Journal of Business Research xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Journal of Business Research

A model for unpacking big data analytics in high-frequency trading☆

Jonathan J.J.M. Seddon, Wendy L. Currie *

Audencia Business School, 8 route de la Joneliere, Nantes Cedex 3, France

ARTICLE INFO

Available online xxxx

Keywords:
Big data
HFT
Latency
Asymmetry
Strategies

ABSTRACT

This study develops a conceptual model of the 7 V′s of big data analytics to gain a deeper understanding of the strategies and practices of high-frequency trading (HFT) in financial markets. HFT is computerized trading using proprietary algorithms. Empirical data collected from HFT firms and regulators in the US and UK reveals competitive asymmetries between HFTs and low-frequency traders (LFTs) operating more traditional forms of market trading. These findings show that HFT gains extensive market advantages over LFT due to significant investment in advanced technological architecture. Regulators are challenged to keep pace with HFT as different priorities to the 7 V′s are given in pursuit of a short term market strategy. This research has implications for regulators, financial practitioners and investors as the technological arms race is fundamentally changing the nature of global financial markets.

1. Introduction

Financial markets have become increasingly fragmented characterized by globalization and technological change (Funk & Hirschman, 2014). Three important activities include: *fund management* for long term investors; *low-frequency trading* (LFT) by the traditional brokers; and *high-frequency trading* (HFT) by proprietary financial firms (Blocher, Cooper, Seddon, & Van Vliet, 2016). An important research challenge is to unpack the big data implications of HFT as industry and academics question whether HFT has positive or negative impacts on financial markets (CFTC-SEC, 2011; Cooper, Davis, & Van Vliet, 2016).

In 2012 algorithmic trade instructions sent by both LFT and HFT accounted for over 1.6 billion shares every day (Shorter & Miller, 2014, p. 14). HFT is a form of algorithmic trading which relies on advanced technological infrastructure to compete on speed, rapid turnover rates, and high order-to-trade ratios as they leverage vast amounts of financial data (Aldridge, 2013). While big data is defined as, "data sets with sizes beyond the ability of commonly used software tools to capture, curate, manage, and process data" (Snijders, Matzat, & Reips, 2012) this study extends the concept through empirical research on how modern HFT firms deploy big data analytics in their market strategies to compete with other HFT's and more traditional forms of financial trading carried out by LFT's.

For this special issue on *Big Data and Analytics in Technology and Organizational Resource Management*, the theoretical model on the 7 V′s of big data has been utilized to provide further illustration of the HFT phenomenon. So far, much of the academic literature on HFT is found in the finance field, but an opportunity exists for business researchers to address the gap "between the academic research on HFT and its perceived impact on markets in the public, media and regulatory" fields (Gomber, Arndt, Lutat, & Uhle, 2011, p. 1). By applying each of the 7 V′s to HFT supported by empirical data collected from HFT firms in the US and UK, this research is relevant for regulators, industry practitioners and investors, as these groups continue to question whether HFT has positive effects on financial markets, or negative impacts on traditional forms of market trading (Cooper et al., 2016).

The relationship between HFT and big data is crucial as the material features of the technology, coupled with the trading algorithms in the software, help to define how judgments are made in financial markets (Angel & McCabe, 2010). So far, the main body of academic work on HFT is found in finance, with many studies using mathematical models to simulate HFT activities. For example Hoffmann (2014) shows how HFT sequential bargaining disadvantages the slower LFT traders whilst Aït-Sahalia and Saglam (2013) illustrate how HFT exploit data asymmetries to disadvantage other traders. Other academic studies show that HFT has a positive effect on financial markets and liquidity, reducing the probability of adverse selection when a trade does not occur at a fair market price (Brogaard, Hendershott, & Riordan, 2014; Brogaard, Hagströmer, Nordén, & Riordan, 2015; Boehmer, Fong, & Wu, 2013; Menkveld, 2013). A recent study looked at every message sent to the NASDAQ exchange for the S&P 500, and concluded that the

☆ The authors are grateful to Audencia Business School, Nantes, France, for a grant to study big data and high-frequency trading in financial markets. Thanks are also given to Professor Ricky Cooper and Professor Ben Van Vliet (Stuart School of Business, Illinois Institute of Technology) for comments made on prior drafts of this article.

* Corresponding author at: Audencia Business School, 8 route de la Joneliere, Nantes Cedex 3, France.
E-mail addresses: jseddon@audencia.com (J.J.J.M. Seddon), wcurrie@audencia.com (W.L. Currie).

http://dx.doi.org/10.1016/j.jbusres.2016.08.003
0148-2963/

high levels of order cancellations were essential for HFT firms to establish a true market price (Blocher et al., 2016).

In this paper the generic literature on big data has been situated in the context of HFT as a sub-set of algorithmic trading in financial markets. While many contributions discuss big data in the business, organizational and management literature, more empirical work is needed to provide theoretical insights and analysis in specific business sector contexts (George, Haas, & Pentland, 2014). The HFT model of the 7 V′s of big data illustrates how HFT embodies a paradigm shift in the financial markets fueled by deregulation and unprecedented technological change. Unlike the LFT which represents traditional financial trading, HFT is speed and data-intensive. The architecture used to process data, the speed of execution, software tools used and how orders are generated from complex mathematical modelling fundamentally differentiates them from other traders (Aït-Sahalia & Saglam, 2013). For example, some HFT strategies are based on strategic sequential trading in event time (Easley, Lopez de Pradom, & O'Hara, 2012), driven by volume clocks (buy 1000 securities and then sell them all) rather than the time clock which underpins LFT (sell 100 securities every minute until 10,000 have been sold). The big data implications are that HFT collect trillions of trade records to process real time events to identify LFT trading activity, giving them a technological and time advantage over their much slower competitors.

