UNIVERSITY OF SOUTHAMPTON

FACULTY OF BUSINESS, LAW AND ART

Southampton Business School

Employing web-based information in financial decision-making

by

Lawrence Miles Green

Thesis for the degree of Doctor of Philosophy

March 2018

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF BUSINESS, LAW AND ART

SOUTHAMPTON BUSINESS SCHOOL

Doctor of Philosophy

Employing web-based information in financial decision-making

By Lawrence Green

This thesis, which is divided into three papers, explores the use of web-based information in financial decision-making and identifies how web information has improved forecasting. New online information is easily accessed and constantly available to the public, potentially enabling decision-makers to make decisions that are more accurate. The academic literature has proclaimed that the web has transformed decision-making but there is little understanding of how increased information availability and transparency can lead to improved forecasting accuracy and enhanced decision-making.

The three empirical papers herein exemplify how web-based information can be employed in decision-making models related to financial markets and particularly, speculative markets, to show the added value of web-based information in decision-making models in a real-world setting.

In order to understand how web-based information affects decision-making, this

thesis is separated into three papers. The first paper explores how new geospatial information improved forecasting accuracy of performances of racehorses and how quickly unprecedented information derived from new Information Technology (IT) is discounted at the aggregate market level. The second paper shows how distance information, which is freely available and easily accessed over the internet, helps explain the decision-making behavior of experts and novices, highlighting how expect knowledge can be elicited from trainers to improve forecasting accuracy. The third paper examines how the sentiment in online news information affects individual-level decision-making behavior and performance.

Taken together, the three papers provide empirical analysis exemplifying how online information can improve forecasting in the real world. The results of the three papers have important contributions to the literature. Paper 1 highlights market convergence with respect to geospatial information and the horse race betting market, showing that improved web-based information availability provides unprecedented information to improve forecasts and ultimately, how the market adapts to this information becoming efficient. Paper 2 identifies how distance information informs the behavior of distinct sub-groups of decision-makers (experts and novices) and, how the elicited knowledge from experts improves forecasting decisions for a limited time before the betting crowd discount such information. Finally, in contrast to the majority of literature on how market prices respond to online information, paper 3 isolates the effect of sentiment on individual behavior, showing how individuals act in a sentiment contrarian fashion providing fine-grained analysis of the effect of online information at the individual level.

This thesis shows how improved access to online information improves forecasting abilities at various levels by showing how web-based information is discounted at the aggregate market level, how distance information informs expert and novices behavior, and how information affects individual behavior and performance.

The web has transformed decision-making and this thesis exposes the benefit of web information to improve forecasting accuracy. Online information can improve forecasting and, the rate at which the information diffuses into financial markets is an important research area as new information becomes available and markets constantly adapt to such information.

Table of Contents

List of Tables

List of Figures

Academic Thesis: Declaration of Authorship

I **Lawrence Green** declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

Employing Web-based information in financial decision-making

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;

2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;

3. Where I have consulted the published work of others, this is always clearly attributed;

4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;

5. I have acknowledged all main sources of help;

6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;

7. Part of this work has been presented at:

   Green, L., Ma, T, Johnson, J.E.V., Sung, M and Tiropanis, T. “Determining the Value of Web-based IT in Financial Decision Making”, *2016 World Conference on Information Technology*, Bali, Indonesia, 10-12 September, 2016.

   Green, L., Ma, T, Johnson, J.E.V., Sung, M and Tiropanis, T. “The economic value of information derived from a new web-based technology and its rate of diffusion in a financial market, *2016 Annual Meeting of the Decision Sciences Institute,* Austin, Texas, November 19-22, 2016.

Signed: …………………………………………………………………………………….

Date: ……………………………………………………………………………………

Acknowledgements

I would like to thank my friends and family for their continual support throughout.

I am massively grateful to my supervisors Johnnie Johnson, Tiejun Ma, Ming-Chien Sung, Thanassis Tiropanis and Wendy Hall for their invaluable advice and time during my PhD.

This work could not have been done without the Web Science DTC that provided funding[1].

I have been incredibly lucky to have the opportunity to pursue a PhD and I have given it my all.

[1] The Digital Economy Theme is a Research Councils UK cross council initiative led by EPSRC and contributed to by AHRC, ESRC, and MRC. This work was supported by the EPSRC, grant number EP/G036926/1.

Chapter 1 Introduction

The World Wide Web (web) has revolutionized how individuals use information and how information spreads. New Information Technology (IT) has arguably been one of the most important drivers of economic and social value in the last 50 years transforming organizations, markets, industries, societies and the lives of individuals (Clemons et al. 2013). Businesses can create and capture value from IT to improve their existing strategies and create economic profit (Drnevich and Croson 2013): data-driven companies are reported to be 10% more productive than companies that fail to use their data but most companies are only using about 12% of their data (Science and Technology Committee 2016). The widespread impact of new information on the web was predicted to contribute £216 billion to the UK economy from 2012-2017 or 2.3% of GDP (Science and Technology Committee 2016). More long-term estimates predicted that the world-wide data technology and services market will grow 23.1% annually to $48.6 billion in 2019 (Nadkarni and Vesset 2015).

Transformation in web-based information availability has increased both the amount and diversity of information that can be used in forecasting models. Proliferation of information from social media, cloud computing and mobile data has created an abundance of detailed information that individuals have access to which drives faster and more effective decision-making (Bharadwaj et al. 2013). New methodologies are needed to make sense of the heterogeneous and fragmented data which exist in varying formats (text, image, sound and quantitative) and is largely unstructured web-based data (Agarwal and Dhar 2014).

The transformation in web-based information availability is so great that researchers have defined it a paradigm shift that opens up opportunities for new research directions (Chang et al. 2014). Within this paradigm, one of the key specific research directions, as identified by Chang et al. (2014), is the emerging capabilities to collect data from the real world to enable insights into the transformational aspects of IT at the aggregate market and individual level. Indeed, event and technology-focused analysis opens up new applications of research methods to capitalize on the plethora of data available, which in turn, opens up novel insights for business. As increasing amounts of information become available from technological advancements, significant opportunities arise for research on the impact of technology on forecasting.

Specifically, the value of new IT is an important research area as organisations adopt innovative technologies (Drnevich and Croson 2013; Ren and Dewan 2015; Tambe et al. 2012). While the majority of literature has focused on the impact of IT on organisations and firms, this thesis contributes by exploring how IT affects decision-making at both the market and individual level. Previous research on technology adoption has been primarily driven by users' perceptions of a technology and largely ignored a user's actual experience with and actual use of a technology (Aggarwal et al. 2015). In order to understand the value of new web-based technology, research must move beyond initial perceptions of technology and explore actual use in an empirical setting.

The financial market literature provides a useful framework to explore the use of information within empirical analysis. The efficient market hypothesis (EMH), and the idea that a market is efficient in the sense that the stock prices adjust rapidly to new information (Fama et al. 1969), provides a means to test to what extent new information available on the internet is reflected in prices. A central tenet of the EMH is the testing of 'the speed of adjustment of prices to *specific kinds* of new information' (Fama et al. 1969, p. 1). Originally, the notion of efficiency was related to the speed of adjustment of market prices to textual news information related to stock specific companies, financial reports, and security issues and, whether prices fully reflect available information at any given time (Fama 1970).

As the web transforms information availability, the diversity of new information enables more types of information to be tested in lieu of efficiency, for example geographical and social media data. The proliferation of information online and the various types of data offer fresh complexities to the EMH with an abundance of textual information to incorporate into prices as well as images, search trends and click data (Yu et al. 2018).

Although the EMH suggests that prices are instantaneously and unbiasedly incorporated into prices, in practice prices tend to adjust to new information after a certain amount of time and traders are able to exploit temporary profitable opportunities as prices ultimate push towards efficiency (Reboredo et al. 2013). The time period during which information is made available and is fully incorporated into the market is known as market convergence and, information availability and distribution channels are critical to market convergence (Nassirtoussi et al. 2014). Therefore, research on the distribution channels used to share information on the web and, how new information becomes available is necessary in light of the market convergence literature. As technology advances and improves the rate at which information is released, our abilities to capture and analyse such data improves, resulting in markets becoming more efficient. Information transparency, defined as the level of availability and accessibility of market information to participants, has a fundamental impact on markets (Yang et al. 2015). The amount of information available is highly correlated with market efficiency and forecasting performance. Since information availability if perpetually increasing, further research is needed to understand how the markets converge and develop theoretical basis of how the internet is linked to market efficiency.

Market convergence occurs at varying speeds depending on the form of information (distribution channel) and the market. Diffusion is the process by which new information is shared between participants: ‘a process of convergence […] as two or more individuals exchange information in order to move towards each other in the meanings they ascribe to certain events' (Rogers 1995, p. 6). The diffusion of technical innovations is a process of convergence where individuals gain new information that removes some level of uncertainty in a situation where a choice of alternative exists (Rogers 1995). There are four elements which affect diffusion of innovations: the innovation itself (or the new information technology), which is communicated using certain channels, over time, to participants of a system (Rogers 1995). As such, information diffuses at different rates depending on the composition and interplay of these aforementioned elements. That is to say, the same technology will diffuse differently in one market system compared to another market depending on the market participants. Similarly, the same innovation will diffuse differently depending on the communication channels used – for example, information in text format may be more readily accessible and interpretable than images.

There are no set rules that dictate how quickly, and to what extent, information will diffuse. Some researchers have found that faster-diffusing information is associated with quicker and less noisy profits, while increasing competition impounds more information into prices, eroding profits (Manela 2014). The value of information and how fast information can be incorporated into the market is a contemporary and important research area. Although the internet has radically changed financial markets, the literature explaining the theory behind such transformation is inadequate (Zhang and Zhang 2015). IT improves information access and reduces costs of trading. Understanding new technology and the people who use them will provide insight into the market itself (Boehmer et al. 2008).

As information access and availability increases though, complications arise because there is more information than could be processed by any one individual. With the increasing amounts of data available, a potential challenge arises as information overload occurs resulting from diverse data sources, multiple data formats and data volume (Chen et al. 2012). Indeed, the idea that investors are rational investors and able to incorporate all information is at sharp contrast with 'bounded rationality' (Simon 1955). Simon (1955) suggests that individuals are limited by their cognitive resources and instead of making optimal decisions that are costly and timely, individuals alternatively make decisions that are satisfactory based on the bounds of rationality and the amount of information they are able to process at any one time. Such a model of human behavior as being limited in their information processing capabilities fits with how individuals make decisions using information available online. The cost and time to process all information available on any subject is such that individuals will process a small fraction of information and suffice in making a decision based on the limited knowledge they have. The idea of investors being perfectly rational is dubious considering the complexity of web decision-making and, as one of the necessary conditions of efficient markets, the lack of rational investor poses a theoretical debasement to the EMH.

In order to account for some of the short comings of the EMH, the adaptive market hypothesis (Lo 2004; AMH) was developed. The AMH extends market efficiency theory to suggest that prices reflect as much information as dedicated by the combination from environmental conditions and the number and nature of market participants in the economy (Lo 2004). Markets adapt to changing environmental conditions, for example, with increasing technological advancements more data is available of various formats, making markets more efficient over time. Innovations make available new information and profitable opportunities arise as the market adjusts to the new information. In this way, 'markets are never completely efficient or irrational – they are simply adaptive' as the market participants learn to adapt and use new practices that are better suited to the economy (Lo 2012, p. 11). Exploring how the AMH helps contextualize market efficiency in the era of the web helps underpin how technology has changed since Fama's inception of the EMH.

The internet has made traders more informed with constant access to up-to-date information that is freely available (Rubin and Rubin 2010). Individuals learn, adapt and evolve to the new environment and bounded rationality offers a means of explaining that the market is as efficient as its component participants who are able to process limited and select information from the myriad available on the internet. As individuals learn and evolve new practices that are better suited, the forecasting performance gradually improves. However, the nature of competition is such that new best practices become the level benchmark necessitating new approaches. In this fashion, the AMH captures how markets continually adapt to the extant information and evolving processing capabilities of market participants.

In order to understand how markets learn to adapt, it is necessary to understand the distinct sub-populations of traders that exists in the demand side of financial markets and how each sub-group behave. Market participants can be broken down into various groups which typify trader behavior, including experts and the crowd. Experts are classified as market participants who have significant knowledge of the market in which they are trading, and have considerable skills and experience that makes them qualified decision-makers in their market (Van Wesep 2016). The notion of experts is similar to the idea of professional bettors as experts typify a heightened level of skill which is shared with professional bettors (Bruce et al. 2012). There are further similarities between experts and informed traders in that part of their advantage can come from the ability to analyse publically available information (Engelberg et al. 2012). As such, the aforementioned literature recognises a certain group of traders, typified by their expertise and skill, and existing across a number of financial and betting markets. The expert group are theoretically identified by their ability to use information available to them and produce superior forecasts to the market.

Whether experts are able to achieve superior accuracy compared to the market is disputed in the literature. There is conflicting evidence on the accuracy and performance of expert forecasters and various examples have shown that experts may not be outperform non-experts in forecasting (Lawrence et al. 2006). The conflicting research results found in the literature (Önkal et al. 2003) raise doubt about the practical abilities of experts to achieve superior forecasts in practice despite their heightened ability and access to information.

At the same time, recent theory has identified 'the wisdom of crowds' (Surowiecki 2005) which proposes that the crowd is able to predict just as well as any single expert. This phenomenon provides an interesting comparison in that the average estimate of the crowd is as accurate as an expert's estimate. Research has shown that the wisdom of the crowd is such that crowds achieve similar and even superior forecasts to experts in the real world (Budescu and Chen 2014; Mollick and Nanda 2016).

Therefore, the quoted literature offers a conflicting view of whether experts actually fail to achieve superior forecasts in practice, or whether the crowd are actually skilled forecasters themselves due to the dynamics of aggregating individual forecasts.

Comparing the market aggregated decisions of the crowd with decisions from experts is necessary as crowds and experts differ in their ability to interpret information (Chen and Zeng 2016). Research identifying how experts and the crowd use web-information is well placed to contribute to the literature on the existence of distinct sub-groups and their varying characteristics, which further defines the characteristics of their sub-group.

Betting markets have many similarities to financial markets, making betting markets the ideal experiment to offer new insight into financial behavior of market participants and the market as a whole. Betting markets provide the ideal context to explore how the distinct sub-groups interpret information and the various behaviors which typifies each group (Andersson and Nilsson 2015). Both betting markets and financial markets permit money to be wagered on future and uncertain events where the price of securities is representative of the complex interdependence of factors that determine and asset's value (Sung et al. 2012). Betting markets consist of distinct sub-groups of bettors who differ in terms of how informed they are and analysis of which offers valuable insight into financial market contexts on how such sub-groups differ in terms of the aggregate performance and risk-attitudes of each group which has an underlying effect on the market efficiency (Bruce et al. 2012).

