself (closing prices of the BIST-30 index) is not a suitable instrument to detect relevant outliers, information wise[1]. What we search for is an outlier that uniquely characterize the major reversal day which is defined as a more than two standard deviation in the series considered. This outlier must also be unique in the sense that it is not observed in the days prior to reversal but shows itself only at the point of reversal. One such candidate is the daily returns. Relevant outliers can be observed in the daily return series (which simply is the day’s rate of change) when the market reverses itself from a major bullish (bearish) trend to a major bearish (bullish) trend. Another candidate is the volatility which may demonstrate a sudden burst during major reversal days. The outlier identification at reversal days may also utilize the volatility series. However; such unique outliers are observed essentially in the volatility of the rate of change rather than the volatility of price series[2].

The two tools mentioned above are easy to implement and can effectively be utilized by fund managers. Some major reversals, however, are harder to detect using the above mentioned tools though they represent a minority of the cases considered. They do not reveal a more than two standard deviation outlier in either the daily return (rate of change) or the volatility of daily return series. In such cases; detection of outliers may require a more sophisticated approach. An effective tool to use in such disguised cases may be to analyze the wavelet components of the daily rate of change (daily return) series. As it will be evident in the next section; the wavelet components of daily return series have a good record for revealing outliers at the reversal days. Even when daily return or volatility of return series does not display outliers at the critical major reversal days; such behavior may be detected by checking the wavelet components of the daily return series.

Wavelet analysis provide a complete representation of a data series from both time and frequency perspectives simultaneously. Thus they permit to break down the market activity into different frequency components and to study the dynamics of each of these components separately. This analysis is especially useful to analyze nonstationary data (data whose mean and variance change over time) which is a typical characteristic of financial data.

A wavelet is essentially a small wave which grows and decays in a limited time frame. This small wave is characterized by a mathematical function \( \psi(t) \) which is called a mother wavelet. There is a variety of different functions which may be used to characterize the small wave. The chosen function \( \psi(t) \) can be dilated and translated as:

\[
\Psi_{u,s}(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{t-u}{s}\right)
\]

where \( u \) and \( s \) are the location and the scale parameters, respectively and \( \sqrt{s} \) is the square root of \( s \). The wavelet coefficients \( W(u, s) \) which is a function of \( u \) and \( s \) are
obtained by projecting the original function $x(t)$ or the data set onto the mother wavelet $\psi_{u,s}(t)$ yielding:

$$W(u, s) = \int x(t)\psi_{u,s}(t)dt$$

(2)

The $W(u, s)$ is a continuous wavelet transform of the original data. In financial applications; we rather use the discrete wavelet transform (DWT). The DWT uses only a limited number of translated and dilated versions of the mother wavelet to decompose the original signal (Masset, 2008). DWT selects $u$ and $s$ so that information contained in the signal (data) can be summarized by a minimum of wavelet coefficients. This is achieved by setting $s = 2^{-j}$ and $u = k2^{-j}$ where $j$ and $k$ are integers representing the set of discrete translations and discrete dilatations. The DWT can only be applied to series of dyadic length (the number of observations used must be equal to the powers of 2).

The DWT is based on two discrete wavelet filters which are the mother wavelet and the father wavelet. The mother wavelet (differencing function) is a high-pass filter which measures the deviations from the smooth components. The father wavelet (scaling function) captures the long scale (low frequency) components of the series and generates the scaling coefficients. The application of both the mother and the father wavelets allow separating the low frequency components of the time series from the high frequency components. The wavelet and scaling coefficients at the first level of decomposition are obtained by convolution of the data series with the mother and father wavelets. After having applied the mother and father wavelets on the original data series, one has a series of high frequency components and a series of lower frequency components. The low frequency components (scaling coefficients) can be further decomposed into high and low frequency components by recursive application of a pyramid algorithm up to scale $J = \lceil \log_2(T) \rceil$.

In this study; the wavelet Daubechies-3 is used to obtain the DWTs. Choosing a level of 3 with dyadic data is sufficient for our purpose. This DWT generates three levels of approximation coefficients which display the average behavior of BIST-30 index and three levels of detail coefficients which indicate short bursts in the data and being more responsive to the jumps generated by new information. Rather than focusing on all approximation and detail levels; I primarily used A1 (approximation level 1) and D3 (detail level 3) to detect the outliers of the reversal days. Empirical tests showed that the type of wavelet chosen for the analysis has only a minor impact on the results.

III. Empirical results

In what follows; I presented evidence of extreme outliers at major reversal points by analyzing several different cases. The cases presented cover almost all of the major reversal points of BIST-30 index during the period from November 22, 2013 to June 15, 2016. I also tested if such extreme outliers are detected in case of minor reversals. The purpose here was to highlight the fact that the extreme outliers are a specific footprint characterizing only the major reversals. Then a brief evaluation follows which discusses the importance of outlier information for the fund managers.

Case 1 – Observation 775 (Date: January 24, 2013)

This is the end of a major bullish move which started on observation 730 (date: November 22, 2013) when the BIST-30 index was 86,993. The index value of observation 773 (daily close) was 108,090 representing a gain of 24.3 percent during the bullish rally. It is technically the first head of a major double top. In fact; it was the latter part of a bigger bullish move which started on observation 609 (date: May 30, 2012) when the index value was 65,956 thus representing a significant gain of 63.8 percent. The market closed down
next day at 105,800 with a daily decline of 2.8 percent at which point a downward move started that continued until observation 794 (date: February 22, 2013) leading to a loss of 13.3 percent. The chart in Figure 1 with a small square at reversal point illustrates vividly the transition from a major bullish trend to a bearish trend.

The change from the bullish phase to the bearish phase is signaled by the close of observation 775 (after a further drop to 101,110). When we enter the observations (closes of the BIST-30) from 1 to 775 and calculate the daily rate of changes (daily returns); we see that the last entry corresponding to the daily rate of change from observation 774 to 775 represents a more than 2 standard deviation change (an extreme outlier)[3]. Please take note of the fact that the most recent outlier (represented by a more than two standard deviation change in the rate of change series of the BIST-30) prior to the one at observation 775 is recorded at observation 648 which was in fact also a reversal point from a bearish trend to a bullish trend[4].

The aberrant behavior of observation 775 is also confirmed by further evidence. When I enter first 775 observations, calculate the daily return (rate of change) of the closing prices and then calculate the volatility of the returns; it is observed that the last entry (observation 774)[5] is an extreme outlier with a deviation of more than 3 standard deviations. The most recent outlier prior to observation 774 is detected at observation 479 of the volatility series which actually corresponds to a major bullish reversal. Furthermore; when we enter the first 775 observations; calculate the daily rate of changes, delete the first 262 observations and extract the A1 (approximation level 1) and D3 (detail level 3) wavelet transforms from the remaining dyadic series of 512 observations; we observe that the last entry[6] displays extreme outliers (a deviation greater than two standard deviations) both in A1 and D3 components[7]. The reversal from a major bullish rally to a major bearish rally is signaled effectively by the aberrant behavior of rate of change series, by the volatility of the rate of change series as well as by the A1 and D3 wavelet transforms. This aberrant behavior displays itself as more than two standard deviations (as extreme outliers) in all of these series[8].