This paper is structured as follows. First, a review of the information systems and management literature on big data in financial markets is presented. Research on big data is attracting interest from management disciplines including, information systems (Agarwal & Dhar, 2014; Baesens, Bapna, Marsden, Vanthienen, & Zhao, 2014; Constantiou & Kallinikos, 2015) accounting and finance (Bhimani & Willcocks, 2014; Bennett, 2013) and general management (George et al., 2014). Having built this model, a discussion of methods used and data collection is then given. Primary data is collected from in-depth interviews with multiple informants from HFT firms, regulators and industry analysts. Secondary data is collected from reports, articles, websites, conferences and other relevant material on HFT strategy and practice. The model of the 7 V's of big data in relation to HFT firm strategies is then discussed and analyzed. Finally, the implications of this research for practitioners is considered with suggestions for potential areas of future business research.

2. Big data analytics and high-frequency trading

The origins of the term, big data emerge during the 1990s with one of the first academic articles appearing at the end of the decade (Bryson, Kenwright, Cox, Ellsworth, & Haimes, 1999; Lohr, 2013). Today, interest in big data spans numerous industrial and not-for-profit sectors (finance, healthcare, manufacturing, social media) and offers new opportunities and challenges for researchers across several fields (Agarwal & Dhar, 2014; Baesens et al., 2014). A recent editorial in a leading management journal noted that big data is now widely used in the business community, but "there is very little published management scholarship that tackles the challenges of using such tools – or, better yet, that explores the promise and opportunities for new theories and practices that big data might bring about" (George et al., 2014, p. 321). As a topic for research enquiry, big data is still new and has so far generated significant interest from business analysts, management and IT consultancy firms and government agencies.

A UK government report looks at three areas around big data. (1) Defining big data. (2) High level trends in big data. (3) Opportunities for big data applications. A working definition is also provided, "Big data refers to both large volumes of data with high level of complexity and the analytical methods applied to them which require more advanced techniques and technologies in order to derive meaningful information and insights in real time" (HMG, 2014, p. 2).

Despite the potential opportunities for researchers, the theoretical and methodological challenges surrounding big data need to be considered in the context of prior work, to determine whether this is indeed a new topic worthy of research enquiry (and grant funding) or just the latest fad to generate more business for the consulting industry. Observations from the extant literature show that the multiple definitions of big data lead to confusion rather than clarity. Common examples of big data companies are Amazon, Facebook and Google, where big data analytics services enable clients to analyse large data sets using cloud computing (HMG, 2014). While these companies are relatively new in the digital age, more traditional examples include banks, with several decades of experience in managing large volumes of complex, structured and unstructured data (IBM, 2013).

Business and information systems research has a long history of examining data analytics where technological changes from mainframe computers, personal computers and, more recently, the Internet present business managers with new data and information management challenges (Yoo, 2015). As an emerging phenomenon, big data arises from technological changes that facilitate the collection, storage and transfer of data, and also how individuals, or agents, create, interpret, analyse and manipulate data as part of the business strategy toolbox (Woerner & Wixom, 2015).

In the financial markets, the increased momentum to move away from manual actions to full electronic trading following the "big bang" in 1986 has generated studies on the use of data by fund managers and traders (Preda, 2007a,b), trading technologies (De Goede, 2005), the global coordination of currency traders (off and online) to interpret pricing screens (Knorr-Cetina & Bruegger, 2002), the stock ticker as a recording technology (Preda, 2006), visualization software to present complex market data (Pryke, 2010) and technologies to facilitate both trading conducted in open outcry pits and electronic trading (Zaloom, 2003).

The literature review reveals that "data analytics" and especially, "big data analytics" is under-theorized and empirically under-represented in business research. Big data in financial markets covers many relevant areas including regulation and compliance, global trading strategies and infrastructure, transactions between institutions, networks and firms, risk management and trading algorithms (MacKenzie & Millo, 2003). Over the past three decades, the pace, volume and origin of change in financial markets is unprecedented (Wheatley, 2014). Computerization has been at the forefront of these changes with the emergence of HFT, digital money transfers, payments technology, peer-to-peer finance, and portfolio analysis, all of which rely on the speed and accuracy of data flows in an increasingly networked and international financial industry (MacKenzie, 2006).

Recent calls invite researchers to examine the micro-foundations of big data in business and organizational strategies or behaviours in real time rather than snapshot quarterly outputs (George et al., 2014). This research theorizes the concept of big data in the context of financial markets. The relationship between data and finance has been a perennial topic in business research for decades. Prior work looks at data and information for financial trading (Clemons & Weber, 1990; Weber, 1999, 2006) data analytics and inter-organizational standards in the mortgage industry (Markus, Steinfield, Wigand, & Minton, 2006). Yet today's financial markets have become increasingly fragmented and complex, with exchanges no longer having the monopoly on trading specific securities. In the US, for example, financial trading may occur at any of the exchanges, alternative trading systems (ATS), electronic communication networks (ECN) or dark pools (Shorter & Miller, 2014; White, 2014).

Although no legal definition for HFT exists opinion is generally focused around trading using proprietary algorithms (NASDAQ, 2016), not owning any security at the end of the day, sending orders which are quickly cancelled (sub-second) or a having a low ratio of executions to orders sent (ESMA, 2014). Typically, three types of organization are regarded as behaving in such a way: proprietary trading desks of broker-dealers and banks, independent proprietary trading firms, and

Please cite this article as: Seddon, J.J.J.M., & Currie, W.L., A model for unpacking big data analytics in high-frequency trading, *Journal of Business Research* (2016), http://dx.doi.org/10.1016/j.jbusres.2016.08.003