One advantage of betting markets if that the eventual outcome and final state of the security is resolved and the degree to which information was fairly reflect in the price can be estimated. As such, betting markets have been the subject of a large body of research that measures the extent to which betting markets are efficient, and the composition of market participants present in the market that influences the operation of betting markets (Lessmann et al. 2011; Oikonomidis et al. 2015).

The British horse race betting market provides a useful experimental set-up to explore how markets use information and the behavior of the sub-group populations within the market. Estimates show that the horse race betting market had turnover in excess of £12 Billion in 2015. [2] The horse racing betting market is determined by a wide array of factors that make it an interesting and insightful area to test financial market theory. Different markets have varying levies enforced that stem from the operating levels of various sub-groups within the market depending on the racetrack location, day of the week and type of race (Bruce et al. 2014). Races themselves are determined by a complex interdependence of factors relating to the horse, the jockey and the track amongst over things, which allow for testing market (information) efficiency (Lessmann et al. 2011).

This thesis demonstrates the impact of web-based IT in decision-making and the economic value of such information in the real world. Ultimately, research has tried to explore how online information affects decision-making behavior but there is a research gap

[2] http://www.horseracingintfed.com/resources/Annual_Report_2015.pdf

in the literature concerning how decision-making behavior at the individual level is affected (Lillo et al. 2015). Furthermore, a survey of firm level research has identified that the web has had a significant effect in the findings reported in studies, with the internet era being associated with significant increase in business value (Sabherwal and Jeyaraj 2015). Collectively though, current research fails to highlight the transformational impact of the internet in decision-making and how new IT affects the underlying decision-making processes of individuals. As such, novel research is emerging on how the internet transforms individual decision-making and the wider financial markets (Xu and Zhang 2013).

This thesis offers new understanding of how IT improves decision-making performance by exploring how information creates value through superior forecasts. Moving beyond technology adoption and initial perceptions, the analysis demonstrates the impact of new IT and highlights the use of information from a technology in decision-making related to financial markets in order to estimate the value of new IT. It is fundamentally important that web information is used and used efficiently in order to realize the value of IT to the UK and global economy (Science and Technology Committee 2016).

The three studies provide new contributions on how information from IT diffuses through the market place and how information affects distinct sub-group decision-making. By focusing on a specific technology as the unit of analysis, the three grounded empirical studies draw implementation-orientated evidence of the rapid rate of change in information systems relative to overall business practices (Drnevich and Croson 2013). That is, the three papers describe how three different technologies are transforming decision-making processes and how these improved processes affect the wider business market. Since previous literature has failed to demonstrate how technology transforms financial markets and decision-making (Xu and Zhang 2013), this thesis provides import contributions that are centred on the use of technology in a market context. By focusing on the transformational impacts of IT in decision-making in relation to financial markets, and by exploring the practical and economic consequences of IT enhanced decision-making, this thesis highlights how technology is contributing to efficient markets.

Paper 1 explores information diffusion, the rate at which information derived from new IT spreads through a market. Information from (geospatial) web-based IT allows individuals to solve complex decision-making tasks with real monetary reward/penalty. However, discovering the economic value of information is a challenge as new IT are adopted and used in new ways to enhance value creation (Melville et al. 2007) and, thus, decision-makers will inevitably have a learning curving for this information discovery process. Therefore, the first paper, "**The economic value of information derived from a new web-based technology and its rate of diffusion in a financial market**", addresses the theme of information diffusion and economic value of information in a market context. In particular, showing how Virtual Globes (VG) technology such as Google Earth can be used to derive the topology of racetracks, which is very difficult to achieve without IT aids. The layout of a track influences how horse's run and therefore, the outcome of a race, providing a new type of information that was not available before (Self et al. 2012). Tracking how this information became incorporated into the betting market reveals how markets become more efficient throughout time and how markets adapt to new information.

Analyses uses 18 years of data (1997 to 2014, inclusive) from the UK horse racing markets, with 75,750 races (incorporating 76,406 different horses) run at all 34 UK racetracks. This period encompasses the release of VGs. Geocontext[3] was used to collect the elevation data at each of the 34 racetracks providing 300 different courses to estimate the

[3] www.geocontext.com

topology.

The results show how this geospatial information provides superior estimates compared to public odds to estimate a horse's probability of winning and therefore, the value of this information in a financial market context. Furthermore, by showing how quickly the market adjusts to this new information, the results show how quickly markets incorporate information from new IT and how markets adjust over time. The paper explores how information from VGs can be used in decision-making models, highlighting the link between new IT and economic value. The findings contribute to the literature by providing a concrete real world example of market convergence, using VGs to improve financial decision-making in discrete choice situations characterised by a high-level of competition and uncertainty.

This paper offers three contributions to the literature. i) Demonstrating how information derived from VG technology can be used to improve predictive accuracy of performances of racehorses. This is a methodological contribution as the paper documents how value can be extracted from an information source and how such information supports decision-making processes. VGs provide unprecedented IT-enabled information which show how information availability can be used to improve decision-making performances. ii) The rate at which technology-enhanced information diffuses through a financial market is depicted. Improvement in prediction scores may be diluted over time as the information becomes public knowledge to the market. Results show that information from VGs is not immediately diffused through a financial market but rather takes a number of years to be fully discounted by the public suggesting the market are not efficient in the sense that they completely and immediately incorporate all new information, but adapt to a changing environment. iii) The clear link between technology-enhanced decision-making and economic value is highlighted. Often it is difficult for economists to quantify the economic value of accurate forecasts (Lessmann and Voß 2017) but this paper highlights the direct link between exploiting new information from IT in making decisions and the relative profitability in a market context. Furthermore, the longitudinal analysis shows how economic value changes over time, a finding that has not been made explicit in the literature (Papagiannidis et al. 2015), and is fundamentally important in future research exploring the economic value of improved forecasts.

Understanding how individuals use information available online and behave in the real world (outside the laboratory) is important since there are an increasingly wide range of decision aids available to decision makers (Kocher et al. 2013). Individuals frequently make decisions using online tools such as Google Maps to calculate the best route in journey planning (Constantiou et al. 2014) and these decisions will involve distance-based trade-offs. As such, teasing out the underlying role of distance in decision-making reveals the fundamental aspects of human behavior. Although this distance related information might be known by aficionados and professionals, only in recent years has distance information become more transparent and accessible to the public via the web.

Using geolocation and distance related information will become increasingly important as mobile technology both empowers consumers to find (local) information and businesses are able to use location data to provide highly targeted content (Chen et al. 2012). Distance information available online will enhance research on decision-making and belief elicitation – that is inferring subjective beliefs based on individual behavior.

Thus far, no research has explored how access to distance information informs the decision-making of experts and the betting crowd, and how elicited expert knowledge can improve forecasts in the context of horseracing. In belief elicitation, conventional methods present respondents with a range of options and require them to choose their preferred observation (Schlag et al. 2013). Research on eliciting subjective beliefs fails to take into account that respondents do not always behave as they themselves state even when there are incentives (Schotter and Trevino 2014) and as such, it is problematic to assume that subjects will behave according to their own stated beliefs or intentions (Costa-gomes and Weizsäcker 2008). One assumption in the extant research is that elicited beliefs are consistent with behavior. Observing individual actions as opposed to stated intentions enables subjective beliefs to be inferred. As such the second paper, '**How new distance information that is readily available online can be used to elicit knowledge from experts and produce improved forecasting accuracy compared to the betting crowd**', explores how new distance information can be used to elicit knowledge from the behavior of experts to obtain more accurate probability estimates.

Observing UK horse race trainers over a period of 12 years, and analysing the decisions they made on which races to travel to and compete in, reveals their (expert) subjective beliefs. Data from the UK horse racing market was collected involving 82,703 races (involving 84,939 different horses) run in the UK between 1998 and 2010 with 495 active English and Welsh trainers. Postcode level data was also collected for the 495 active trainers to calculate the distance from each trainer to each racetrack venue. Over this period, if all trainers had only sent their horses to their nearest racetrack when their horses ran, then they would have collectively sent their horses 10,153,979km in the period 1998-2010 rather than the 61,866,824 km they actually travelled, showing that the extra distance can be used to understand their decision-making.

Paper 2 shows how these professionals make decisions and the betting public can elicit subjective beliefs and update their prior probabilities. In uncertain decision-making tasks objective probabilities may be difficult to ascertain and therefore, expect knowledge elicitation is important for forecasting (Bolger and Wright 2017). This paper exemplifies a technique to elicit subjective information from experts (i.e., horse trainers), and how this can be combined with objective information to produce more accurate probability estimates. As such, paper 2 contributes to the literature threefold: i) how new web-based information from Google maps can be used to inform decision-making of experts and highlight how they make distance-based trade-offs under uncertainty. ii) How distance information can be used to elicit expert knowledge from the actions of horse race trainers. iii) How chronicling the rate at which distance-based information becomes incorporated in the aggregate decision-making of the betting-crowd shows that the crowd learn to improve their forecasts in relation to distance information.

Consequently, this paper contributes to the growing forecasting literature and addresses the growing need to understand how new web-based information can be used to improve forecasting accuracy (Kim et al. 2015; Schneider and Gupta 2016; Yu et al. 2018). As well as using distance information to inform and contextualise the decisions made by experts and novices, the paper also shows exemplifies eliciting expert knowledge from real world experts. Expert knowledge elicitation is an important pursuit in forecasting, whereby the ability to directly extract predictive information from experts serves as a useful tool to improve forecasting accuracy (Bolger and Wright 2017). There are two basic strategies that have emerged: one is to anticipate future events by extracting information from wise crowds such as those in financial prediction markets, and the other is to extract information from a small group of experts (Bolger and Wright 2017). This paper explores both strategies simultaneously by exploring how expert elicitation can improve forecasting accuracy but also, how the crowd learn to discount complex elicited information over time.

Extracting expertise from experts is an important endeavour for forecasting as it has been shown that expect knowledge elicitation lies at the core of judgemental forecasting and there are important issues of expert knowledge elicitation in a real world forecast setting (Alvarado-Valencia et al. 2017). Experiments to date have been mostly conducted in the laboratory, devoid of experts and in such cases, the act of observing ‘experts’ may even influence how they respond, biasing the results and offering little relevance to behavior in the real world (Werner et al. 2017). By observing actual horse race trainers and the decisions they make in relation to which races to enter, all of which incur various costs and impact the probability of success, paper 2 overcomes these aforementioned shortfalls in the forecasting literature.

The wider implications of these contributions are important not just in the context of the forecasting literature but also online search and purchase behavior where individuals frequently use mobile devices to access local information (Cachon 2014). For organisations to actualise the 60% increase in operating margins using location data (McKinsey Global Institute 2011), new research must illustrate how individual behavior is affected by distance and how increased information availability through geographical mapping services can be harnessed for decision-making.

Finally, paper 3 focuses on the effect of online news information on individual level decision-making. The internet has transformed decision-making particularly for financial traders who frequently use the web to find new information and to share their own ideas (Sabherwal et al. 2011). Previous research has assumed that markets respond to online news information directly. However, evidence suggests that information affects individual traders who, in turn, drive price changes (Peress 2014). Research at the individual level is warranted to elaborate on how individuals respond to online news information (Kearney and Liu 2014).

Previous literature on news information and financial decision-making has failed to address how individuals have constant and time relevant news information available to them on the internet. In the few studies on individual level trading, the studies have been limited to exploring newspapers (see for example, Engelberg and Parsons 2011; Yuan 2015) and consider how information from the previous days affects future trading in the next number of days or even the next month. The only other study (Lillo et al. 2015) to consider individual level trading and online news sentiment examined the relationship at time horizons spanning days and even months.

The internet has had a fundamental effect on how information spreads and diffuses among investors. Investors have access to constantly available and up-to-date information which has changed the information environment and emphasised the speed and immediacy of information (Barber and Odean 2003). As such researchers have called for future studies to explore how information diffuses among the investor population at the inter-day rather than intra-day (or month) level to show the speed at which online information actually diffuses (Peress 2014). Exploring financial decision-making permits new understanding of how behavior and performance of individuals is related to online information where there are considerable rewards to be earnt by individuals who use accurate and up-to-date information.

Paper 3, titled '**The long and short of how online sentiment signals impacts financial trading behavior and performance',** shows how sentiment affects individual trading at much finer time horizons than days. This paper combines two datasets: i) individual level trading records related to the FTSE 100 and whether individuals bought or sold the market in 190,363 trades from 3583 individuals over 6 months from 1st April 2012 to September 30th. ii) Sentiment (positive/negative) scores related to news for the FTSE 100 constituents over the same period from Thomson Reuters News Analytics. Combining these two datasets shows how sentiment in financial news affects individuals prior to and during their trading. The analysis focuses on the influence of sentiment to drive trading behavior and, how sentiment affects performance.

Importantly, positive sentiment has a different and asymmetric effect on behavior compared to its negative counterpart (Akhtar et al. 2013). Positive information is more likely to affect immediate decision-making and impulsive buying behavior, whereas negative information affects more long term decision-making and is likely to increase resistance to buying (Sul et al. 2016). In the stock market, individuals decision-making is limited to stocks they already own (Barber and Odean 2008), therefore, the effects of sentiment are limited.

In spread trading however, individuals can profit from predicting that the market price will go up or down and provides a holistic environment to explore how positive and negative sentiment affects buying and selling decisions respectively. Previous research has failed to consider that individuals will behave differently with positive news if they believe the market price will increase or decrease (Sul et al. 2016) and therefore, spread trading provides a suitable experimental set-up to explore the full effects of sentiment.

The results in paper 3 show that positive and negative sentiment affects buying and selling behavior differently. Sentiment has a significant impact on the buy-sell imbalances in the market and this is the first paper to show how individuals act in a sentiment-contrarian fashion, buying the market around negative sentiment and selling the market around positive sentiment. Finally, trading short around prevailing negative market sentiment can lead to profitable returns whilst prevailing positive sentiment has a negative effect on short trading returns. There is no similar relationship for long trading, however their risk-taking levels are significantly impacted by sentiment, with positive sentiment decreasing risk and negative sentiment increasing risk.

The sentiment-contrarian behavior that paper 3 exposes is important in the wider contexts of behavioral finance. Research has found that individuals behave in a price-contrarian behavior, buying the market when the price decreases and selling when the price increases (Lillo et al. 2015). After controlling for price changes though, these results show that individuals trade in opposition to sentiment, which could contribute to price drifting and the slow adjustment of prices to earnings announcements.

Also, short traders have been found to achieve superior performance in the literature (Boehmer et al. 2008) and it has been suggested that it is their ability to interpret public news information better than long traders. Paper 3 shows exactly how short trading can achieve superior performance by showing how sentiment affects their returns: short traders achieve significantly higher returns around negative news and significantly lower returns around positive news. Research has shown that short traders are skilled information processors and it is after news is that they achieve their advantage (Engelberg et al. 2012). Paper 3 shows that a momentum type strategy associated with short trading can lead to improved returns, which is one reason why short traders may have a reputation for being informed traders.