Is that extreme behavior unique in the sense of characterizing only the reversal points? A simple test can clarify this issue. I entered ten previous observations of daily closes starting with observation 764[9]. The market during this period was displaying a strong uptrend without any corrections. Neither the rate of change of closing prices nor the

![Figure 1. Case 1](image)
volatility of the rate of changes displayed any extreme outlier during this ten day period. The D3 (detail level 3) displayed an outlier for observation 765 but it can be ignored given the highly responsive nature of high frequency data to daily changes. An interesting case is the case of observation 774 which displayed more than two standard deviation outliers in both A1 and D3 series but unable to find an extreme outlier in either the rate of change or the volatility series. The observation 774 corresponds to the first reversal day after the strong bullish move with index dropping from 108,090 to 105,800. The major reversal was confirmed one day later by rate of change and volatility series after a further drop to 101,100. This suggests that the wavelet components may identify a major reversal earlier than the rate of change or volatility series in some cases. The essential point, however, is the fact that the series used for outlier detection reveal extreme behavior at the point of reversal but not before.

Case 2 – Observation 860 (Date: May 29, 2016)
The bearish reversal that started after observation 775 continued until observation 794 (February 22, 2013) after which a new bullish rally started that continued until observation 855 (May 22, 2013). The index value moved up from 93,664 to 115,341 representing a gain of 23.1 percent (see Figure 2). The market started to reverse at observation 856 with a daily decline of 2.2 percent. The downward move that started at this point continued until observation 922 (August 27, 2013). The index fell from 115,341 to 79,952 with a loss of 30.6 percent. The extreme outliers detect this major reversal at the slightly later date of May 29, 2013 when the index value is 107,136[10]. The rate of change series displays a deviation of more than two standard deviations on this date though volatility series do not display extreme behavior. The wavelet components A1 and D3, however, confirm the aberrant behavior displaying deviations of more than two standard deviations on that particular day. The most recent outlier prior to observation 860 is detected by both rate of change series and wavelet components at observation 840 which was the starting point of a strong rally. In fact; wavelet component A1 signals an extreme behavior (a deviation of more than two standard deviations) at observation 856 (the first day of reversal). A test of previous days starting with observation 849 do not reveal any extreme behavior prior to observation 860 (or observation 856 in case of A1).
Case 3 – Observation 938 (Date: September 19, 2013)
This case represents the end of a strong bullish move which started on August 27, 2013 (observation 922) when index value was 79,952 and reaching to a value of 98,375 (at observation 938) with an overall gain of 23 percent (see Figure 3). The observation 938 represents the end of the bullish rally and corresponds to the first head of a significant double top (see Figure 3). The index then declined to 91,670 on September 27, 2013 (observation 944). A second rally to 97,333 on September 27, 2013 (observation 961) created the second top after which index fell down to 77,856 on December 27, 2013 (observation: 1,004).

All the series signal a major reversal of index at observation 938. The rate of change series display a more than three standard deviation on that particular day and volatility series confirm the extremity by a more than two standard deviation in the same day. The A1 series display a more than three standard deviation on that particular day and D3 series also signal aberrant behavior represented by a more than two standard deviation change. The extreme outlier characteristic of all the four series on September 19, 2013 is quite unique since the previous days starting with observation 930 do not display such a character (the only exception is observation 935 with extreme outliers in the rate of change series and A1 series but not in the other two series. This date, in fact, was also a bullish reversal day after a correction)[11].

Case 4 – Observation 1,005 (Date: December 30, 2013)
A very sharp decline started on December 16, 2016 (observation 995). The index value plummeted from 91,399 on December 16, 2016 to 77,586 by December 27, 2013 (observation 1,004)[12]. The index then turned up sharply on December 30, 2013 with a 6.5 percent daily gain starting a rally which continued until January 7, 2014 (observation 1,010). The rally led to a 7.4 percent gain in six days (see Figure 4). This strong bullish reversal is typically characterized by extreme outliers. The rate of change series, volatility series and A1 series displayed more than three standard deviations on that date (the previous outliers detected by the data set ending at observation 1,005 were at observations 938 and 995). The previous entries starting with observation 996 do not display any extreme outliers with the exception of observation 997 where more than two standard deviations are observed in the rate of change and A1 series (observation 997 corresponds to a neckline breakdown plausibly confirming the onset of the bullish trend).
Case 5 – Observation 1,050 (Date: March 4, 2014)
A major bearish move started on October 28, 2013 (observation 961) when the index value was 97,333. The bearish move continued until March 3, 2014 when the index close was 74,173 representing a loss of 23.7 percent. We then see a typical double bottom signifying the end of the major bearish move after which a big bullish rally started eventually leading to an index value of 103,953 on July 25, 2014 (observation 1,150) representing a gain of 40.1 percent[13].

The observation 1,049 is the dip of the second bottom. The bullish rally started on March 4, 2014. The bullish rallies, however, seem to be more subtle than the bearish ones in the sense that their major reversal points are not characterized by easily recognizable footprints. The reversal that started on March 4, 2014 fails to display extreme outliers (deviations exceeding two standard deviations) in the rate of change or the volatility series. We do not observe such outliers at observation 1,050 or at the three consecutive observations (also not in the preceding ones). In such cases; wavelet components which are more responsive to reversals may provide valuable information about the impending change in the market.

When we check the outlier status of the entries starting with observation 1,041; we do not observe any outliers in A1 and D3 components in the first four days. Observation 145 corresponds to a minor bearish reversal and this reversal generates outliers (deviations exceeding two standard deviations) on that day. Following two more quiet days; we observe an extremity (a deviation of more than two standard deviations) of D3 series at observation 1,048. Then we also observe an extremity (a deviation of more than two standard deviations) in the A1 series at observation 1,049. Following these early signals; the reversal day 1,050 is characterized by extreme behavior (deviations exceeding two standard deviations) in both A1 and D3 series.

Case 6 – Observation 1,276 (Date: January 28, 2015)
The bullish trend that started on December 16, 2014 (observation 1,246) when the index value was 97,978 continued until January 26, 2015 (observation 1,274) reaching to a value of 113,614 with a gain of 15.9 percent. The market reversed and the downward trend continued until March 13, 2015 (observation 1,308). The index declined from 113,614 to 93,828 with a loss of 17.4 percent.
The footprints of this major reversal are detected at observation 1,276 which was characterized by a strong daily decline. This reversal, however, cannot be detected by extreme behavior of daily rate of change or volatility series. The latest outlier detected in these series is at observation 1,245 which in fact was a significant bullish reversal. As in the previous case (case of observation 1,050); the reversal is signaled by extremity of wavelet components. The A1 approximation and D3 detail series display outliers at observation 1,276 with deviations exceeding two standard deviations on that particular day. It is a noteworthy fact that none of the previous observations starting with observation 1,266 display any outliers in the A1 and D3 wavelet components. The aberrant behavior of wavelet components is detected only and uniquely at the major reversal point (at observation 1,276).

Case 7 – Observation 1,366 (Date: June 8, 2015)
A bearish trend that started on May 18, 2015 when the index value was 108,609 ended on June 8, 2016 when the index value was 95,205 after a 12.8 percent decline. The bullish reversal started on June 9, 2015 (observation 1,367) leading to a value of 102,928 on June 24, 2015. The immediate bullish reversal was signaled by the significant daily decline (5.71 percent) on June 8, 2015 which led to observations of extreme outliers in all the relevant series. The daily rate of change series and the volatility series display a change of more than three standard deviations on that day. The wavelet components A1 and D3 also display changes of more than three standard deviations on that day confirming an aberrant behavior at the point of bullish reversal. Such an aberrant behavior is not observed in the previous observations (in consecutive data sets starting with observation 1,360). The more responsive A1 and D3 series detect an outlier at observation 1,362 which was a minor (though quite significant) bearish reversal of 8.33 percent though the extreme outlier characteristics of all the four major series is uniquely confined to the observation 1,366 which is immediately followed by the onset of a bullish reversal.