Paper 3 contributes to the literature threefold: i) how the sentiment from aggregated online news articles is useful for explaining individual trading behavior. In particular, the underlying order flow (total buy stakes minus sell stakes) at the individual level shows that individuals act in a sentiment-contrarian fashion, buying the market in relation to negative sentiment and selling the market in relation to positive sentiment. This finding is robust after controlling for price-change behavior. Currently, the literature provides only one other example of sentiment contrarian behavior at the aggregate market level (Yang et al. 2017) and therefore, this paper on individual sentiment contrarian behavior has important implications in context of prices slowly drifting to news announcements because individuals do in fact trade in opposition to prevailing sentiment. ii) The effect of news sentiment on individuals' behavior and performance at the intra-day level provides much finer-grained levels of analysis than has previously been undertaken. Specifically, how trading volume is affected by sentiment at the hour level and how performance is related to a range of different time periods from 15 minutes. The only other paper to look at individual trading found that sentiment measured each day was not significantly correlated to buyers and sellers behavior (Lillo et al. 2015). As such, research measuring sentiment at the minute level may be better suited to reflect the contemporaneous nature of sentiment and the rate at which online news information diffuses among investors. iii) The prevailing market sentiment prior to an individual's trade is related to the returns and variability of the individual's returns. In particular, for individuals who predict the market price will decrease (i.e. execute short trades after the release of sentiment), increased positive sentiment prior to the trade is associated with decreased returns and negative sentiment prior to the trade is associated with increased returns. For individuals who predict the market price will increase there is no significant correlation between sentiment and average returns. However, the variability of their returns is significantly impacted by sentiment: positive/negative sentiment prior to long trading is associated with decreased/increased variability in returns. These findings on the superior ability of short trading relates to the wide literature on short traders advantage (Engelberg et al. 2012) and shows that a momentum type strategy following the release of prevailing negative sentiment can lead to profitable returns.

In summary, this thesis explores the impact of information derived from IT on decision-making in a financial market context. The first paper shows how information from new IT improves predictions and how the market takes a number of years to fully incorporate information from a new IT source. The second paper contributes to understanding of how distance-based information affects behavior and how subjective beliefs from experts can be elicited in the real world to improve forecasting accuracy. Finally, the third paper shows how sentiment in online news information affects individual-level trading behavior and performance.

Research on how web-based information impacts investors, companies and financial markets is becoming increasingly important in the era of 'Big Data' (Liu and Ye 2016). Taken together, this thesis shows that new information derived from IT available on the web improves the quality of decision-making. Experts and novices use online information to improve their predictions, facilitating information dissemination and expediting the rate at which information diffuses into financial markets. Paper 1 shows that it can take years for new geospatial information to diffuse into the market in the 2000s but paper 3 shows that sentiment information from online news articles begins to diffuse in the same day in 2012.

The rate at which information diffuses is rapidly accelerating and individuals are learning to use information available on the web almost instantaneously. As such, financial markets are adjusting to new information and becoming more efficient over time as individual level trading drives market prices fluctuations. Therefore, this thesis provides new contributions of the value of web information in financial decision-making and empirically shows how individuals could use such information.

Specifically, this thesis highlights that IT has an integral role on how, and what types of information are available to decision-makers. By focusing on how individuals and markets respond to online information, this thesis employs web-based information in financial decision-making to provide a holistic view of how IT is transforming decision-making in general.

Chapter 2 The economic value of information derived from a new web-based technology and its rate of diffusion in a financial market

ABSTRACT

As the rate of information availability increases, the ability to use web-based technology to aid decision-making becomes increasingly important. Previous research has explored how new information available on the internet has been used to improve decision-making. By focusing on Virtual Globe technology, this paper shows how unprecedented information enhances forecasting of racehorse performance and leads to significant and measurable economic benefits. Specifically, elevation data from Virtual Globes (VG) enables improved forecasting decisions and this paper shows the rate at which VG information diffuses through the betting market over an eighteen-year period. The results directly demonstrate how markets adapt over time to new information giving rise to profitable opportunities, albeit these profits are short-lived as the market converges towards the new information.

Keywords: Forecasting, Market Diffusion, Diffusion of Information, Virtual Globes, Market Convergence, Symbolic Decision-making.

The economic value of information from a new web-based technology and its rate of diffusion in a financial market

2.1 Introduction

The information environment surrounding financial markets has rapidly developed in recent years due developments in Information Technology (IT). In particular, the internet has made available a rich and diverse source of information that is constantly available to decision makers and can be used to improve forecasting (Yu et al. 2018). Data sources such as Wikipedia, Twitter, Facebook, Google and YouTube offer ubiquitous and incessantly generated data that has played a central role in decision-making literature in a wide array of domains including business analytics (Wang et al. 2016).

Rapid adoption of web-based technology and the shift from proprietary information systems of the 90s to more open and standardized web technologies offers prospects for novel analysis of the relationship between IT and decision-making (Chae et al. 2014). In addition, as technology becomes more affordable, universally available and simple to acquire, the link between IT and decision-making may become apparent (Chae et al. 2014).

Information availability has enabled superior forecasting in relation to diverse applications: social networking data has been used to forecasts box office statistics (Kim et al. 2015), twitter data has been used to predict elections (Huberty 2015) and product reviews have been used to predict sales of new and existing products (Schneider and Gupta 2016). These examples show that new information online has improved decision-making in a wide range of fields.

The extent to which new information online has improved decision-making is highlighted by a recent review paper on how the scope of research related to Google search has burgeoned (Jun et al. 2017). Research on Google search has increased dramatically and the review of 650+ research papers shows that the focus of research has moved from ‘nowcasting’ and a focus on describing the present, to forecasting (Jun et al. 2017). The shift towards using information for predictive analytics underlines how online web information is frequently used to better understand real world events. Despite the diverse wealth of data that is available online, research on how new information available online improves forecasting has centred on forecasting with Google Search data (Joseph et al. 2011; Yu et al. 2018).

The recent developments in technology has led to a rapid growth in location analytics, spatial analysis and geographic information systems (Pick et al. 2017). New information is available through tools like Google maps which provide unprecedented access to spatial decision information yet research on Google mapping technology is sparse (Jun et al. 2017).

This paper focuses on “virtual worlds” and the extent to which they can create and capture value by providing information that can enhance decision-making capabilities (Drnevich and Croson 2013). The focus on virtual worlds is important since knowledge of how geospatial technology can transform operational decision-making and lead to profitability is limited (Habjan et al. 2014).

Specifically, Virtual Globes (VG) emerged in 2000 and have become one of the most popular and influential web technologies (Pick et al. 2017). VGs provide a digital, three-dimensional representation of the earth and are useful for a wide array of tasks from everyday decision-making (e.g., finding locations) to business logistics and military planning. In principle, the use of three-dimensional visualisations and digital maps can enhance decision-making capability, having the potential for improving accuracy and efficiency in location/distance related problem solving and allowing more complex topographical tasks to be undertaken (Mennecke et al. 2000). Laboratory studies have demonstrated that two- and three-dimensional representations support complex tasks (Shen et al. 2012). However there is a lack of empirical evidence concerning how VGs improve decision-making (Liu et al. 2011).

To show how VGs improve decision-making and their value, analysis must be situated where profitability originates – the market – as value is inextricably determined by the market setting (Schryen 2012). In order to measure the value or profit that can be derived from VG enhanced decision-making, this paper focuses at the market level.

The aim of this study is to identify how VGs improve decision-making and to examine the rate at which new information from VGs diffuses through the market. In particular, elevation data from VGs can be used to enhance decision-making in a financial market.

IT has been transforming the financial sector for decades, yet little research has shown the extent to which IT affects financial markets (Zhang and Zhang 2015). Measuring the speed and extent to which information from VGs diffuses through the market highlights the economic value of employing information from VGs changes. Since new information has been shown to have an unpredictable effect on share prices it is important to avoid using shares prices as a measure of value (Johnstone 2016). This paper explores how IT is linked to market efficiency by examining a setting where the geospatial information provided by VGs offers the prospect of more effective three-dimensional decision-making. It is then possible to measure directly the economic impact of the effective application of this information.

Specifically, results measure objectively how information from VGs can be used to estimate more accurate winning probabilities of racehorses and, how betting strategies based on these probabilities produce economic value.

This paper makes three important contributions: First, the core contribution of this work lies in how geospatial information can be utilized to improve the prediction of performances of racing horses. VG technology can improve decision-making capabilities by providing more accurate and complex spatial information that can aid decision-making. VGs provide information that can be used in a geospatial-based model to obtain superior forecasts compared to odds information. Second, this improvement in prediction scores may be diluted over time as the information becomes public knowledge to the market. The speed with which technology-enhanced information diffuses through a financial market varies and there is a clear link between technology-enhanced decision-making and the profit that can be achieved in a financial market as the market converges towards new information. Finally, the paper estimates the value of using geospatial information in decision-making by conducting betting strategies to highlight potential profits from employing the information. These contributions illuminate aspects of the under-explored link between VGs and decision-making (Drnevich and Croson 2013).

Exploring the value of IT using data-driven approaches that account for how technologies are used in real world decision-making processes (e.g., using case studies) are essential for explaining how technologies improve model performance (Martens and Provost 2014). This paper supports the validity of data-driven approaches for measuring improved forecasting accuracy by highlighting the economic value of such innovations.

The remainder of the paper is organized as follows: section 2 describes VGs, their history and how VGs might aid decision-making. In this section the literature related to the speed of technology diffusion is used to develop hypotheses. In section 3, the data and methodology used to test the hypotheses are described. Section 4 presents and discusses the results. Sections 5 and 6, draw some conclusions and discuss the broader contributions of these findings in terms of information diffusion and value of IT.

2.2 Hypotheses

2.2.1 The nature and history of Virtual Globes

Digital Elevation Models (DEM) provide a 3D digital representation of the terrain's surface showing elevation detail. Elevation data have been freely available over the internet since 1999, when GLOBE was released. This dataset resulted from a combination of various existing DEM products. A series of DEM product releases have followed. The most significant resulted from the Shuttle Radar Topography Mission (STRM) mission, a joint collaborative effort by multinational agencies including NASA. On the 11th January 2003 the resulting dataset, SRTM, was released online and made available in VGs, providing near global coverage to the public (Rabus et al. 2003). Initially, only individuals with the necessary programming tools were able to use the raw data. Amongst these were academics who developed Geographical Information Systems (GIS). GIS scientists used specialist software called ArcGIS to perform geographic analysis, initially for regional planning. Those with the programming knowledge could incorporate the raw data from DEM products in ArcGIS 8.0 through the command line interface, but the technical skills necessary to manipulate the elevation data were considerable.

The significant technical barriers to using elevation data were lifted when more commercial VG products, such as KeyHole (which later became Google Earth), were released in June 2001. These were more user-friendly and intended for public use. GIS software was specialised, expensive to use, had high functional capacity and complexity, and was intended to be used by professionals. By contrast, VGs were easy to use, free, had less analytical functionality and were widely used by the public (Goodchild et al. 2012).

SRTM data, released in 2003, enabled the public to use accurate geospatial data more effectively. This elevation data was made available through various VG products: (i) to KeyHole users in 2003 and subsequently to Google Earth users from its first release in June 2005; (ii) ArcGis incorporated the data in February 2003 and (iii) Nasa World Wind, first released in 2003.[4] The advent of VGs and particularly KeyHole, which first made elevation data freely and easily accessible to the public, have facilitated the sharing of geospatial information (Sheppard and Cizek 2009), which has the potential for improving the public's geospatial decision-making capabilities.

Following widespread media coverage and the open availability of elevation data in easy to use tools, the number of geospatial users has risen dramatically. GIS software, the most popular software among GIS specialists when it emerged in the 70s, took 30 years to achieve 1 million users (Flaxman and Vargas-Moreno 2012). More user-friendly VGs have attracted greater user numbers in a far shorter time period: Keyhole, released in 2001, had 250,000 consumers in 2003 (Keyhole 2003). Google Earth, released in 2005, had 100 million users one year after it was launched (Sheppard and Cizek 2009); dwarfing (by 100 times) the user numbers of the most popular social media technologies, Facebook and Twitter in the launch year (Shontell 2012). The context and chronology of important events related to VGs are summarised in Table 1.

---

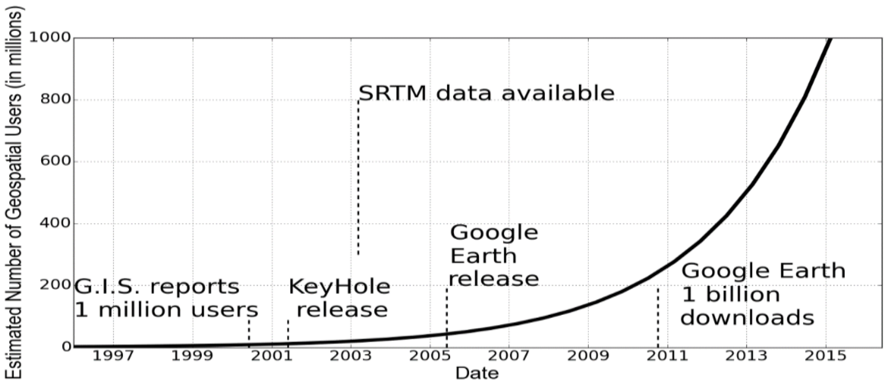
[4] Further information, historical context and media coverage related to VGs is provided in Appendix A in the supplementary file.

Table 2.1 Context and Chronology of Important Events related to VGs Software

Date	Event
1969	Tomlinson (1968) coined the term “Geographic Information Systems in his paper “A Geographic Information System for Regional Planning”
1970	Burgeoning of GIS software
1997 November	Microsoft released virtual globe Encarta Virtual Globe 98 (offline CD version); one of the first virtual globes allowing users to navigate and experience the world in 3-D
1999 December	ArcGIS 8.0 released: Professional software providing a command line interface for (experienced) users to create, combine and analyse statistical mapping information
2000	1 Million estimated Geospatial users
2001 June	Keyhole Inc. released Earth Viewer with elevation data
2003 February	ArcGIS incorporated elevation data in software update
2003	Widespread media coverage of VGs, including articles in Wall Street Journal, NY Times, PC World and USA TODAY.
2003-4	NASA World Wind released with elevation data and made available through open source.
2004	Google acquire Keyhole Inc
2005 June	Google Earth released
2007	100 Million estimated users of Google Earth
2011	Google Earth reaches 1 billion downloads
2016	1 Billion estimated users of Google Earth

The rise of VGs and their improved user friendliness had a profound effect on decision-making in a variety of contexts, from everyday decision-making based on location searches (Constantiou et al. 2014), through military planning associated with flight routes (Meeks and Dasgupta 2004), to planning time-critical routes for emergency services (Shen et al. 2012). The pervasive nature and popularity of VGs (see Figure 1), make them a significant technology. Measuring their economic impact and rate of effective adoption is, therefore, an important goal.