Case 8 – Observation 1,420 (Date: August 24, 2015)
A significant bearish move started on May 18, 2015 (observation 1,352) when the index value was at 108,609. The index then declined 19.3 percent reaching to a dip value of 87,545 which is observed on August 24, 2015. This dip was immediately followed by a bullish reversal which continued until November 2, 2015 (observation 1,467). This bullish reversal carried the index to the peak value of 103,595 with a significant gain of 18.3 percent. The bullish reversals are harder to detect on the basis of the outlier information as we had also seen in the case of observation 1,050. The observation 1,420 is characterized by the same problem. Neither the rate of change series nor the volatility series exhibit an outlier character (deviations of more than two standard deviations) on August 24, 2015 despite a daily decline of 3.13 percent (outliers are not detected in the following three bullish days as well). The immediate bullish reversal, however, can be detected by the more responsive wavelet components. The A1 (approximation) component of daily rate of changes display a typical outlier at observation 1,420 with more than three standard deviation change which is also confirmed by a more than two standard deviation change in the D3 series. This outlier characteristic of wavelet components signaling the impending bullish reversal and the termination of bearish trend is not observed in the preceding days starting with observation 1,414. It is only detected at observation 1,413 (both A1 and D3 displaying more than two standard deviations on that day) which in fact was the initial day of a sharp decline[14].

Case 9 – Observation 1,467 (Date: November 2, 2015)
A bullish trend that started on September 14, 2015 (observation 1,435) when the index value was 87,593 continued until November 2, 2015 by reaching to a peak value of 103,593 on that day.
This bullish move was characterized by a gain of 18.2 percent from the dip to the peak. The observation 1,467 is the final day of the bullish rally marked with a significant daily return of 5.33 percent. The index turned bearish by the next day (March 11, 2015) initiating a long decline until December 14, 2015 (observation 1,497) on which day it reached to the dip value of 84,893 after a loss of 18.05 percent.

All the four series display outliers on November 2, 2015 signaling the major bearish reversal. The rate of change series and volatility series display more than three standard deviation changes on November 2, 2015. Similarly, the A1 and D3 components also display more than three standard deviations on that particular day. This observation of outliers in all the four series on December 15, 2015 is a distinctive event since such behavior is not observed in any of the preceding days starting with observation 1,459. Neither the rate of change and volatility series nor the wavelet components display any aberrant behavior (deviations of more than two standard deviations) in any of these preceding days.

Case 10 – Observation 1,498 (December 15, 2015)
The bearish move that started on March 11, 2015 continued until December 14, 2015 as mentioned above. The observation 1,497 is the first dip of a big double bottom that formed in the consequent days. The major bearish move effectively ended on December 14, 2015 and the market started reversing itself by a daily up move of 4.26 percent on December 15, 2015 which was the starting point of a new bullish rally which eventually reached to a value of 105,912 on April 18, 2016.

The bullish reversal in this case deviates from the previously analyzed bullish cases and displays outliers in all the four relevant series. The rate of change series and the volatility series display more than two standard deviations on December 15, 2015. Similarly, the A1 and D3 components also display more than three standard deviations on that particular day. This observation of outliers in all the four series on December 15, 2015 is a distinctive event since such behavior is not observed in any of the preceding days starting with observation 1,487. The wavelet components A1 and D3 however start signaling a possible reversal at earlier days by starting to display outliers after observation 1,495 due to their higher sensitivity (we detect wavelet outliers in observations 1,496 and 1,497 but not before).

The analysis so far reveals a close relationship between the major reversal points of the index and the extreme outliers. Extreme outliers emerge at the end of a major bullish (bearish) trend or at the initial stages of a major bearish (bullish) trend. They can be detected in the rate of change and/or volatility series. They are also detected in the A1 (approximation) and/or D3 (detail) components of the rate of change series. All the four series may be characterized by extreme outliers in some major reversal points. There are, however, also cases (mostly in the case of bullish reversals) in which only the outliers of the A1 and D3 components signal a possible reversal. The A1 and D3 wavelet components seem to be more sensitive to reversals in the market. An important point is the fact that the outliers are observed at the major reversal points but not in the previous observations.

Are these outliers a unique event characterizing only the major reversal points? A partial answer was provided in the ten cases analyzed above. They do emerge at major reversal points but not before. A further test may clarify the issue better. In what follows, I randomly selected six minor reversals and repeated the same procedure to find outliers at these minor reversal points if there are any. The selected reversals have the common property of being preceded by a relatively limited price move and also displaying only a limited price move after the reversal. They are generally temporary corrections or reactions observed during a major bullish or bearish rally.

The first one is observation 719 which is a minor correction that was observed during a major bullish rally. The index fell from 90,659 on November 8, 2012 (observation 719) to 86,932 on November 22, 2012 (observation 729) after a small triple top. The bullish trend
then resumed after the limited downward move of 4.11 percent. The observation 890 represents a minor upward reaction observed during a major bearish move (from observation 856 to observation 922). It is the second dip of a double bottom which however failed to lead to a bullish rally. It was the end of a limited bearish move from 93,228 on July 1, 2013 to 85,912 on July 10, 2013. This downward move of 7.8 percent reversed on July 10, 2013 leading to a reaction reaching to 93,886 on July 17, 2013 after a gain of 9.2 percent. Bearish trend then resumed with a further 14.6 percent drop. The observation 1,085 was a minor correction during a major bullish rally characterized by a fall from 89,460 on April 22, 2014 (observation 1,085) to 87,118 on April 25, 2014 (a downward move of 2.6 percent).

The fourth case is observation 1,214 which is a minor correction during a major bullish rally. This correction led only to a limited downfall from 98,957 on October 31, 2014 (observation 2,014) to 94,749 on November 6, 2014 (observation 1,218). Bullish trend resumed after this downward move of 4.2 percent. The observation 1,425 corresponds to the top of a double bottom (after the completion of which a major bullish rally started). The major bearish episode was terminated at observation 1,420 (case 8). A bullish reversal that started on August 24, 2015 when the index value was 87,545 led only to a 5.5 percent increase reaching to 92,360 on August 31, 2015 (observation 1,425) at which point it turned bearish leading to a downward move of 5.1 percent at the end of which it marked the second dip of the double bottom on September 14, 2015 (observation 1,435) with an index value of 87,593. The observation 1,508 also corresponds to the top of another double bottom. The observation 1,497 was the end of a major bearish rally (case 10). A bullish reversal that started on December 15, 2015 (observation 1,498) carried the index to 90,785 on December 28, 2015 (observation 1,508) representing an upward move of 6.9 percent but went out of steam fairly quickly and was followed by a downward move from 90,785 to 85,981 on January 4, 2016 (observation 1,512). The downward move (a 5.3 percent decline) was a limited one as the preceding upward move[15].

The rather detailed description of these six cases above serves the purpose of highlighting an important point. The previous ten cases are major reversals in the sense of being preceded by substantial changes in the index value (usually over 10 percent) which are followed by a substantial change in the opposite direction. All of them are characterized by a significant mark at the reversal points which is fairly easy to identify. The rate of change in the prices displays outliers either at the final day of the previous bullish/bearish trend or at the first day of the opposite trend. In some cases; these outliers are also detected in the volatility of the daily returns. If both of them fail to identify a reversal, the A1 and D3 components of the daily returns (price rate of changes) display such outliers which are important warning signals of the impending change in the market mood.

The minor reversals covered by the six cases above are different in the sense of being preceded by rather limited changes in the index value and also followed by a limited change in the opposite direction. The outlier characteristics of these minor reversals show a remarkable difference from the major reversals. They do not display outlier characteristics at all. We cannot detect outliers in any of the four series (in rate of changes, volatility of returns and wavelet components). This negative result is valid for all the six cases considered (except for an outlier in the highly responsive D3 component detected at observation 890). The empirical evidence so far seems to indicate that there may be a way of distinguishing major reversals from the minor reversals at the reversal point (at least in the context of the BIST-30 index[16]).

The importance of identifying the major reversals at the initial stage of such a reversal is quite evident. Its importance for portfolio management may easily be demonstrated by the following two simple cases. During the bullish rally from December 15, 2015 to April 18, 2016; the price of Akbank (a stock trading in Istanbul Stock Exchange) increased from 6.42 (closing price on December 15, 2015) to 8.48 (closing price on April 18, 2016) which is an increase of 32 percent. Akenerji (another stock trading in Istanbul Stock Exchange)
increased from 0.86 to 1.19 (an increase of 38.3 percent). During the bearish phase from March 11, 2015 to December 14, 2015; Akbank’s price dropped from 7.70 to 6.16 (20 percent) and the price of Akenerji dropped from 1.00 to 0.85 (15 percent).

There is another important aspect regarding the importance of identifying the major reversals on time. Index futures is an essential hedging instrument for the portfolio managers. The hedging operation requires assuming short positions in index futures when a major drop in the index is anticipated. The rules for choosing the correct number of contracts to be shorted for a given portfolio is well established[17]. The major problem in the implementation of a hedging operation by using the index futures is to determine the correct time for shorting the index futures which usually is not an easy task in the volatile environment of the markets. Identification of a major bearish reversal by outliers at a peak point signaling the high probability of an impending substantial drop in the index value may be important to choose the correct time for initiating the hedge position (as well as for the correct timing of covering the shorts).

IV. Conclusion

This paper is about the identification of the major reversal points in the context of the BIST-30 index. The empirical analysis based on daily data for a six year period strongly suggests that the major reversal points of the BIST-30 index are characterized by extreme outliers. The extreme outliers are defined as more than two standard deviation changes observed in the vicinity of major reversal days. Major reversal days are interpreted as days when a substantial bullish (bearish) move in the index is reversed leading to a substantial move in the opposite direction. Such outliers can be detected in the daily return (rate of change) series, volatility of return series and in the level 1 approximation (A1) and detail level 3 (D3) wavelet components of the daily return series. Wavelet decomposition is achieved by using dyadic observations and Daubechies-3 wavelet. Such outliers seem to emerge on the day of the major reversal and not before.

A plausible explanation of this observation may be the following: A major reversal day is a day when the mood and expectations of the market suddenly change probably due to some new information. It seems that such a change is also characterized by emotional excesses coupled with some degree of herding behavior leading to aberrant price volatility and outliers. The empirical analysis also suggests that such outliers characterize the major reversals rather than the minor reversals (with the possible exceptions of highly responsive D3 outliers which may indicate the minor reversals as well).

This empirical fact detected in the BIST-30 index may be important for fund managers. When the market changes direction; it is generally very difficult to judge if the observed reversal is a short-term phenomenon or the starting point of a long-term major trend in the opposite direction. The empirical tools suggested in this paper may provide an answer to this question. Apparently; a more robust conclusion requires further analysis and the extension of this kind of analysis to the prices or values of other indices and financial asset prices. The preliminary evidence, however, implies the importance and possible benefits of considering the outliers in the somewhat chaotic environment of financial markets.

Notes

1. When markets are characterized by a strong rally prior to reversal; prices exhibit a successive series of outliers whose information value is negligible. What we search for is a unique footprint (outlier) which characterizes the major reversal day but is not detected before the reversal day while the rally continues.

2. That simply means that we check if the squared deviation of the daily rate of change from its mean value is a more than two standard deviation change (an outlier) during the reversal day.
3. Please take notice of an important fact. I do not use a backward analysis such as entering all the data (1,627 observations) and then checking for extreme outliers. I enter only the data up to observation 775 and search outliers on the basis of this data set which has 775 observations. This procedure is also used in all the remaining cases. Apparently this is what a fund manager will be faced with in the real life. Another brief note is in order. The first reversal considered was the one at observation 775. There were other major reversals before that date. However, this type of analysis requires a fairly long data set to be statistically robust. Also the use of discrete wavelet transforms compels us to use dyadic data. In the case of observation 775, I used the dyadic 512 observations (by deleting the first 263 observations of the data set and thus considering the 512 observations prior to observation 773). The number 512 is dyadic since it is equal to $2^9$. The discrete wavelet approach also requires a fairly long series to be robust enough.

4. The list of previous extreme outliers detected by using a data set of 775 observations are at observations 486, 488, 504, 521 and 525 all being earlier reversal points.

5. The last entry of the series is 774 instead of 775 because the price rate of change operation automatically deletes the first observation. The entry 774 corresponds to the rate of change from the close of observation 774 to the close of the observation 775.

6. The last observation in the dyadic series of 512 observations corresponds to the rate of change from the close of observation 774 to the close of observation 775 whose wavelet transforms A1 and D3 are extreme outliers.

7. They also represent a behavior quite unique to the entry 512 since the earlier most recent outliers are detected at observations 386 and 388.

8. All the statistical operations and wavelet analysis is evaluated in MATLAB environment.

9. As usual, I entered only the observations from 1 to 764 in order to search the extreme outliers and conducted the tests on this observation set to see if the last entry qualify as an extreme outlier. Then I entered the data set from 1 to 765 repeating the test with this data set and so on.

10. The 100-day moving average which is commonly used in technical analysis to identify the changes in the primary trend gives the signal (price crosses below moving average) on March 6, 2013 when the index value (close) is 94,827.

11. The bullish move that ended on September 19, 2013 actually started on September 19, 2013 (at observation 930) when the index value was 85,434. The fact that a bullish rally started on that date is also signaled by extreme outliers since rate of change series display a more than two standard deviation on that date confirmed by more than three standard deviation change in A1 and more than two standard deviation change in D3 series.

12. Though it is not covered in the main body of the text, the sharp drop that started on December 16, 2016 is also signaled by the extreme outliers detected on that date. The rate of change series and the volatility series both display deviations in excess of three standard deviations as well as the A1 component.

13. The charts after Figure 4 are not displayed for the sake of brevity. They are available upon request.

14. The index following an indecisive horizontal period during which a small double top formed started to decline on August 13, 2015. The index value declined sharply in a few days from 94,589 on August 13, 2015 to 87,545 on August 24, 2015. The wavelet components by revealing outliers on August 13, 2015 successfully signal the sharp decline at its initial stage.

15. A more detailed analysis of two cases may further illustrate the benefit of checking outliers. When a major bearish trend is over, this is usually signaled by an early bullish reversal. Then the market enters into a horizontal phase during which certain formations (e.g. double bottoms) are formed. A strong bullish rally starts after the neckline of the double bottom is broken (usually after a neckline test). In case 8, major bearish trend was over on August 24, 2015 with an initial bullish reaction. The outliers successfully detected this change. In fact; outliers detected on August 24, 2015 were signaling the end of major bearish move rather than identifying a new
bullish episode. The first initial bullish move led only to a limited upward move reversing direction on August 31, 2015. As indicated in the main text, this reversal is not characterized by any outliers implying that the reversal was a minor one. The index initiated a second bullish move on September 15, 2015 (corresponding to the second dip of the double bottom). The A1 and D3 components remarkably identify the starting point of this second bullish move by displaying deviations of more than two standard deviations in both components on that day. The preceding four days, for example, do not display such an event. The onset of a really strong bullish move was, however, somewhat later starting on October 5, 2015 (observation 1,448) which roughly corresponds to the date when the prices crossed over the neckline of the double bottom. The outliers successfully detected the onset of this strong rally by displaying outliers (deviations exceeding two standard deviations in the rate of change series and deviations exceeding three standard deviations in the A1 and D3 components). The outliers appeared exactly on October 5, 2015 but not in the previous five days. The rally carried the index from 94,671 to 103,593 in less than one month. Another major bearish move was over on December 15, 2015 (case 10). The initial bullish reaction was rather limited reversing on December 28, 2015. This led to a minor bearish move whose onset was not accompanied by outliers. The Index turned bullish again on January 22, 2016 after forming the second dip of the double bottom. The A1 and D3 components displayed outliers exactly on that date though they are not observed in the previous days. Following a possible neckline test; a strong bullish rally started on February 17, 2016 (observation 1,543). That particular day is characterized by outliers of both A1 and D3 components though we do not observe them in the previous days. The result was an upward move from 87,104 to 105,912 in nearly two months.