Figure 2.1 Estimated growth in Geospatial users



2.2.2 Decision-making advantages of Virtual Globes

In order to recognize how information from technology can lead to economic benefits it is necessary to understand how the IT improves capabilities, i.e. how human and technological resources work together (Bulchand-Gidumal and Melián-González 2011). By adopting a resource-based view, research has sought to identify the unique capabilities of IT that can result in superior performance.

It has been suggested that geospatial technology can provide more factual based information for decision-making and that this can improve firm performance (Habjan et al. 2014). Since it is difficult to discern this impact in an organisational context, how VG's improve capabilities of individuals operating within a financial market enables research to directly assess the economic value which these enhanced capabilities provide.

It has been suggested that 'tasks' can be broadly described as 'spatial' or 'symbolic'

(Vessey 1991). These are, characterized as visual (graph or map) and tabular representations (tables), respectively. Spatial tasks are those where users must acquire information and make a simple decision that is generally qualitative in nature, whereas symbolic tasks involve discrete data values and are generally quantitative. Spatial and symbolic representations are suited to varying problem tasks. The former involve decision makers viewing information at a superficial level and are largely perceptual while the latter provide access to the underlying data, enabling analytical processes (Vessey 1991; Vessey and Galletta 1991). For example, if data are displayed on a map with geographic areas in different colours to indicate their data values, then this is a spatial representation, while data displayed with specific numeric values for each geographic area are symbolic (Dennis and Carte 1998). The former presentation allows individuals to identify which of two locations is more elevated ('spatial task'), while the latter enables one to quantify differences in height ('symbolic task'). Consequently, three-dimensional digital map-based presentations allow individuals to address symbolic tasks using analytical processes (Speier 2006).

Spatial and symbolic tasks largely align with the two systems of thinking described by Kahneman (2003): System 1, which is associated with intuitive, perceptual responses to problems that require very little conscious thought, and System 2, which is associated with reasoning and is undertaken deliberately, requiring effort. Most everyday decisions are made intuitively (Kahneman 2003) and as such require little conscious thought. For example, Location Based Services (LBS) have become ingrained in individuals' decision systems and many smart phone owners use LBS services to locate the nearest restaurant (Constantiou et al. 2014). LBS have altered the decision-making processes by which individuals use VGs to solve problems related to the physical world: a subtle change, whereby direct reference to the physical world has been superseded by reference to a digital representation. This has had an effect on the cognitive systems of thinking. In fact, for many individuals, LBS have become the first reference for location-based problems and are largely framed by system 1 modes of thinking (Constantiou et al. 2014).

This paper argues that the value of VGs for decision-making will not be fully realized until the geospatial data is used in symbolic tasks. Using the raw elevation level data allows decision-makers to solve complex decision-making problems, where reasoning and specific data values are required to make accurate calculations. Web-based maps are useful for 'fast and frugal decision-making' that do not require analysing relationships among geographic areas (Jarupathirun and Zahedi 2007). However, elevation data provides the means of analysing symbolic tasks. VGs can affect the performance of both spatial and symbolic tasks. Previous studies examining how geospatial technology is used to support decision-making have focused on spatial tasks, without empirical evidence (Liu et al. 2011). It has been proposed that information derived from geospatial technology can improve decision-making (Habjan et al. 2014), but no evidence has been provided to support this view. Therefore, the following hypothesis is tested:

Hypothesis 1 (H1): VG information improves upon the predictive performance of predictions generated from market information

2.2.3 Speed of market diffusion

Technological change occurs in three stages: invention, innovation and diffusion (Stoneman 1995). The diffusion stage, defined as the "process by which individuals and firms in a society/economy adopt a new technology", is arguably, the most important (Hall 2004, p. 2). Identifying impact at the diffusion stage involves measuring "how the economy changes as new technologies are introduced and used" (Stoneman 1995, p. 2). The diffusion process is thus an intrinsic part of technology adoption, focusing on the actual impact of IT on a population. However, this is a topic that has been somewhat neglected (Hall 2004) and this paper focuses on the diffusion rate of the VG information that can improve symbolic decision-making tasks.

While models of technology adoption help us understand user perceptions and attitudes towards technology adoption, they offer little to explain the rate and extent of diffusion. In order to obtain the benefits from a technology, the end user must employ it appropriately. Understanding the degree to which technology is used appropriately, is a critical gap in theories of adoption (Aggarwal et al. 2015; Venkatesh et al. 2012). Furthermore, the technology adoption theories are deterministic and do not account for environments where there are competing technologies and non-technological means of acquiring similar information (Constantiou et al. 2014). Technology adoption is not deterministic in the real world as user adoption is not guaranteed; the rate of diffusion will depend on a multitude of factors, including, for example, competing technologies and user friendliness. The limitations of many of the adoption theories and the adoption literature is their neglect of the extent to which, and the manner in which, an adopter actually uses the technology (Comin et al. 2006). Exploring how elevation information from VGs and how this diffuses through the market, will contribute to knowledge of how technology is actually used.

Typical models already exist for exploring diffusion patterns, such as Bass Diffusion Model (Bass 1969). Such models are useful for showing how products or information are adopted by the underlying population over time, and the characteristics of early adopters and innovators compared to the late majority and laggards who are the last to use a new technology. Without information on the individual betting behavior and their reasons for betting on a specific horse, it is unclear whether bettors used information from VGs.

The purpose of this paper is to show the convergence aspect implied by the nature of diffusion as communications between participants is a process by which participants exchange information and converge towards a general understanding of a certain event (Rogers 1995). Conceptualizing diffusion in a framework based on information and uncertainty helps represent how information represents one of the main means to reduce uncertainty regarding a specific event and information is exchanged between participants allowing them to gain information and remove uncertainty towards the eventual outcome of an event (Rogers 1995). Convergence then is the process for information to be exchanged between participants, which leads to a better understanding of eventualities. Market convergence refers to the time frame between which information is released and the market reacts to that information, becoming efficient by fully reflecting the new information in prices (Nassirtoussi et al. 2014). Therefore, such models as the Bass Diffusion model are unable to help show how new available information becomes incorporated into the market.

It is necessary to track how VG information itself diffuses through a financial market since 'diffusion does not take place in a vacuum, but within market systems' (Papagiannidis et al. 2015, p.2). Market forces undoubtedly affect the rate of adoption and it has been suggested that it is necessary to study how technology is used and how technology impacts the market and vice versa (Benbasat and Zmud 2003). Directly focusing on how information from VGs is employed in a market context will show the tangible link between information and the economy.

Diffusion in this context is the complex process by which VG information is eventually fully incorporated by a population. Adoption is the first step in the process which ultimately leads to the full value of a technology being realized by the market (Fuentelsaz et al. 2003). Consequently, it is necessary to explore diffusion over time to fully capture the process that starts with the first individual using VG information, to the final stage where the VG information if fully accounted for by the market (Fuentelsaz et al. 2012).

It has been shown that individuals learn to use new technologies in ways that increase their long term welfare (Stillwell and Tunney 2012). In addition, the Efficient Market Hypothesis (EMH; Fama (1970)) suggests that in efficient markets all available information is discounted in prices instantaneously. Consequently, horserace bettors could use the information that can be discerned from VGs regarding racetrack topology to maximise their returns. As a result, this information could quickly be fully discounted in market odds. Elevation data was available to experts from 1999 in its raw format but SRTM data was first available in VGs to the public in 2003 and this enabled them to employ it for symbolic decision-making. This time difference helps set up the following hypothesis to systematically measure and quantify the time lag for complex forms of information to be disseminated, understood and assimilated by market participants:

Hypothesis 2 (H2): technology enhanced information that facilitates symbolic decision-making, diffuses through a financial market completely and immediately once available to the public

2.3 Design of empirical study

The analysis explores how the aggregate market responds to VG information and how the predictions based on this new information generate superior predictions, which decline over time. By focusing on the information derived from VGs and using an information-centric approach, there are no overhead costs to consider and the value of appropriately applying the information derived from the IT can be directly linked to economic impact.

The horserace betting markets was identified as an environment, which provided these advantages. In particular, bettors in these markets make symbolic decisions concerning each horse's chance of winning (which inform their betting decisions) and these are more likely to be accurate if they account for the extent to which various gradients and cambers at different tracks afford advantages to certain horses. These decisions are likely to be enhanced by data concerning topology available from VGs. Furthermore, the potential advantage of applying the VG data appropriately is measured by contrasting the betting profit achievable from decisions (a) informed by judgements based on the symbolic representation of these gradients and cambers from VG data with (b) based on simply observing each racetrack to estimate the spatial topography perceptually.

Bettors must assess the probability of each horse's chance of winning, in order to decide which horse to bet. Consequently, the input and output are clearly defined. This provides a less ambiguous setting than previous studies that attempt to assess the extent to which IT is exploited. Furthermore, each race is a separate decision-making task and there are several thousand races each year. Consequently, races provide an historical record of the changing degree to which individuals incorporate VG information into their betting decisions. The process for learning the value of VG-based information is facilitated by immediate feedback (i.e. the winner of each race is announced immediately after the event). This helps to overcome one of the limitations of previous studies, namely, the time lag between adopting new IT and when it is possible to observe the economic benefits (Devaraj and Kohli 2003).

Prices within betting markets represent the betting public's combined view of the winning probabilities of each horse. By comparing these probabilities with results of a number of races, the calibration of bettors' probabilities to the VG information can be assessed. Equally, betting strategies using this VG Information are used to estimate winning probabilities. In this way, the extent to which geospatial data from VGs provides superior estimates and the economic value (in terms of betting profit) of effectively employing this symbolic information is highlighted.

In general, markets have been shown to effectively aggregate information (Xu and Zhang, 2013) and the EMH suggests that in efficient markets all available information is discounted in prices instantaneously and no abnormal profits are possible. In betting markets, rewards are instantaneous and relate directly to the outcome of a given race. Consequently, this overcomes another limitation of some previous studies which have tried to examine the economic value of IT; namely that a redistribution or dissipation of profits arising from the IT occurs and it is difficult to directly allocate the benefits that the innovation offers a particular decision-making task.

Betting markets also have the advantage that they capture the 'real world' decision-making of individuals who have strong (financial) incentives to make accurate judgements. This is an important advantage over the majority of research on technology diffusions conducted in laboratory settings, where participants are offered either no incentives or artificial incentives of limited value.

2.3.1 Data

Analysis of 18 years of data (1997 to 2014, inclusive) from the UK horse racing markets including 75,750 races (incorporating 76,406 different horses) run at all 34 UK racetracks is used to test the hypotheses. This period encompasses the dates of release of VGs. Geocontext (www.geocontext.com), a web service available from 2010, provides access to SRTM elevation data from the Google elevation API. This data provides topological profiles of the 34 UK racetracks. Races of different lengths at a given racetrack are often run over different sections of the track – defined as 'a course'. Elevation data above sea level data was collected at 'measuring points (MPs)' at 50m intervals from the finishing post to the start of each course. At these points, the camber was also measured by taking elevation readings across the track. The starting line and finishing posts were determined via Geocontext in conjunction with a published source that explicitly shows the start and finishing posts for each course.[5] This procedure produced a dataset of topologies related to 300 different courses.

Spence et al. (2012) found that horse races are largely decided by performance in the final section of the race. A leading racehorse trainer and breeder who prefers to remain anonymous was consulted. They confirmed that the topology in the final section of the race was crucial. Consequently, topology variables were created relating to the final quarter of the race distance and in the last furlong (200m) of the race. To identify suitable variables that capture the advantage a given horse gains from topological features, the equine literature provided insight into the kinematics of horses. For example, the gradient of a track has a significant effect on horse speed. Self et al. (2012, p. 606) found that "during racing, horse maximum speed is less on both inclines and declines, with top speeds being achieved during level running". Consequently, the profile of the track and the gradient of the slope (along the track) can affect the speed of horses. The physiological characteristics of some horses might give them an advantage on flat, upward or downward sloping tracks or on undulating tracks.

In brief, four variables capture four different topological features of courses that might confer an advantage/disadvantage on particular horses:

(i) the camber of track k (i.e. degree of slope towards the centre of the track) at various points, because speed is influenced by adaptation to curved motion (Hobbs et al. 2011). The number of MPs with 'flat' (cf. 'steep') cambers in the last quarter of the race (*CAMBERS*); where 'steep' is defined as a gradient greater or less than 10 degrees (based on Hobbs et al.'s (2011) definition). The number of MPs in the last quarter of the race where the cambers were

[5] http://www.racingpost.com/horses/course_list.sd

‘flat’ were measured. The proportion of MPs with flat cambers in the last quarter as compared to the the total number of MPs in the last quarter:

$$CAMBERS_k = \frac{\sum_{f=1}^{\left(\frac{n_f}{4}\right)} \begin{cases} 1 \text{ if } \delta_f < 10° \text{ and } \delta_f > -10° \\ 0 \text{ else.} \end{cases}}{\frac{n_f}{4}} \tag{2.1}$$

where $1,2 \ldots n_f$ is the total number of MPs for track k, $\frac{n_f}{4}$ is the number of MPs in the last quarter, n_1 is closest MP to the finish line, n_f is furthest MP from it and δ_f is the angle of incline (decline) along $camber_f$. The Proportion of flat MPs was measure in relation to the total number of MPs in the last quarter of each track ($\frac{n_f}{4}$) to nullify effects of distance within the variables, ensuring that long and short tracks are treated equally.

(ii) The cumulative drop in elevation in the last quarter of the race (*DOWNSLOPE*); since downward sloping tracks reduce a horse’s maximum speed differentially, depending on the horse’s physique (Self et al. 2012). Therefore, the cumulative decline in the last quarter of the race for track *k*:

$$DOWNSLOPE_k = \frac{\sum_{f=1}^{\left(\frac{n_f}{4}\right)} (Df - Df_{+1})}{\frac{n_f}{4}} \tag{2.2}$$

where $1,2 \ldots n_f$ are defined as above and Df is the elevation at the inside rail at MP_f, and Df_{+1} is the elevation at MP_{f+1} .

(iii) The undulation of track *k* in the last furlong (*UNDULATION*); as this has been shown to differentially affect a horse’s galloping speed (Self et al. 2012). This is calculated as the standard deviation (SD) of the elevations along the inside of the track at the various MPs, in the last furlong. Since the last furlong is 201 meters, there are 4 readings (50 meters apart) and the SD for track *k*:

$$UNDULATION_k = \sqrt{\frac{1}{4}\sum_{f=1}^{4}(Df - \mu)^2} \tag{2.3}$$

where Df is the height above sea level at MP_f and the mean height is given by $\mu = \frac{1}{4}\sum_{f=1}^{4} Df$.