16. Though I analyzed only six cases in the main body of the paper; I also checked seven additional reversals which may qualify as minor reversals in addition to the six reversals already mentioned. These reversals were located at observations 884, 1,040, 1,083, 1,520, 1,535, 1,567 and 1,623. We do not observe outliers of rate of change or volatility at the reversal points in any of them. Only one of them displayed an A1 outlier while five of them displayed D3 outliers. It seems that an outlier detected only in the highly responsive D3 series but not in the other three series (or if we do not observe any outlier at all during a reversal) tend to mark a minor reversal while the additional observation of outliers in A1 and daily return (price rate of change) series point out to the high likelihood of a major reversal.

17. The formula used to determine the number of contracts to be shorted is (portfolio value/(index futures price \times contract multiplier) \times portfolio \beta) (Erol, 2014). A hedging based on this formula guarantees perfect hedging (a gain in short futures fully compensating the loss in the portfolio value) if the difference between the index futures price and the spot index value (basis) remains unchanged during the hedging period.

References


Corresponding author
Ümit Erol can be contacted at: umit.erol@eas.bau.edu.tr
The flash crash: a review

Ali N. Akansu
Department of Electrical and Computer Engineering,
New Jersey Institute of Technology, Newark, New Jersey, USA

Abstract

Purpose – The purpose of this paper is to present an overview of the flash crash, and explain why and how it happened.

Design/methodology/approach – The author summarizes several studies suggesting various perspectives on the flash crash and its causes. Furthermore, the author highlights recently proposed and introduced improvements and regulations to reduce the risk of having similar market collapses in the future.

Findings – It is an overview paper that highlights the state of the art on the subject.

Research limitations/implications – Paper does not report any research findings of the author.

Practical implications – High-frequency trading (HFT) along with its pros and cons is the new normal for most of the current electronic trading activity in the markets. It is well recognized by the experts that HFT may have its important shortcomings whenever the rules and regulations are not up to date to match the technological progress offering faster computational and execution capabilities.

Social implications – HFT has created a societal discussion about its benefits and potential deficiencies as the common practice for trading due to potentially unequal access to market data by various categories of participants. Such arguments help the regulators to develop improvements to reduce the market risk and nurture more robust and fair markets for all.

Originality/value – The paper has a tutorial value and summarizes the current state of HFT. The readers of more interest are guided to the most relevant literature for further reading.

Keywords Algorithmic trading, Electronic trading, High-frequency trading (HFT), Limit order book (LOB), The flash crash, LOB imbalance, Security Information Processor (SIP), National best bid and offer (NBBO), Regulation national market system (Reg NMS), United States Securities and Exchange Commission (SEC)

Paper type Research paper

I. Introduction

US equity markets with about 8,000 tickers literally collapsed within a few minutes and quickly recovered by the end of the day on May 6, 2010. They were down in the range of 5-15 percent, and a few hundred of them lost 60 percent of their values in this short period of time. The markets recovered most of their losses at the end of the day, and closed the day 3 percent down from the close of the previous day. The market has lost about $1 trillion of its value during this episode due to the chaotic nature of volatility and disappearance of liquidity that made limit order books (LOBs) of highly correlated equities extremely unstable and dysfunctional, and the entire system collapsed. This event, called “the Flash Crash”, was the first of its kind experienced by the financial industry and generated serious concerns, and highlighted the weaknesses of financial markets and some of the regulations (SEC, 2010a, b; Lewis, 2015). Mostly, high-frequency trading (HFT) has been presented as the culprit by many experts and writers that created a societal discussion about the use of the-state-of-the-art technology in the financial sector. The jury is still out although HFT has become the new normal for the trading infrastructure in the USA.

In this paper, we present an overview of the flash crash, and explain why and how it happened. We summarize several studies suggesting various perspectives on the flash crash and its causes (SEC, 2010a, b; Bouchaud et al., 2002). Furthermore, we highlight...
recently proposed and introduced improvements and regulations to reduce the risk of having similar market collapses in the future.

The price formation process of an asset in an exchange and its consolidated version for the multiple exchanges based trading infrastructure are discussed in the following section. In Section 3, HFT is presented and its unique characteristics are highlighted. Then, we detail the underlying events of May 6, 2010 and dramatic regime change in the financial markets that triggered the collapse in Section 4. The experiences gained and lessons learned from the flash crash are summarized in Section 5. Then, we conclude the paper.

II. LOB and price formation of an asset

LOB is the list of ask and bid prices along with attached quantities and order types of a tradable financial asset. The information posted on an LOB is continuously checked and updated by the computationally powerful matching engine of an electronic trading platform that executes and reports trades of a single ticker in real time at an exchange. Commonly, a single ticker is simultaneously traded at multiple exchanges. In the USA, these exchanges periodically report their best quotes (bids and offers) and executed trades to the Security Information Processor (SIP) (Consolidated Tape Association (CTA)) that functions as the benchmark of the market. These data are consolidated and displayed in real time as the National Best Bid and Offer (NBBO) governed by the Regulation National Market System (Reg NMS) that is described and enforced by the United States Securities and Exchange Commission (SEC) (Consolidated Tape Association Plan/Consolidated Quotation Plan (CTA/CQ); Regulation National Market System (Reg NMS), 2005). Regulation NMS sets the rules to ensure that trades occur at the best available prices at any time in order to achieve a robust, transparent and fair market for all participants.

Mathematically speaking, the price formation process of a stock through the market microstructure as highlighted is sophisticated with its intricacies. In theory, the price of a stock is usually modeled as the geometric Brownian motion in which the rate of return (price change) for a stock is defined as Bachelier (1900), and Akansu and Torun (2017):

\[
\frac{dp(t)}{p(t)} = \mu dt + \sigma dw(t), \quad t \geq 0
\]

where \(p(t)\) is the price of the stock, \(\mu\) the drift, \(\sigma\) the volatility or the standard deviation of price change (return) and \(w(t)\) the standard Brownian motion. This stochastic differential equation has its solution obtained by using Ito’s Lemma and expressed as Ito (1946):

\[
p(t) = p(0) exp \left[ \left( \mu - \frac{\sigma^2}{2} \right) t + \sigma w(t) \right]
\]

It is noted that geometric Brownian motion model assumes constant volatility. Since prices of stocks are affected by economic news and many relevant events that happen almost randomly with their unique impact, constant volatility assumption may not always hold. The flash crash is the case in point thoroughly investigated in the paper. More advanced price models incorporate the very fact that the volatility itself varies in time, and they include time local volatility in the formulation. Therefore, more realistic model is shown as:

\[
\frac{dp(t)}{p(t)} = \mu dt + \sigma[p(t), t] dw(t), \quad t \geq 0
\]

where \(\sigma[p(t), t]\) is the volatility at time \(t\).
Discrete-time counterpart of geometric Brownian motion model is obtained by sampling the price and given as (Akansu and Torun, 2017):

\[ s(n) = s(n-1) + \mu + \sigma \xi(n) \]

where \( s(n) = \ln p(n) \) is the log-price of a stock at discrete-time \( n \) with price \( p(n) \), \( \mu \) the drift, \( \sigma \) is volatility of the stock, and \( \xi(n) \) the white Gaussian with zero mean and unit variance, \( \xi(n) \sim \mathcal{N}(0, 1) \).

Limit orders sent by the market participants with their price, volume and trade limiting requirements are aggregated by the price and time priority (PTP) protocol and posted in the LOB. These limit orders are matched with market orders of the same price by the matching engine in the order execution process that creates trades. Trading cycle is completed through the settlement steps and recorded permanently as a trade at the exchange (Hasbrouck, 2007). The order executions to buy and sell assets causes price fluctuations that are formulated as the market impact (Cont, 2011). Execution strategies are developed to lower market impact at favorable prices and times identified through the real-time analysis of LOB structure and its variations (Akansu and Torun, 2017).

With the help of the example given in Akansu and Torun (2017), we explain herein the inner-workings of the LOB step by step for the limit and market order types, and track how they interact and impact the book to create a dynamically changing market for an asset in an exchange. Four LOB snapshots for a ticker are shown in Figure 1 to describe the LOB states just before and right after a limit order to buy 100 shares of let’s say a stock at $98.75 per share is placed by a trader. There are five limit orders at four different price levels prior to this order as seen in Figure 1(a). The lowest priced two limit orders are on the buy side (bid) while the remaining three orders are placed at the two highest price levels of the sell side (ask) where $98.77 price range has two independent orders sent to the LOB. The bid-ask spread (spread) of the LOB at this time and state is $0.01. Then, 100 shares of limit order to buy at $98.75 arrive and placed in the book based on the PTP protocol as depicted in Figure 1(b). Later in time, a market order of 200 shares to sell at $98.75 arrives at the LOB, and it is matched with the first arrived limit order of 200 shares sitting in the front of the best bid level, $98.75. Therefore, the available shares at that price level of the LOB are depleted by 200 as displayed in Figure 1(c). Next, the trader who placed the limit order to buy 100 shares at $98.75 cancels its order as depicted in Figure 1(d). It is noted that the bid-ask spread of the LOB at this time becomes $0.02. Similar to this example, such trading activities in an exchange create the ask (supply) and bid (demand) sides of the market for an asset through its LOB at that exchange. Typically, an asset is simultaneously traded at multiple exchanges with their own LOBs for the ticker. The status of these LOBs for the same ticker are periodically reported to the centralized SIP (Consolidated Tape Association (CTA)) by all the exchanges in order to post the NBBO that is utilized as the market reference by the participants and regulators (Consolidated Tape Association Plan/Consolidated Quotation Plan (CTA/CQ); Reg NMS, 2005). The trading dynamics of LOBs and the stability of the resulting NBBO lead to the market price formation and volatility (risk) of the asset under consideration where its market maker(s), with their privileges and responsibilities, play(s) a crucial role. The number of price levels and the number of orders (with quantities) in each level (complete book) are quite larger (in hundreds for some liquid stocks) in a real-world scenario than what we have in this example. The price level (bin) resolution of LOB and available order types may vary based on asset type and trading rules set by regulatory bodies, financial and legal frameworks that are implemented and serviced by exchanges.

It is desirable to describe the inner workings of LOB through the models where other significant metrics like the probability of upward/downward price moves (trends) based on the current and most recent LOB states can be analytically expressed (Cont et al., 2010).
Note that a thorough understanding of LOB along with the technological and procedural specifics of the trading venue always help us to have better insight and intuition to build more complex proprietary HFT strategies (Akansu and Torun, 2017).

Interested readers are referred to Hasbrouck (2007), Bouchaud et al. (2002), Potters and Bouchaud (2003), Jain et al. (2016), and Menkveld (2016) for further discussions on LOBs, and market microstructure in general. The applications of LOB models in optimal order execution are covered in Alfonsi et al. (2010), Alfonsi and Schied (2010), Bayraktar and Ludkovski (2014). Moreover, references (Avellaneda and Stoikov, 2008; Guilbaud and Pham, 2013) present applications in HFT. Furthermore, an overview on modeling of high-frequency financial data is found in Cont (2011).

III. HFT

The advances in technology and its deployment in financial sector have caused a historical disruption in several aspects of the financial activity. Warren Buffet remarked on HFT as “one of the most significant market structure developments in recent years (SEC, 2010a, b; Crippen, 2014).” This paradigm shift made leading firms of old models fall behind, losing capacity and market share as new models evolve and flourish. This “new normal” with its fierce competition and technology driven nature along with the unpreparedness of the
regulatory bodies to upgrade the rules and regulations, and fueled by various interest
groups generated public outcry and a negative social reaction against HFT for the last few
years (Lewis, 2015). The pros and cons of HFT has been extensively debated by many
people arguing different aspects of such a complex issue including Buffet who stated his
view on HFT as “It’s not a liquidity provider. It may create more volume but that’s not the
same as being a liquidity provider. To the extent that it is front running, I think society has
generally been against front running for good reasons. Here they’ve gained an advantage by
figuring out how the system worked and getting there first and that adds nothing to
economic activity. On the other hand, high-frequency trading isn’t costing the average
investor a penny. They’ve never had it so good” (Crippen, 2014). Today, most of the tradings
in the financial markets are HFT, and it is a quite matured field with its evolving business
models since the speed is a commodity now (Parsons, 2017).
Technology Advisor Committee’s working group of the US Commodity Futures Trading
Commission (CFTC) identified the following features of HFT as its unique descriptors
(CFTC, 2013):

(1) Algorithms for decision making, order initiation, generation, routing or execution,
for each individual transaction without human direction.
(2) Low-latency technology that is designed to minimize response times, including
proximity and co-location services.
(3) High-speed connections to markets for order entry.
(4) Recurring high message rates (orders, quotes or cancellations) determined using one
or more objective forms of measurement, including cancel-to-fill ratios; participant-
to-market message ratios; or participant-to-market trade volume ratios.
And, effectively, HFT is a form of automated trading, but not all automated
trading is HFT.

According to the SEC, the main characteristics of HFT firms include:

(1) their use of extraordinarily high-speed and sophisticated computer programs for
generating, routing and executing orders;
(2) their use of co-location services and individual data feeds offered by exchanges and
others to minimize network and other types of latencies;
(3) their use of very short time-frames for establishing and liquidating positions;
(4) their submission of numerous orders that are canceled shortly after submission; and
(5) their ending the trading day in as close to a flat position as possible (i.e. not carrying
significant, unhedged positions over-night) (SEC, 2010a, b).

A typical HFT trade account satisfies the three metrics: trading more than 5,000 contracts
per day (a notional value of more than $300 million in August 2010); holding end of day
inventory positions of no more than 5 percent of its total daily volume; and maintaining
intraday inventory positions of less than 10 percent of its total daily volume (Baron et al., 2012).
The literature is rich in studies on market microstructure and HFT. The SEC staff
published excellent white papers reviewing these publications. Part I entitled “Market
Fragmentation” covers reviews of papers published between 2007 and 2013 that address the
issue of market fragmentation – both visible and dark by analyzing financial market data
and reach empirical findings (US Securities and Exchange Commission, 2013). Similarly,
Part II is entitled “HFT” and discusses papers that address the particular issues of HFT.
These papers analyze non-public data sets where market activity can be attributed to
trading accounts that have been identified as engaging in HFT (US Securities and Exchange
Commission, 2014). HFT is a large subset, but by no means all, of algorithmic trading.'
For example, algorithmic trading describes a broad range of trading activity, also including the execution algorithms of large orders. This type of algorithm takes large size orders with the potential of causing high market impact and breaks them into many small size orders that are spread in time for their execution. Such smart execution algorithms to liquidate large positions through many small size high-frequency trades with the overall longer time horizons for order completion are not classified as HFT in this white paper.