(iv) Average width of the track in the last furlong (*WIDTH*), as this can affect horses differently depending on their various running styles (Spence et al. 2012). Even though the rails can be moved during meeting to give fresh ground, the width of the track is limited to the available track. This is calculated by determining the width W_f at each MP_f in the final furlong of the track. Since there are four MPs in the final furlong, the average width for track k is:

$$Width_k = \frac{\sum_{f=1}^{4} W_f}{4} \tag{2.4}$$

A number of other track features relating to the topology of the track were tested: whether the curve of the track and the change in height from start to finish, interactions with weather, ground conditions and distance of the race affected horse performance. However, none of these features significantly affected winning probabilities. The final set of four variables provided reliable and consistent variables related to racetrack topology that might affect a horse's winning probability. Further details of these variables are provided in the supplementary Appendix B.

There are a vast range of potentially influencing variables related to the track, the horse, the jockey and even weather and ground related conditions as well as market-making and demand side conditions that are known to impact on the pricing in horse racing betting market (Bruce et al. 2009). Due to the inherent difficulty in controlling for every known factor that can influence the race, the analysis uses an information-centric approach of tracing VG information and market convergence in relation to this specific form of information. The nature and complexity of the horse race betting markets make it difficult to isolate the effect of anyone factor. Therefore, given the turbulence, diversity and complexity of the broader environment, the methodology identifies a number of VG-related factors and the analysis focuses on to what extent this information is discounted in odds.

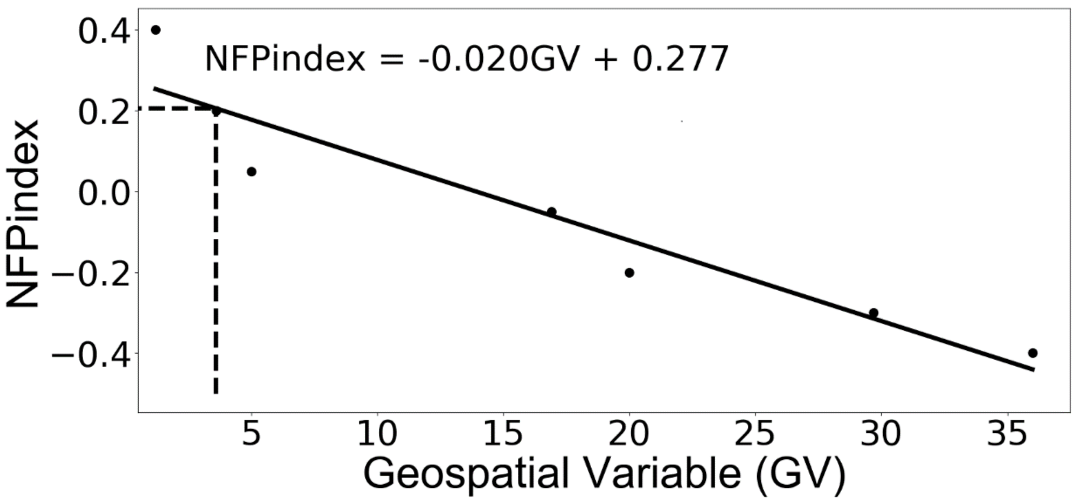
2.3.2 Methodology

In order to test whether the four variables associated with racetrack topography (*CAMBERS, DOWNSLOPE, UNDULATION and WIDTH*) confer advantages/disadvantages to particular horses in relation toeach horse's past performances, measured by its normalized finish position (Brecher 1980); defined for horse i in race j as:

$$NFP_{ij} = 0.5 - \frac{Ordinal\ Finishing\ Position_{jk} - 1}{Number\ of\ runners_k - 1} \tag{2.5}$$

where the first and last placed horses are assigned values of 0.5 and -0.5, respectively. The NFP_{ij} provides a means of assessing how well the horse performed compared to the other horses in the race. For example, if a horse had performed better in races where the cambers were generally flat, then this might indicate the horse has an advantage in those conditions. In order to find such patterns, the preference variable technique introduced by Benter (1994) was used. Specifically, for each horse, a linear regression is constructed with one of the geospatial variables (e.g. *CAMBERS*) as the independent variable and horse i's normalized finish position index, $NFPindex_{ik}$ (NFP_{ik} – average NFP_{ij} in previous races) as a dependent variable. $NFPindex_{ik}$ measures the degree to which the horse over/under-performs in race k compared to its average performance in past races (i.e races 1… k-1). This technique captures how much worse/better than normal the horse will perform, ceteris paribus, when encountering a track with particular values of a geospatial variable (e.g. a track with completely flat *CAMBERS)*; consequently, it captures its 'preference' for particular geospatial conditions. The application of the technique is depicted in Figure 2. Here, the horse's previous performances over the history of races up to today's race (*j*) are shown as dots. The estimated linear regression line shows the horse's preference for lower values of the geospatial variable. Consequently, the dotted line in figure 2 indicates that if today's race *j*, were run at a track with a geospatial measure of 3.58, then for horse *i* the predicted $NFPindex_{ij}$ is 0.20 (i.e. the horse is predicted to perform better than its average performance under those geospatial conditions). The regression is re-estimated prior to each race of each horse, accounting for all its runs up to that time. Since it is impractical to fit a model with less than 4 points, the preference indicator is set to zero for the first four races for each horse.

Figure 2.2 Preference Variable Technique



2.3.3 Two Stage conditional logit model

The preference variables were used to predict the finishing position for horse *i* in the next race, as follows:

$$predictedNFP_{ij} = \beta_1 prefCAMBER_{ij} + \beta_2 prefDOWNSLOPE_{ij} + \beta_3\, prefUNDULATION_{ij} + \beta_4 prefWIDTH_{ij} + \beta_5 averageNFP_i \qquad (2.6)$$

3 years of data (1997-1999) is used to estimate the coefficients for *predictedNFP*$_{ij}$ using ordinary least squares regression. The horse's *predictedNFP* for the races in the one year hold-out sample is estimated using the estimated coefficients based on the training data to predict how the horse will perform in the out-of-sample races.

The conditional logit model (CL, McFadden, 1974) is used to estimate winning probabilities for each horse using the same three year training set (1997-1999). CL is the most widely used model for competitive event prediction (Lessmann, et al. 2012) and is applicable because it takes into account competition between horses (unlike ordinary logistic regression). The output of CL is a vector of estimated winning probabilities p_{ij} for each horse in race *j*:

$$P_{ij}^{e} = p_{1j}^{e}, p_{2j}^{e}, \dots p_{n_j j}^{e} \tag{2.7}$$

where n_j is the number of horses in race j. CL estimates P_{ij}^{e} based on the effects of the independent variables employed in the model. In particular, a 'winningness index' W_{ij}, is defined for horse *i* in race *j*, as follows:

$$W_{ij} = \sum_{k=1}^{m} \beta(k) x_{ij}(k) + \varepsilon_{ij} \tag{2.8}$$

where *β(k)* (for *k= 1,2…m*) are the coefficients which measure the relative importance of the input variables $x_{ij}(k)$. W_{ij} provides a measure of the relative strength of each runner in a race. The error term ε_{ij} represents the information that is unknown in the model.

Assuming that the error term follows the double exponential distribution (which has been shown to be a sensible assumption for horseraces: Benter, 1994), the probability of horse *i* winning race *j* is given by:

$$p_{ij} = \frac{\exp(\sum_{k=1}^{m} \beta_k x_{ij}(k))}{\sum_{i=1}^{n_j} \exp(\sum_{k=1}^{m} \beta_k x_{ij}(k))} \tag{2.9}$$

I use the *predictedNFP*$_{ij}$ as the sole input for a CL model in order to estimate the winning probability of each horse in each race, based on data for the previous 3 years (1997-1999). This model is referred to as $Geospatial_{simple}$:

$$Geospatial_{simple} = \frac{\exp(\sum_{k=1}^{m} \beta_k predictedNFP_{ij}(k))}{\sum_{i=1}^{n_j} \exp(\sum_{k=1}^{m} \beta_k predictedNFP_{ij}(k))} \tag{2.10}$$

The predicted probabilities from the $Geospatial_{simple}$ model are then used as independent variables, alongside market odds implied probabilities, in a second CL model (second stage), as follows:

$$Geospatial_{Full} = \frac{\exp(\alpha \ln\left(p_{ij}^{s}\right) + \gamma \ln(p_{ij}^{f}))}{\sum_{i=1}^{n_j} \exp\left(\alpha \ ln\ (p_{ij}^{s}\right) + \gamma \ln(p_{ij}^{f}))} \tag{2.11}$$

where, $\ln(p_{ij}^{s})$ is the natural logarithm of the probability estimates derived from $Geospatial_{simple}$ and $\ln\left(p_{ij}^{f}\right)$ is the natural logarithm of the market odds implied probability, for horse *i* in race j. The parameters α and γ are estimated using maximum likelihood procedures using the three year sample period.

The market odds implied probability for horse *i* is given by 1/(decimal odds for horse *i*). For example, market odds of 2.0 mean that a winning bet of £1 produces a profit of £1, and this suggests that the market participants as a whole believe the horse has a 50% chance of winning the race. The natural log of these probabilities is used in the CL model, as this transformation provides a better fit to winning probabilities (Benter 1994).

The two-stage procedure adopted here has been shown to provide more accurate winning probabilities than those derived by simply incorporating odds-implied probabilities and other variables (geospatial, in this context) in a one-stage CL model (Benter, 1994). The advantages of this method arise from the fact that the two-stage procedure captures more information contained in the independent variables (Sung and Johnson 2007), allowing more information from the complex geospatial variables to be included in the model.

To explore how the information arising from the geospatial data diffuses through the betting market over time, a sliding test sample window is used. In particular, a model is fit onto a three-year sample of data to predict winning probabilities for races run over the next year. For example, $Geospatial_{simple}$ and $Geospatial_{Full}$ are estimated using data from a 'training window' consisting of the first race in January 1997 to the last race in December 1999. These models are used to predict the winning probabilities in 'out-of-sample races' run in January to December (inclusive) 2000. The training window then moves forward to include races run from January 1998 to the last race in December 2000. The model developed is used to predict winning probabilities in 2001, and so on.

2.3.4 Testing hypotheses

To test H1that VG information improves upon the predictive performance of predictions generated from market information, statistical models derived from the geospatial variables are used to accurately predict horses' winning probabilities. First, Examining whether the model with geospatial preference variables ($Geospatial_{simple}$) can improve probabilities estimates over those based on random choice shows the validity of the constructed variables in estimating winning probabilities. Testing whether the variables add predictive power over a random choice model by examining if the variable coefficients β are significantly different to 0 (using the standard normal test statistic $z(l) = \frac{\beta(l)}{S.E\,[\beta(l)]}$).

McFadden's pseudo-R^2 value, a goodness-of-fit index, shows how much variation in the win probabilities are explained by the model (Bolton and Chapman 1986):

$$R^2 = 1 - \frac{LL(model)}{LL(random)} \tag{2.12}$$

$LL(model)$ is the log-likelihood of the $Geospatial_{simple}$ and $LL(random)$ is log-likelihood of the random choice model, where each horse is assigned an equal winning probability. The log-likelihood of $ln(model)$ is given by:

$$LL\,(model) = \sum_{j=1}^{N} \sum_{i=1}^{n_j} y_{ij} \ln p_{ij} \tag{2.13}$$

and $LL(random)$ is given by

$$LL(random) = \sum_{j=1}^{N} \ln(\frac{1}{n_j}) \tag{2.14}$$

where y_{ij} = 1 if horse i won race j and otherwise 0 and N is the total number of races in the dataset. The maximum likelihood procedure estimates the 'best-fitting' parameters of a statistical model, namely, those that maximize the probability of observing the actual data.

The $z(l)$ statistics for the $Geospatial_{simple}$, provide the means of testing whether the information provided by the geospatial variables is more useful than a random model. The R^2 score is a measure of how much more variance is explained in winning probabilities derived from the $Geospatial_{simple}$ compared with those derived from the random model.

As shown in the results section, the geospatial variables can help to predict winning probabilities. To what extent information contained in these variables is discounted at the aggregate market-level is evidenced by the improved estimates of winning probabilities over and above those estimated by the public. The odds of each runner provide the best estimate of the public's combined view of each horse's prospects of success. Whether winning probabilities derived from a CL model including the geospatial variables and odds as independent variables ($Geospatial_{Full}$) are significantly more accurate than winning probabilities derived from a CL model simply including odds (*Odds*) highlight whether the geospatial information is discounted. Log Likelihood Ratio (LLR) test compares how well two models fit the data. In particular, the maximum likelihood values (LL) of the $Geospatial_{Full}$ and the $Odds$ (null model) models show whether the model incorporating geospatial information significantly fits the race results data better than the model incorporating only odds, using the following statistic:

$$LLR = -2\frac{LL(\,Geospatial_{Full})}{LL(Odds)} \tag{2.15}$$

This statistic is $\chi 2$ distributed with degrees of freedom (d.f.) equal to the difference in the number of parameters between the models (d.f. = 1). The associated p-value tests whether null hypothesis is rejected and that the $Geospatial_{Full}$ model predicts winning probabilities more accurately.

This process is repeated via a series of sliding windows of consecutive three year periods, for all years from 1997 to 2014. The sliding window approach is similar to that adopted in recent studies (e.g. Charles et al. 2012) and allows us to explore the degree to which market prices discount symbolic information over time.

To test H2, that technology enhanced information that facilitates symbolic decision-making, diffuses through a financial market completely and immediately once available to the public, the winning probabilities estimated using $Geospatial_{Full}$ are compared to a model containing only the odds information. In 2003 VGs made elevation data available to the public. Although elevation data have been freely available over the internet since 1999, it was only when SRTM data was made available in VGs in 2003 that the public could harness this data to make probability estimates. If the information diffused immediately then the LLR test for 2003 will be insignificant.

The public will take time to understand fully the value of the data. This would be confirmed if the LLR test statistic were significant for periods immediately after the release of the data and later, when the information was fully exploited in the public's decisions, the LLR test statistic became insignificant.

To show the economic value of VG enabled symbolic decision-making capabilities betting simulations are conducted. In particular, to discover if a betting strategy based on the winning probabilities derived from a CL incorporating geospatial information and market odds ($Geospatial_{Full}$), can provide profitable returns. To develop an appropriate betting strategy, the probability estimates from $Geospatial_{Full}$ are used to calculate the expected return for a £1 bet on each horse in each race :

$$E(R_{ij}) = p_{ij} * G_{ij} - 1 \quad (2.16)$$

where p_{ij} is the estimated probability of horse i winning race j derived from $Geospatial_{Full}$ and G_{ij} is the return to a £1 bet placed on the horse (based on its market odds). The Kelly betting strategy (Kelly, 1956) has been proven to be the optimal betting strategy where one's predictions are more accurate than those indicated by the market (Maclean et al. 2010; Sung and Johnson 2010). Assuming an initial wealth level of £1000, the following is simulated:

1. Kelly Criterion : determines how much to bet, x_i, over all n horses in a race in order to maximize the log of expected wealth. Let r_{ij} be the return on a bet of £1 if horse i wins race j and let b_{ij} be the fraction of current wealth that is bet on horse i. Then, given that horse h wins race j, the current wealth increases by the following proportion:

$$1 - \sum_{i=1}^{m_j} b_{ij} + b_{hj} . r_{hj}. \quad (2.17)$$

Specifically, the Kelly strategy determines how much to bet to maximize the expected log payoff across all potential winners h:

$$\max b_{hj} \sum_{h=1}^{m_j} p_{hj} . \ln(1 - \sum_{i=1}^{m_j} b_{ij} + b_{hj} . r_{hj}) \quad (2.18)$$

Research has suggested that using a bet limit provides optimal returns in the long run (Maclean et al. 2010). Consequently a maximum limit per bet of 1% of current wealth is imposed.

If the profit derived from the Kelly strategy using probabilities from $Geospatial_{Full}$ is positive (and greater than the profits derived from a betting strategy based on probabilities derived from a CL model incorporating only market odds implied probabilities) this will show that the improved decision-making capabilities enabled by the VG data provide significant measurable value.

2.4 Model results

2.4.1 Extent of market diffusion

Results of estimating a CL model including *predictedNFP* as the only independent variable ($Geospatial_{simple}$) for each of the 3 year samples from 1999-2014 (inclusive) are displayed in Table 2. The z statistic for *predictedNFP* in each year is significantly different from zero, indicating that *predictedNFP* is useful for estimating winning probabilities. The β for *predictedNFP* is positive for each year (range: 6.0051 to 6.4308; mean: 6.2264), indicating that greater values of predicted NFP are associated with a higher probability of winning. The mean R^2 of the annual models is 0.0230, suggesting that the models, on average, explain 2.3% more of the variation in win probabilities than the null model. Importantly, in each year, the $Geospatial_{simple}$ model is significant at the 1% level, leading us to reject the null hypothesis that a random model fits the data more accurately than one that incorporates predictions of win probabilities based on horses' preferences for various aspects of racetrack topology (determined using VG data). In addition, estimated CL models for each three-year sample incorporating the single independent variable, *averageNFP* were tested. None of these models were significant at the 1% level, further confirming that it is the horses' *preferences* for features of racetrack topography, rather than simply their previous success rate, which are the key predictive ingredients of the variable *predictedNFP*.

Table 2.2 Results for $Geospatial_{simple}$ for 1999-2014

Date	Variable	β	S.E	Z	Sig	R^2	Prob
1997-1999	*predictedNFP*	6.4228	0.2566	25.0282	0.0000	0.0227	**0.000****
1998-2000	*predictedNFP*	6.2821	0.2476	25.3682	0.0000	0.0224	**0.000****
1999-2001	*predictedNFP*	6.1065	0.2355	25.9302	0.0000	0.0225	**0.000****
2000-2002	*predictedNFP*	6.0921	0.2207	27.6067	0.0000	0.0242	**0.000****
2001-2003	*predictedNFP*	6.0051	0.2089	28.7433	0.0000	0.0237	**0.000****
2002-2004	*predictedNFP*	6.2364	0.2030	30.7205	0.0000	0.0248	**0.000****
2003-2005	*predictedNFP*	6.2596	0.1951	32.0840	0.0000	0.0241	**0.000****
2004-2006	*predictedNFP*	6.1964	0.1813	34.1697	0.0000	0.0241	**0.000****
2005-2007	*predictedNFP*	6.2582	0.1680	37.2572	0.0000	0.0253	**0.000****
2006-2008	*predictedNFP*	6.3602	0.1656	38.4031	0.0000	0.0251	**0.000****
2007-2009	*predictedNFP*	6.4308	0.1726	37.2537	0.0000	0.0233	**0.000****
2008-2010	*predictedNFP*	6.3253	0.1717	36.8358	0.0000	0.0223	**0.000****
2009-2011	*predictedNFP*	6.1961	0.1708	36.2791	0.0000	0.0213	**0.000****
2010-2012	*predictedNFP*	6.0439	0.1694	35.6697	0.0000	0.0203	**0.000****
2011-2013	*predictedNFP*	6.1553	0.1722	35.7374	0.0000	0.0206	**0.000****
2012-2014	*predictedNFP*	6.2521	0.1760	35.5194	0.0000	0.0208	**0.000****

** denote significance at 1% level in a 2-tailed test.

2.4.2 Rate of market diffusion

Having confirmed that VGs provide the opportunity to make more accurate winning probability estimates, it is necessary to explore how quickly the market discounts this information. Consequently, the predictive accuracy of a model ($Geospatial_{Full}$) incorporating market odds-implied probabilities and probabilities derived from the $Geospatial_{simple}$ model are compared with that of a CL model simply incorporating market odds-implied probabilities ($Odds$), using an LLR test. This test is repeated for each of the three-year training sets from 1999-2014 and the results are reported in Table 3. In each of the training sets covering years 1999-2005 (inclusive), the LLR test statistic is significant at the 5% level, suggesting that $Geospatial_{Full}$ better predicts the winning probabilities than the ($Odds$). These results provide clear support for H1, that VG information improves upon the predictive performance of predictions generated from market information.

However, from 2006-2014, the LLR test statistic is insignificant at the 5% level. These results confirm that prior to 2006, the market odds do not fully reflect the information regarding horses' topographical preferences (that could be determined from data available in VGs). However, VG information was fully discounted in odds thereafter.

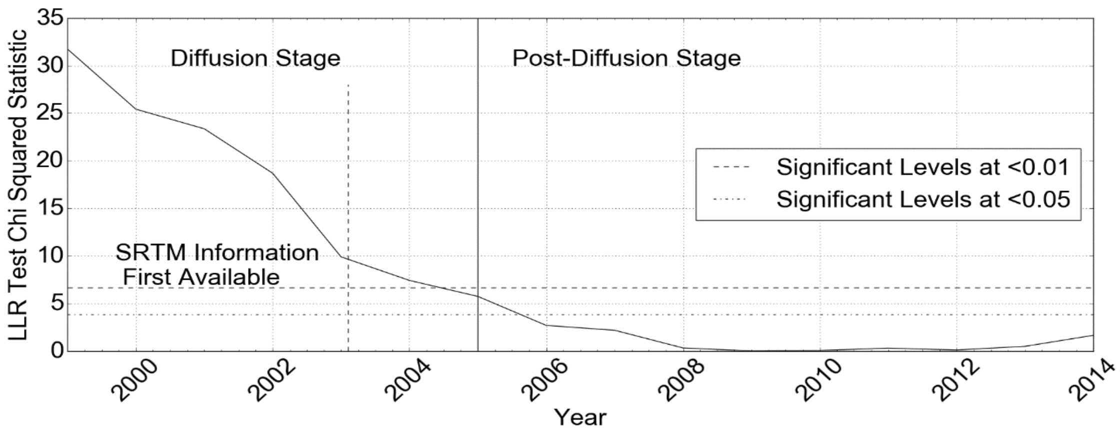
Table 2.3 Model summary for $Geospatial_{Full}$ and *Odds* with LLR test statistic

Date	Model	Observations	LL value	R^2	LLR test
1997-1999	$Geospatial_{Full}$	66496	-11877.7342	0.1432	**31.74****
	$Odds$		-11893.6018	0.1420	
1998-2000	$Geospatial_{Full}$	70082	-12340.9592	0.1430	**25.42****
	$Odds$		-12353.6695	0.1421	
1999-2001	$Geospatial_{Full}$	74253	-12884.7748	0.1458	**23.36****
	$Odds$		-12896.4496	0.1451	
2000-2002	$Geospatial_{Full}$	80271	-13486.8197	0.1560	**18.69****
	$Odds$		-13496.1641	0.1555	
2001-2003	$Geospatial_{Full}$	89214	-14874.5080	0.1597	**9.90****
	$Odds$		-14879.4565	0.1594	
2002-2004	$Geospatial_{Full}$	96461	-16209.5784	0.1609	**7.42****
	$Odds$		-16213.2879	0.1607	
2003-2005	$Geospatial_{Full}$	105928	-18329.5944	0.1546	**5.74***
	$Odds$		-18332.4655	0.1544	
2004-2006	$Geospatial_{Full}$	119098	-21001.9451	0.1507	2.69
	$Odds$		-21003.2886	0.1506	
2005-2007	$Geospatial_{Full}$	132700	-23927.3628	0.1458	2.18
	$Odds$		-23928.4541	0.1458	
2006-2008	$Geospatial_{Full}$	140078	-25521.0345	0.1493	0.32
	$Odds$		-25521.1932	0.1493	
2007-2009	$Geospatial_{Full}$	139055	-25873.8529	0.1496	0.01
	$Odds$		-25873.8539	0.1496	
2008-2010	$Geospatial_{Full}$	139814	-26484.8798	0.1456	0.06
	$Odds$		-26484.9121	0.1456	
2009-2011	$Geospatial_{Full}$	139431	-26971.5328	0.1405	0.30
	$Odds$		-26971.6834	0.1404	
2010-2012	$Geospatial_{Full}$	140752	-27513.8371	0.1401	0.11
	$Odds$		-27513.8932	0.1401	
2011-2013	$Geospatial_{Full}$	138286	-27059.9046	0.1422	0.49
	$Odds$		-27060.1484	0.1422	
2012-2014	$Geospatial_{Full}$	134344	-26541.3963	0.1423	1.65
	$Odds$		-26542.2232	0.1423	

** and * denote significance at 1% and 5% levels respectively in a 2-tailed test.

Furthermore, the LLR test statistic $\chi 2$ values decrease steadily over time (see Figure 3), demonstrating the diffusion of symbolic information into market prices. In Figure 3, a vertical line at 2005 separates the 'Diffusion Stage', a period when $Geospatial_{Full}$ provides significantly more information than $odds$, and the 'Post- Diffusion stage', when market odds fully incorporate the valuable information regarding racetrack topology offered by VGs.

Figure 2.3 Diffusion of technology enhanced information over time



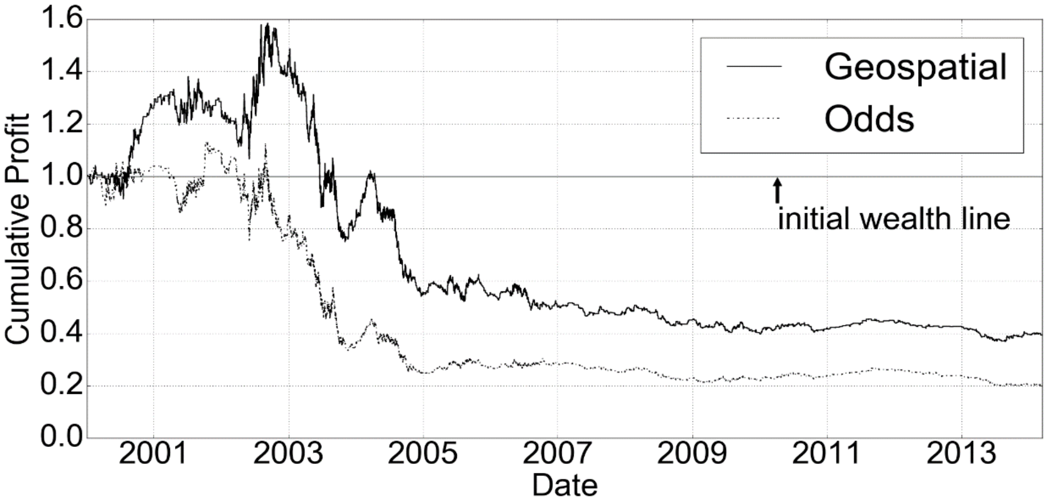
These results lead us to reject H2, namely that VG information diffuses through a financial market completely and immediately once available to the public. In particular, since topographical data was available to the public from 2003, the fact that the LLR test statistics for all years up until 2005 were significant, demonstrates that this information was not fully discounted in odds for two years following its release.

The results, therefore, demonstrate a delayed diffusion process, where the full value of information is only realised over a number of years. This finding highlights the need for research that explores the value of IT to consider longitudinal timeframes.

2.4.3 Economic value of technology enhanced decision-making

Betting simulations based on probability estimates derived from employing VG information to account for horses' topological preferences show the economic importance of VG information. The profit achieved by betting strategies will provide a measure of the economic value of the symbolic-enabled decisions of bettors using VG data. In particular, estimated winning probabilities from $Geospatial_{Full}$ instruct a Kelly betting strategy over each one year hold-out sample. For example, 3 years of data (1997-1999) are used to estimate the $Geospatial_{Full}$ coefficients and then these coefficients provide winning probabilities throughout the following year (2000). A Kelly betting strategy for 2000 is based on these predicted probabilities. Subsequently, the next three years of data (1999-2001) are used to estimate $Geospatial_{Full}$ model and to predict winning probabilities throughout the following year (2001). This process continues for each out-of-sample dataset from 2000-2014. The results are displayed in Figure 4.

Figure 2.4 Cumulative wealth from a Kelly betting strategy based on winning probabilities estimated using information available in VGs and the Kelly betting strategy based on Odds



Between 2000 and 2003, the Kelly betting strategy based on Geospatial information produced positive profits. The cumulative wealth rises to a maximum of 1.59 times initial wealth in July 2002. However, when the elevation data was made available to the wider public in VGs in 2003, the profit levels decrease. After 2005, the VG information was completely diffused into the market as the cumulative wealth based on a Kelly betting strategy dropped below the initial wealth level.

To confirm the additional benefits offered by employing VG information when estimating winning probabilities, the results of betting simulations based on probability estimates derived from employing two models are compared: (i) using VG information to account for horses' topological preferences, together with market odds (i.e. $Geospatial_{Full}$) and (ii) using market odds information alone (i.e. $ODDs$). The $Geospatial_{Full}$ model provides consistently higher returns compared to those from $ODDs$, as is depicted by the difference between the two lines in Figure 4. In 2001, 2002, 2003, 2004 and 2005 the VG information provides 22.12%, 19.56%, 60.14%, 48.06% and 30.25% higher returns than those generated simply with $ODDs$, before plateauing at about 20% higher from 2006-2014. Therefore, using the VG information, betting strategies achieve consistently higher returns than using market odds information.