Although HFT encompasses a variety of strategies, we briefly discuss a few of these methods here and the readers of more interest are referred to the references for further reading (Akansu and Torun, 2017; US Securities and Exchange Commission, 2014). Market making is one of the most common HFT strategies. It is based on submission of non-marketable orders providing liquidity at pre-defined prices, and at both ask and bid sides of the LOB. This strategy profits from the spread between bids and offers along with rebates paid by exchanges as the incentive to provide liquidity. It is noted that a traditional market maker is required to provide liquidity while an HFT market maker does not have any obligation to provide liquidity for the market. The optimal pricing algorithms of market makers consider the built-in asset price risk and information risk due to informed traders in the market. The statistical properties of LOB are tightly coupled in the development of such market making strategies. A variety of HFT arbitrage strategies are utilized by traders. An arbitrage strategy seeks to capture price inefficiencies between related and correlated products, venues, and markets. For example, one may arbitrage pricing of an exchange traded fund (ETF) against its underlying basket of stocks with the assumption that market efficiency causes mean reverting asset returns. Structural strategies attempt to exploit the built-in inefficiencies in the market, in practices of certain market participants or in certain products. As an example, traders with access to the low-latency market data and powerful processing tools may profit by low-latency (speed) trading with market participants using a venue offering executions at stale prices. Similarly, directional strategies have their variations including momentum ignition and order anticipation types. These directional HFT strategies have been a major concern with their potential to disturb the natural price formation process through their manipulative and predatory versions. For example, a momentum ignition strategy may initiate a sequence of orchestrated orders and trades in order to cause a rapid price change up or down to benefit from. This kind of manipulative (predatory) trading strategy is an illegal practice with its consequences (Akansu and Torun, 2017; US Securities and Exchange Commission, 2014). Indeed, there are claims among others that such a trading pattern caused the flash crash of May 6, 2010 that discussed in the next section.

IV. The flash crash
On May 6, 2010, trading day started with political and economic news from overseas regarding the European debt crisis. Therefore, risk premiums rose to protect against likely default by the Greek government on their sovereign debt. The Euro began its sharp decline around 1:00 p.m. against US Dollar and Japanese Yen (SEC, 2010a, b).

Report of the staffs of the CFTC and SEC to the Joint Advisory Committee on Emerging Regulatory Issues provided an in-depth analysis of the flash crash of May 6, 2010 (SEC, 2010a, b). We summarize herein the findings of the report and significant events that triggered this historical market collapse as follows:

(1) Around 1:00 p.m., broadly negative market sentiment was already affecting an increase in the price volatility of some individual securities. And, the number of volatility pauses, Liquidity Replenishment Points, triggered on the New York Stock Exchange in individual equities began to substantially increase above their average levels (SEC, 2010a, b).
By 2:30 p.m., the S&P 500 volatility index was up 22.5 percent from the opening level, yields of ten-year treasuries fell and selling pressure had pushed the Dow Jones Industrial Average down about 2.5 percent. Buy-side liquidity in the E-Mini S&P 500 futures contracts (the "E-Mini"), the S&P 500 ETF ("SPY"), the two most active stock index instruments traded in equity markets, had fallen 55 and 20 percent, respectively. Some individual stocks also suffered from a decline in their liquidity (SEC, 2010a, b).

At 2:32 p.m., in addition to already unusually high volatility and thinning liquidity, a large fundamental trader initiated a sell program to sell a total of 75,000 E-Mini contracts (valued at approximately $4.1 billion) as a hedge to an existing equity position.

This large fundamental trader chose to execute this sell program via an automated execution algorithm. It was programmed to feed orders into the June 2010 E-Mini market to target an execution rate set to 9 percent of the trading volume calculated over the previous minute without checking market price or time:

- It is expected from such an execution algorithm for large size sell program to take into account of price, time and volume information at the current market in order to keep its market impact within a justifiable range.

- The execution of this sell program was extremely fast, in just 20 minutes compared to five hours for a prior execution of the same size sell order (SEC, 2010a, b), and resulted in the largest net change in daily position of any trader in the E-Mini since the beginning of the year in 2010.

This strong sell pressure was initially absorbed by:

- high-frequency traders ("HFTs") and other intermediaries in the futures market;
- fundamental buyers in the futures market; and
- cross-market arbitrageurs who transferred this sell pressure to the equities markets through their opportunistic buy of E-Mini contracts and simultaneously selling highly correlated products like SPY, or selling individual equities in the S&P 500 Index (SEC, 2010a, b).

What happened next is best described in terms of the two strongly coupled liquidity crises. Namely, first one is at the broad index level in the E-Mini, and the second one is with respect to individual stocks as described in the following:

1. Liquidity crisis in the E-Mini (SEC, 2010a, b):
   - Between 2:32 p.m. and 2:45 p.m., as prices of the E-Mini rapidly declined, the sell algorithm sold about 35,000 E-Mini contracts of the 75,000 intended. During the same time, all fundamental sellers combined sold more than 80,000 contracts net, while all fundamental buyers bought only about 50,000 contracts net, for the resulting net fundamental imbalance of 30,000 contracts. This level of net selling by fundamental sellers is about 15 times larger compared to the same 13-minute interval during the previous three days, while this level of net buying by the fundamental buyers is about ten times larger compared to the same time period during the previous three days.
   - At 2:45:28 p.m., trading on the E-Mini was paused for “five seconds” when the Chicago Mercantile Exchange circuit breaker was triggered in order to prevent a cascade of further price declines. In that short period of time, sell-side pressure in the E-Mini was partly alleviated and buy-side interest increased. When trading
resumed at 2:45:33 p.m., prices stabilized and shortly thereafter, the E-Mini began to recover, followed by the SPY. The sell algorithm continued to execute the sell program until about 2:51 p.m. as the prices were rapidly rising in both the E-Mini and SPY:

(2) Liquidity crisis with respect to individual stocks (SEC, 2010a, b):

- The second liquidity crisis occurred in the equities markets at about 2:45 p.m. Automated trading systems used by many liquidity providers temporarily paused in reaction to the sudden price declines observed during the first liquidity crisis (SEC, 2010a, b).
- After their trading systems were automatically paused, they had to assess the risks associated with continuing their trading.
- By approximately 3:00 p.m., most securities had reverted back to trading at prices reflecting true consensus values (SEC, 2010a, b).
- During the 20-minute period between 2:40 p.m. and 3:00 p.m., over 20,000 trades (many retail-customer orders) across more than 300 separate securities were executed at prices 60 percent or more away from their 2:40 p.m. prices (SEC, 2010a, b).

Table I tabulates summary of statistics for the June 2010 E-mini S&P 500 futures contract for May 3-5 and May 6, 2010 (Kirilenko et al., 2017). Similarly, Figure 2 presents minute-by-minute transaction prices and trading volume of the June 2010 E-mini S&P futures contract on May 6, 2010 between 8:30 and 15:15 CT. Trading volume is calculated as the number of contracts traded during each minute. Transaction price is the last transaction price of each minute.

V. Lessons learned

The undesirable events with their underlying causes that triggered the market disorientation and eventually its collapse on May 6, 2010 were explained earlier in the paper. And in this section, we highlight lessons learned by all stakeholders including financial sector, governmental and non-governmental regulatory bodies and the public. We provide the summary of a potpourri of proposals made by several institutions and researchers to improve market stability and robustness in order to avoid similar crashes in the future. These developments have created public debate on HFT, and relevant laws and regulations are duly revised, accordingly, by legislative and regulatory bodies. Moreover, a clear understanding of the flash crash has generated significant know-how, and the leading regulators and exchanges around the globe have benefited from this experience to improve their operations and regulations for HFT practices.

<table>
<thead>
<tr>
<th>May 3-5</th>
<th>May 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daily trading volume</td>
<td>2,397,639</td>
</tr>
<tr>
<td>No. of trades</td>
<td>446,340</td>
</tr>
<tr>
<td>No. of traders</td>
<td>11,875</td>
</tr>
<tr>
<td>Trade size</td>
<td>5.41</td>
</tr>
<tr>
<td>Limit orders % volume</td>
<td>95.45</td>
</tr>
<tr>
<td>Limit orders % trades</td>
<td>94.36</td>
</tr>
<tr>
<td>Volatility (log high-low price range) (%)</td>
<td>1.54</td>
</tr>
<tr>
<td>Return (%)</td>
<td>-0.02</td>
</tr>
</tbody>
</table>

Table I.
Statistics for the June 2010 E-mini S&P 500 futures contract for May 3-5 and May 6, 2010

Source: Kirilenko et al. (2017)
We summarize the major observations and conclusions drawn from the chaotic market conditions of May 6, 2010 as follows:

1. The smart execution of a large size sell order can trigger extreme price movements under stressed market conditions with high volatility. The co-existence of smart execution and algorithmic trading in a typical exchange can quickly deplete liquidity and result in unstable LOB and executed trades that are not economically meaningful.

2. The flash crash was also an eye opener to remind us the integrated nature of derivatives and securities markets. The cross-market trading activity detailed above, in particular, the price discovery cross-products mimicking an index like the E-Mini and SPY was confirmed and verified by extensive investigations and research. The Joint Advisory Committee on Emerging Regulatory Issues was formed prior to the flash Crash due to the need for coordinated regulatory solution for the well observed convergence between securities and derivatives markets including the better monitoring and more judicious recalibration of the existing market-wide circuit breakers that did not turn on even on May 6, 2010.

3. Commonly, each market participant has its own trade stoppers (pauses) to withdraw from trading of a specific ticker, product, basket or other category based on the relevant measurements and automated decisions. Mostly, these withdrawals happen almost randomly and sporadically and do not impact the entire market. Whenever many withdrawals from trading of a single ticker occur, naturally, its LOB is stressed and its spread becomes economically unjustified. Moreover, if such a ticker is highly coupled with a security heavily used by many participants in the market to hedge open positions against market risk, e.g., E-Mini with SPY, then the LOB of the ticker becomes highly stressed and volatile. Hence, the market stoppers need to be devised based on the specifics and roles of tickers in the markets. “One fits for all” approach is proven to collapse LOBs and markets. Therefore, the SEC worked with the exchanges and FINRA to implement a circuit breaker to maintain LOB of each security. Such a circuit breaker stops trading across the US markets in a security for

---

**Figure 2.** Prices and trading volume of the E-mini S&P 500 stock index futures contract on May 6, 2010 in 1-minute time resolution.
five minutes if its price has changed 10 percent during the last five minutes. The SEC approved the use of these circuit breakers to securities that belong to the S&P 500 and Russell 1000 indices and certain ETFs.

(4) The financial commitment of a market participant in a specific ticker is unknown. Similarly, its decision protocols to participate or to withdraw from the trading of a ticker is proprietary. In a stressed and volatile market environment, these uncertainties increase along with the elevated uncertainty of when the ticker base market stopper will kick in to pause trading. The latter systematically impacts the first two sets of decision thresholds emphasized here that impact the available liquidity in the LOB. It is reported that the exchanges and FINRA only broke trades that were more than 60 percent away from the applicable reference price through a process that was not transparent. In order to fill the void, the SEC approved the new trade break procedures similar to the new circuit breakers.

The observations made and the experiences gained by the participants and stakeholders due to the out of the ordinary events and outcomes of the flash crash have motivated various players to come up with their proposals to address the weaknesses in the market structure and regulations for HFT. An incomplete list of the proposed procedures to improve HFT practices by various experts includes order cancellation fees, minimum order exposure times, transaction tax, requiring algorithmic trading firms to register with FINRA as discussed in Vazquez (2017), Dolgopolov (2017), McPartland (2013) and their references. In McPartland (2013), the author discusses the advantages of HFT along with the shortcomings of the current regulations where questionable HFT practices may still happen since the regulatory enforcement in the granular level of trading algorithms is almost impossible. He made nine specific recommendations for improved market microstructure to support more fair HFT spanning from having a unique identifier for each automated trading system (each individual algorithm) to be used for the audit trails to the strict prevention of receiving any trade information directly from the trade matching engines prior to its availability to all market participants.

VI. Conclusions
The HFT infrastructure of the markets in the USA has been disruptive with its built-in efficiencies for the financial sector. And, it is commonly agreed that it is also good for the public when judiciously regulated. This rapid transformation has created its own challenges. The traders, portfolio managers and investment firms along with regulators and investors were caught unprepared and inexperienced for such a paradigm shift. Although prior mini-crashes were experienced in the markets, the scale and impact of the flash crash made it historical. In this paper, we presented an overview of the flash crash and its causes drawn from in-depth investigations and studies reported by experts and researchers. The educative value of this experience is significant, and the regulations for HFT have been continuously revisited and being upgraded by the responsible bodies. Although the mechanics and the events triggered the flash crash are well understood by the experts in HFT, there are still a number of loopholes, questionable practices and business models of the exchanges and traders that need to be streamlined and addressed sooner than later.

References


Consolidated Tape Association (CTA), “Facts about the security information processor (SIP)”, available at: www.ctaplan.com/index

Consolidated Tape Association Plan/Consolidated Quotation Plan (CTA/CQ), available at: www.sec.gov/rules/sro/nms.htm#ctacq


Corresponding author
Ali N. Akansu can be contacted at: Akansu@NJIT.EDU

For instructions on how to order reprints of this article, please visit our website: www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: permissions@emeraldinsight.com
Journal of Capital Markets Studies (JCMS) is a high-quality research publication in economics and finance with a specific focus on capital markets. JCMS intends to explore the interface between academic research and practical application by disseminating new research findings alongside articles related to everyday professional practice. JCMS is a scholarly refereed journal that invites empirical or policy-related manuscripts (research paper, literature review, case studies) as well as theoretical contributions from academia. The journal welcomes quality papers from subjects including, but not limited to:

- Capital Markets and Financial Institutions
- Financial Instruments and Financial Innovation
- Investment and Financial Decision Making
- Financial and Non-Financial Information
- Corporate Reporting & Investment Decision
- Capital Market Governance & Sustainability
- Financial Standards and Regulations
- Financial Risk Management
- Islamic Finance
- Behavioural Finance
- Capital Markets Technology & Cybersecurity

Guidelines for authors can be found at:
www.emeraldgrouppublishing.com/services/publishing/jcms/authors.htm

Journal of Capital Markets Studies
Indexed and abstracted by:
The British Library
Journal of Capital Markets Studies

Volume 1 Number 1 2017

Number 1
1 Editorial boards
2 Editorial
6 Statistical studies of financial reports and stock markets
9 Shyam Sunder
10 Understanding sustainability for socially responsible investing and reporting
Marianne Bradford, Julia B. Earp and Paul F. Williams
36 Leveling the playing field for less-sophisticated non-professional investors: does plain English matter?
Eileen Taylor and Jennifer Riley
58 The performance of US-based emerging market mutual funds
Halil Kiyimaz and Koray D. Simsek
74 Identifying the major reversals of the BIST-30 index by extreme outliers
Yon-Gun Lee, Sohyoun Lee, Mie Kyung Jae and Hyang Ran Jeon
89 The flash crash: a review
Ali N. Akansu

Access this journal online:
www.emeraldinsight.com/journal/jcms