The cumulative returns show that the information provided by VGs was not incorporated into the market between 2000 and 2003 and there were considerable opportunities to profit from the VG information. Once the information became freely available to the public, the market began to incorporate the information and the profitable opportunities were eroded.

2.5 Discussion

These results show that information from VGs can be used to improve symbolic decision-making tasks to improve forecasts, enabling profits. The study, therefore, provides empirical support for the view that 3-Dimensional maps can support complex decision-making tasks. Most studies examining the value of 3-Dimensional maps have underlined the characteristics of map-based presentations of information which can facilitate human decision-making capabilities and have *theorized* how such visualisations can improve decisions associated with complex problems (Shen et al. 2012; Zhu and Chen 2005). This study shows clearly that VG information can be used to support symbolic decision-making in the real world. The study also provides a means of measuring the impact of this information (Benbasat and Zmud 2003). Consequently, the study contributes to the VG/decision-making literature (Drnevich and Croson 2013), by providing empirical evidence that VG information improves symbolic decision-making capabilities and leads to value creation.

The results show how VG-based information diffuses through the market. Although the geospatial data was freely available in 1999, it took the market a number of years to (fully) discount the information. Even when the information became available to the wider public (in VGs), it took two further years to be fully discounted by the market.

2.5.1 Implications for Research

The study helps contribute empirical evidence on how VGs can be used in decision-making (Drnevich and Croson 2013) and the impact that technology-enhancements can have on financial markets (Zhang and Zhang 2015). Similar research proposes that GPS-enabled information provides more fact-based reasoning for decision-making (Habjan et al. 2014). This paper shows *how* information from VGs can be used in symbolic decision-making tasks that require analytical and quantitative precision and that VG information produces more accurate forecasts compared to those forecasts generated from market odds alone.

The majority of research on diffusion has explored to what extent the underlying population have adopted technology-enhanced information (Faria et al. 2003). This paper explored the comparatively less researched aspect of market convergence and exploring that rate at which information is incorporated into the market (Chordia et al. 2005). Few studies have developed theoretical basis on the rate at which information is diffused into the market and one theory is that slower-diffusing information leads to high short-run returns (Hong and Stein 1999). Typically, such models on information convergence have been laboratory based and purely theoretical (Hong and Stein 1999). Therefore, this paper offers empirical evidence on market convergence and the rate at which the horse racing betting market adjusts to VG information over time.

This study highlights how technology can enhance the information environment. This is important because the study demonstrates that IT can create inefficiencies as the market adapts to new information. The assumption of the EMH, that market prices reflect all new information instantly, has been critically questioned by a number of economists (Lim and Brooks 2011).

The Adaptive Market Hypothesis (AMH) has emerged as an extension to the EMH (Lo 2012). AMH assumes that market efficiency is related to the environment and to the adaptability of market participants to emerging technology (Lo 2005). Importantly, when environmental conditions change, AMH indicates that investor behavior should change in response. This results in a period of learning the new heuristics that are better suited to the new environmental conditions (Lo 2004). This learning process is linked to human behavior and complex issues of how individuals make decisions. AMH indicates that as individuals learn and adapt to new processes and as competition drives adaption, the "prices reflect as much information as dedicated by the combination of environmental conditions and the number and nature” of investors in the market (Lo 2005, p.31). This implies that markets are not always efficient and, in fact, vary in their degree of efficiency (Lo 2012). As such, profitable opportunities do arise from time to time but these opportunities are eroded as the IT diffuses and competitors learn to adapt.

This study confirms that technology enhanced information diffuses through a market over time and that the market slowly becomes efficient in respect to this information. The economic value that arises as betting strategies take full advantage of information allow for profits that decline over time as the market adapts. The results are, therefore, very much in support of the AMH.

One important consideration is that horse race betting markets are uncertain and dynamic environments (Johnson and Bruce 2001). Uncertainty related to the uniqueness of each event in terms of the participants, location, conditions and a range of other factors affect the predictability of participant’s performance. The market is dynamic in the sense that information related to these aforementioned uncertainties is constantly updating and the changing odds during the build up to the race reflect the dynamic updating of new information to the market (Johnson and Bruce 2001). Although there are a multitude of factors related to the outcome of any race, it proves inherently difficult to capture every effect. The idea that markets are dynamic contributes to the argument in the paper that new information is released which improves on prior estimates of the winning probabilities. At such times, when new information is available the market is constantly adapting to this new information and profitably opportunities arise. In this sense, markets are in a constantly state of flux, adapting to changing environmental conditions.

In line with the call for longitudinal analyses to assess the impact of technology and financial performance (Chae et al. 2014; Sabherwal and Jeyaraj 2015), the impact through time of VGs is shown. By looking directly at market level, this study enables us to identify clearly how the information diffuses over time and its relative economic value. The study also demonstrates how the potential benefits of technology are competed away as the market learns how to use VG information in symbolic decision-making tasks.

When evaluating the success of technology investment, payoffs of IT are not always fully realized instantaneously and must be estimated in a longitudinal manner. However, there is a lack of longitudinal studies (Sabherwal and Jeyaraj 2015). Consequently, there is a need to examine longitudinal data to observe the lagged effects (Devaraj and Kohli 2003). Results show that information diffusion is delayed and the economic value of IT is linked to how quickly the market learns to adapt to new technology. Clearly, therefore, senior managers and IT executives must explore the value of IT over time if they are to truly estimate IT's payoff.

2.5.2 Implications for Practice

In relation to the diffusion of types of information over time, isolating the effects of one type of information such as VG information is difficult because of the environment within which diffusion takes place is itself subject to significant change through time. Indeed, the information diffusion that is depicted in the results could have been attributed to the rise in online bookmaking services such as Betfair and Bet365, founded in 2000, which changed the nature of the horse race betting market (Johnson et al. 2010). With the information environment constantly changing, new bookmakers may have had an influence on the overall horse race market efficiency however, this would still show to some respect that market discounted geospatial information.

The overall environment within which diffusion takes place is itself subject to significant change through time. Since web information diffuses at varying speeds depending on the region and industries in which the diffusion is taking place, isolating diffusion is non-trivial (Papagiannidis et al. 2015). With regards to the statistical tests on whether the odds contained all information related to the geospatial variables, one might argue that the drop in performance of the prediction model over time might be due to changes in other covariates related to horse racing and not due to the diffusion of geospatial information in the market. For instance, horses along with their riders are not static, which meant that they might improve over time through training to overcome challenges that persist with certain terrains. As such, the dynamic nature of the performance of horses and riders may explain the declining ability of geospatial information in predicting performance. The results in table 2 show that the geospatial variables account for reasonably consistent levels of explained variance in the winning estimates. Although there is a dip in performance after 2008, where the variables are less predictive of winning estimates, the simple model employing just geospatial variables is most predictive in 2005-2008, the same years in which the market becomes efficient. As such, the minor variations in the R^2 performance of the geospatial variables reported in table 2 show that the models do not deteriorate over time and that the declining diffusion rate evidenced was not linked to any reduction in predictive value of the geospatial variables.

Another possibility is that terrains of race courses might change over time with maintenance efforts or with erosion. Elevation data was collected in 2010 and again in 2018 to see how much the data changed over time. These two samples are consistent showing that the prediction model is built on spatial information that does not change considerably over time. Thus, the possibility that potential changes in terrain over time might have resulted in the outcomes observed is negated.

Attributing the declining trend of profit levels to be a consequence of information diffusion in the market is not without limitations. An important assumption inherent in the interpretation of the findings is that bettors are indeed utilizing geospatial information to make betting decisions. Given the complexity involved in the calculation of predicted winning probabilities, it is unlikely that the average individual bettor would be able to utilize the sophisticated prediction method to make informed decisions. Indeed, bettors suffer from bounded rationality (Simon 1955) and they are incapable of processing all the information available to make a decision, but rather, bettors suffice with decisions that are based on the limits of their cognition and time pressures. Rather than using data related to the decisions and betting behavior from individual bettors, the analysis has focused on the rate at which the aggregate-level market incorporates information.

Regarding prediction, there is not clear evidence that bettors did use geospatial information to assist their betting decisions. Even if bettors did use geospatial information to make predictions, whether they received such information from VG or other sources is also a question. It is entirely possible that they made their betting decisions based on past experience or other information sources such as peers, experts, or the media. Further research employing individual level betting patterns would answer such questions and link information availability to information use.

2.6 Conclusion

Information diffusion is important in the context of modern markets since information is constantly available online. Information transfer is key not just within markets but between markets as research has shown that similar commodity futures and equity markets are interconnect to each other over time forming a financial network (Bekiros et al. 2017). Information diffusion is important not just in financial markets, but in a wide range of fields including disaster response, product diffusion and epidemiological studies (Nagarajan et al. 2012). The method employed herein can be used to understand diffusion in different settings and to show how technology can be key to information dissemination.

The full value of the new technology-enabled information may be diffused through the market over a period of years. However, the economic gains from new technology decrease almost immediately when information becomes available and in this way, the market continually adapts to new technology-enhanced information as such information becomes available.

This longitudinal study over eighteen years demonstrates the manner in which information from VGs can be used for improved symbolic decision-making tasks and how the information diffuses through the market. The information was quickly discounted by the aggregate market odds when the information became available to the public (within two years).

This paper shows how information from VGs can be used to generate variables that improve the predictive performance of a model beyond those predictions based on odds. New information is readily available online and offers the potential to improve forecasts. The time for market convergence towards complex forms of information may take considerable time and although VG information can significantly improve forecasts, the market quickly adapts and any market inefficiencies are likely to be short lived.

Chapter 3 How new distance information that is readily available online can be used to elicit knowledge from experts and produce improved forecasting accuracy compared to the betting crowd

Abstract

This paper shows that new web information has the potential to enrich decision-making performance of distinct sub-groups of decision-makers. The internet, and tools such as Google Maps, have provided access to mapping information that can be harnessed to improve decision processes. Particularly, geographical information from online tools contextualises the decision-making of trainers (experts) in horse racing, and the extent to which this nuanced information can be used by 'the crowd' to improve their own forecasts. By exploring to what extent the betting public discerns forecast-relevant information from the decisions made by nearly 500 racehorse trainers, this paper shows the value of incorporating new web information in decision-making. Data related to 82,703 races (involving 84,939 different horses) run in the UK between 1998 and 2010 shows the extent to which the distance information is informative of trainer actions and how the public could have employed this information to improve their betting decisions. The results demonstrate that distance information can be used to imbue predictive information from expert behavior to achieve more accurate forecasts than the betting crowd for a limited time.

Keywords: Forecasting, Decision Analysis, Experts, Wisdom of Crowds, Belief elicitation, Distance

How new distance information that is readily available online can be used to elicit knowledge from experts and produce improved forecasting accuracy compared to the betting crowd

3.1 Introduction

Burgeoning research on how new technology, and particularly the web, have improved information availability and transparency illustrate the new research directions on socio-economic activity that are made possible by the access to increased amounts of information (Chang et al. 2014). As the total volume of data increases, more information related to behavioral and economic activity at the individual and market level is being generated that enables research at the aggregate market level and on group-level behavior.

Recent forecasting literature has been concerned with how new online information can be used in forecasting with a heavy focus on using Google search related information (Jun et al. 2017). Technology advancements in web-based services has both exponentially increased the amount of data produced and, advanced abilities to capture such data. Examples of how new online information is improving forecasts abounds: Google search is being used to predict oil market prices (Yu et al. 2018), twitter data is being used to predict political elections (Huberty 2015) and social networking data is being used to predict box office figures (Kim et al. 2015). Although there is a large body of research on how tools such as Google search can be harnessed for forecasting, the use of Google maps is considerably less dense (Jun et al. 2017). New information online offers the potential to generate more valuable insights by providing hitherto unknown data that adds new perspective on decision-making problems.

The web has transformed how individuals gather information and make decisions

(Wallenius et al. 2008) and location based services and online maps are prime examples of data sources that have transformed individual decision-making and behavior (Constantiou et al. 2014). Geographical information systems provide ready access to distance and mapping information that is used in a wide array of everyday decision-making tasks, from deciding which restaurant to visit, to route planning and emergency service response. The widespread use of these facilities is highlighted by the fact that the global geospatial analytics market was estimated to be £18.71 billion in 2015 and is predicted to grow to £49.61 billion by 2020[6].

This paper specifically focuses on the use of distance-related information from online information services. The role of new information to improve forecasting accuracy and decision-making is often overlooked. Indeed, in a comprehensive review of the forecasting literature, the main categories of research that would contribute to understanding information diffusion for improved forecasting accuracy were identified as multigenerational models and multinational models as well as forecasting with little or no data (Meade and Islam 2006). However, the proliferation of data available on the internet has ushered in a new research paradigms that is characterized by data-driven and data-intensive studies of the range of tools that improve our ability to make more 'felicitous' decisions than ever before (Wang et al. 2016). This paper argues that the such tools as Google maps that are frequently used for everyday decision-making offer a suitable outlet to study how technology improves our forecasting ability and it is, particularly, the new data that affords this 'felicitous' decision-making.

For example, individuals often prefer to shop online rather than going to stores, paying extra for the added convenience of delivery; making a conscious trade-off between

---

[6] http://www.researchandmarkets.com/research/zz7w3b/geospatial [Accessed 07/06/2016]

distance/time and price (Cachon 2014). Equally in holiday destination planning and residential location choice, conscious trade-offs based on distance to local amenities and attractions often influence decisions (Ho et al. 2015). Decision-makers implicitly and explicitly make trade-offs between near/probable benefits and distant/unlikely rewards (She et al. 2012). Therefore, individuals frequently make decisions where they have to weigh up uncertainty and the benefits of various multi-choice outcomes where distance varies.

In uncertain decision-making environments, novices frequently turn to experts to help them make better decisions (e.g., in stock market investment). However, Camerer and Johnson (1991) identified a "process-performance paradox" whereby experts, despite their knowledge, can produce forecasts that are less accurate than 'the crowd'. This paradox has been found in recent studies where experts failed to outperform non-experts in sports forecasting (Andersson et al. 2005) and in stock market prediction (Torngren and Montgomery 2004). Therefore, whether experts are best able to incorporate new web based information to improve forecasting is uncertain.

The 'Wisdom of Crowds' has gained traction in recent years, because it has been shown that the aggregated opinion of the crowd is often at least as accurate as that of any individual expert (Surowiecki, 2005). It is often assumed that this arises because different members of the crowd contribute small but different pieces of information, and the aggregation of all these decisions produces superior forecasts. However, the role of experts in this process may have been under-estimated and that the crowd's forecasting ability may in part stem from its ability to learn from the behavior of experts. As Coussement, Benoit, & Antioco (2015, p. 24) observed, the "impact of integrating expert opinions into the decision-making process has not been sufficiently investigated" and understanding how experts and the wider public differ in their use of information is important (Chen & Zeng, 2016; O'Keefe, 2016). Consequently, the aim of this study is to examine the manner in which distance based information online can be used to frame the decisions of experts and the value of this information if the public were to incorporate distance based information into their own forecasts.

Analysis shows to what extent experts make informed choices related to distance and the extent to which the crowd learns to employ valuable information discerned from experts' distance-related behavior. The horserace betting market provides the ideal empirical setting to examine these issues. Trainers decide in which races to enter their horses and weigh up the distance that the horse would have to travel to the venue. For the average horse, the total cost reported in 2015 for transport was £1,498 per year, which worked out at 6.6% of the total annual costs for the horse, representing the third highest cost after training fees and entry to races. [7]

A trainer has access to privileged information about the quality, fitness of his/her horses and how they are likely to respond to travelling long distances. In addition, trainers are unlikely to incur the additional travel and staff costs of sending a horse a considerable distance unless they believe they have a reasonable chance of winning. Equally, the betting public can observe the distance that a trainer has sent a horse and can attempt to use this to discern the trainer's beliefs concerning his/her horse's chance of success. Therefore, it is necessary to explore simultaneously how experts and the crowd use distance information in a repeated decision-making environment. The advantage of betting market data is that for a given race accurate estimates of the subjective probabilities characterise the public's belief for each of the runners (odds). Importantly, there is a defined end-point at which all uncertainty is resolved and a winner is declared. By observing a large number of races, the

---

[7] https://www.roa.co.uk/en/owners-resources/becoming-an-owner/annual-training-costs.cfm [Accessed 10/02/2018]

appropriateness of individual trainer's distance-trade-off decisions are chronicled and the extent to which the public accounts for this is evidenced by assessing the degree to which it is discounted in their probability forecasts (as displayed by market odds).

The vast majority of research comparing expert and non-expert decision-making has been conducted in laboratories (Hao and Houser 2011; Schotter and Trevino 2014). Laboratory experiments offer the advantage of comparing subject populations that typically include mixture of distinct sub-group populations including experts and novices. While it is of important research interest to elicit information from these two sub-group population, a necessary assumption in such laboratory studies is that people respond optimally to incentives (Hao and Houser 2011). This assumption is somewhat stretched since a number of studies have reported that laboratory subjects respond in dishonest ways to either make explicit idealisations or offer desirable responses with the intention of affirming the research questions inherent in studies (Schotter and Trevino 2014). Also, laboratory subjects may be offered incentives and fake money that bias behavior itself as the responses may be quintessentially different from those actions that would be undertaken if the subject was experiencing and suffering real economic gain and loss.

Indeed, it is necessary to study the consistency between beliefs and actions to affirm that respondents are behaving as they say they would, and in this sense, actions may offer a better indication of what is important to research subjects. A further complication of laboratory studies is that the behavior of subjects may be biased by the laboratory itself as subjects offer different responses than they would if their behavior was not being observed (Schotter and Trevino 2014).

By contrast, this paper explores decisions made in the field by experts, without burdening them during the decision-making process and without the danger that they alter their behavior through being aware that they are being observed (Werner et al. 2017). In addition, the decisions of both the experts (trainers) and the crowd (the betting public) involve the prospect of monetary gains, which are meaningful to them. Further, the horse race setting offers many thousands of decisions over an extended period, enabling the degree to which the public learn to improve their subjective probability estimates to be portrayed over time. The elicitation of experts' judgements is fundamental to building decision-making models that capture real world systems (Werner et al. 2017) and the field experiment provides empirical evidence to supplement the largely experimental (laboratory based) research on how experts' models are used by non-experts to make decisions (O'Keefe 2016).

The paper makes three contributions: First, how new distance information made available online through tools such as Google maps makes it possible to illustrate how experts make informed distance-based trade-offs under uncertainty. Second, the value of using distance information in eliciting the subjective beliefs that underlie expert decisions. Third, how distance-based data can be used to show how the public improve their probability estimates by fully incorporating information from distance in their aggregated decisions. Consequently, this paper contributes to the growing literature on using new web information in forecasting (Jun et al. 2017) by showing how distance information from Google maps can explain decision-making of novices and experts.

The remainder of this paper is organised as follows: The literature examining the decisions of experts and 'the crowd', and the importance of distance-related decisions are discussed in section 2. These literatures are used to establish the hypotheses. The data and methodology employed to test these hypotheses are described in section 3. Results are presented in section 4 and discussed in section 5. Conclusions are drawn in section 6.

3.2 Distance-related decisions and expert and crowd decision-making

First, the literature review shows the importance of distance in many decisions but highlights that the role of distance has been under-researched. Second, literature on expert decision-making illustrates how the performance of experts compares with that of the crowd in uncertain environments. Third, the review shows how experts and the crowd differ in their decision-making by focusing on the information they prioritise. These literatures are used to develop two hypotheses.

3.2.1 Spatial distance

Space and time are fundamental to most economic decisions (Cachon 2014). Construal Level Theory (CLT) promotes the idea that individuals are capable of thinking beyond the 'here and now' by forming mental abstractions of distal objects. Psychological distance is a mental representation of how distal an object is from ourselves in terms of time, space and social distance (Trope and Liberman 2010). Most of the extant research has focused on the importance of time in decision-making. For example, Kocher and Sutter (2006) and Leclerc et al. (1995) have sought to make direct comparisons between time and money, even conflating them under the idea that 'time is money'. Overall, this literature suggests that individuals have similar risk preferences for time and money but have different valuations for monetary and temporal loss (Abdellaoui and Kemel 2014; Ebert and Prelec 2007).

Clearly, when considering travelling distance, the time involved is important. However, this is not the only consideration and despite the wealth of research on time and money, there is little research examining the impact of distance on decisions and decision processes. She et al. (2012, p.266) confirm that “the research on the spatial dimension and preferences in decision-making has not been done”.

Ultimately, all other aspects being equal, humans will choose the proximal option according to Zipf’s law of least effort (Zipf 1949). This preference for local solutions explains why consumers prefer closer shops (Fotheringham 1988), why geographically closer business acquisition targets are often selected (Chakrabarti and Mitchell 2013) and why burglars often commit crimes in locations close to where they live (Bernasco et al. 2013). Clearly, if the rewards at distant and proximal locations are the same, the former are likely to be selected as these maximise the work done/reward ratio. In these discrete location-based choice examples, there is a simplistic mode of evaluation because each choice has similar features and the focus shifts to reducing cost in terms of distance.

However, in most situations, such choices are inherently more complex and a multitude of factors must be evaluated. The nature of uncertainty adds complexity to the distance trade-offs and further research on distance-based decision-making under uncertainty in the real world is needed to understand to what extent individuals make appropriate decisions.

3.2.2 Expert and novice performance

There is mixed evidence concerning whether experts’ forecasts outperform those of non-experts or novices. Despite their apparent superior information, there is evidence that experts’ forecasts often fail to outperform those of novices (e.g., Camerer and Johnson, 1991). This performance-paradox has been show to include the forecasting of competitive sports event outcomes (Andersson et al., 2005; Spann & Skiera, 2009). Equally, Torngren and Montgomery (2004) found that experts were not able to outperform the public in predicting better performing stocks. The experts often used specific information related to stocks in their predictions without understanding the reliability of this information, while novices often used simple heuristics based on the past stock prices. The fact that both groups achieved similar performance, suggests that fast and frugal decisions based on simple information can perform as well as decisions based on specific (and possibly unreliable) information. Similarly, Leitner and Schmidt (2006) showed that professional forecasters were unable to beat naïve forecasts of exchange rates because experts were influenced by irrelevant fundamental information. Collectively these studies suggest that experts often use misleading and inaccurate information as the basis for their decisions and, as a result, they often fail to make better forecasts than novices.

3.2.3 Observing experts and updating prior beliefs

Whilst most decision-making under uncertainty literature assumes that beliefs drive behavior, "it is less obvious how to elicit and measure such beliefs" as much of the literature has bypassed "the thorny problem of how to interpret direct expressions of beliefs" (Fox and Tversky 1998, p. 879). It is commonly held that belief precedes preference (Fox and Tversky 1998) and that the availability of preference data should enable us to predict how individuals will behave (Manski 2004). As such, there is wide interest in predicting people's behavior and a range of literature on how to elicit beliefs from individuals. For many uncertain events, it is difficult to derive objective probabilities and so it is valuable to make decisions based on knowledge of the subjective beliefs of others who are familiar with the situation, especially experts. Research has typically utilized elicitation methods and hypothetical choice data to understand how individuals choose between discrete options. Respondents are typically asked to choose between various prospects or lotteries and their responses are used to predict how they would behave (e.g. see Andersen et al., 2014 and Festjens, Bruyneel, Diecidue, & Dewitte, 2015).

These introspective methods suffer from a number of serious pitfalls. For example, respondents often misrepresent their beliefs due to social desirability and may respond intentionally to influence what they believe is the object of the research (Trautmann and Kuilen 2014). Various incentive schemes have been designed to overcome these difficulties to promote accurate responses and these incentivized approaches have been shown to lead to better predictions of individuals' behavior under certain conditions (Trautmann and Kuilen 2014). Whether or not incentivized methods elicit true beliefs is questioned by Schlag et al. (2013, p. 18), who point out that individuals do not always act in the way they say they will react. Consequently, elicited beliefs may not be a reflection of the actual underlying beliefs (Costa-gomes and Weizsäcker 2008).

In the real world, crowds can overcome these concerns and gain insight into the subjective beliefs of experts by observing their behavior. The pitfalls of laboratory-based enquiry are clearly less acute in these situations. By incorporating theoretic insights related to behavior, empirical work on decision-making can draw on the pragmatic nature of individual behavioral processes (White 2016). Exploring real world behavior also avoids the problem reported by Viscusi & Evans (2006), that "the probability values revealed through individual behavior are not the same as those reported by respondents".

The private information held by experts may not be fully discerned simply by observing the distance-based trade-offs they make. Rather, the trade-offs they make may need to be processed in order to discern the information held by the expert (i.e. to discern its latent value).

Despite the forecast performance-paradox related to experts/non-experts appearing in a number of different fields there is also evidence to suggest that under some circumstances, experts do make superior forecasts. For example, experts have been shown to make superior currency forecasts (Önkal et al., 2003) and to better forecast various other business, economic and political events (Budescu & Chen, 2014). Alevy et al. (2007) found that one of the reasons for this was that experts are better at using public information and discerning price signals. In situations where experts are privy to information not available to the public they are likely to make informed choices. This view is tested in relation to distance-based trade-off decisions by testing:

Hypothesis 1 (H1): Experts make informed distance-based trade-off choices that can only be fully discerned by processing their distance-related trade-offs to understand their preferences and elicited beliefs

3.2.4 Wisdom of crowds

Surowiecki's (2005) book, *Wisdom of Crowds,* popularised the idea that under the right conditions a group of non-experts can make better decisions or more accurate predictions of an unknown variable than any individual (including individual experts). The argument is that there is the potential for each member of the crowd to bring a small but different piece of information concerning the event or decision and when these are combined, they can produce superior decisions/predictions than those of experts. Surowiecki (2005) suggests that there are four key characteristics needed for a wise crowd to form: diversity, independence, decentralization and aggregation. Diversity requires that individuals have their own view of the problem; independence that individuals are not influenced by the opinions of others, decentralization requires that individuals are able to draw on specialized or other available information and aggregation requires a mechanism to turn the diverse views into a collective decision.

'Crowds' have proven to be remarkably accurate in real world prediction; for example, in forecasting inflation and market growth (Budescu and Chen 2014) and in predicting future stock returns (Chen, De, Hu, & Hwang, 2014). Equally, Eickhoff and Muntermann (2016) show that the aggregated opinion on social media of stock market price movements can outperform stock market analysts and Gottschlich and Hinz, (2014) demonstrate that a decision support system incorporating the votes of the crowd related to stocks outperforms public fund managers. This research suggests that experts do not under-perform, rather that the crowd manages to 'over-perform' in relation to the apparent accumulated expertise of the participants. The crowds' wisdom may stem partially from its ability to discern information by observing the behavior of others, particularly experts with access to private information.

How individuals use the opinions of others to update their own beliefs and make decisions is important for understanding the decision-process (Mannes 2009). Lawrence et al. (2006, p. 511) call for more research into "how people acquire and use information when they make forecasts and the effects of differences in the availability of information". One might suspect that access to more relevant information might be associated with improved forecasting accuracy. However, Andersson et al. (2005) found that providing more information had no effect on the forecasting performance of non-experts. In addition, Ecken & Pibernik (2016) found that decision-makers often ignore the advice of experts, which can lead to inaccurate judgements and poor decisions. Ecken & Pibernik focused on individual decision-makers. Consequently, how crowds respond to information gleaned from experts remains uncertain. However, whilst crowds might be able to discern valuable subjective data from experts' choices they may also not fully appreciate the value of this information:

Hypothesis 2 (H2): Privately held information related to expert's distance-based trade-off choices provides improved forecasting accuracy compared to the crowd's best estimates

3.3 Methodology

The hypotheses are tested in a real world repeated decision environment that is characterised by uncertainty. Distance-related decisions made by experts and the crowd are analysed to understand (a) the degree to which the experts make informed distance-related choices, and (b) to what extent the crowd extracts the valuable information contained in the distance-related choices of experts.

The horserace betting market provides a perfect setting in which to test the hypotheses. Betting markets largely fulfil the four conditions required for 'wise crowds', namely a diverse set of decision-makers, each using their own combination of information to make their judgements and their betting decisions. They are also able to draw on specialized information (e.g., from specialist publications) and the market mechanism enables their individual judgements and actions (bets) to be turned into a collective view (market price).The market thus effectively aggregates individuals' judgements and provides an incentive for individual members of the betting public to make decisions based on their information.

Racehorses are trained by experts (trainers) who make repeated decisions on where to send their horses to race in order to maximise their chances of winning. The result of each race provides instant economic feedback to the trainer regarding the success of his/her decision (e.g. in terms of prize money gained and the cost of sending the horse to the track) and s/he is in a perfect position to decide if the horse performed as well as s/he had expected. Racehorse trainers also provide ideal subjects for examining the performance of 'experts' in uncertain environments as they are familiar with dealing with uncertainty (Friedman and Savage, 1948). For example, they make their income and reputation from being able to subjectively interpret multidimensional information in order to accurately assess the probabilities of their horses' chances of success and to weigh this against the cost of running a horse in a particular race. Consequently, they have a strong incentive to make accurate subjective probability judgements concerning their horses' chances of success. In addition, they are likely to make these sorts of decisions more frequently than most members of the betting public, enabling them to gain a level of expertise in this task. Importantly, racehorse trainers also have access to information which is not available to the public (e.g., the well-being of a horse, whether it has sustained an injury, how well it has run in training, whether it is likely to respond well to travelling longer distances etc.). Racehorse trainers make informed decisions regarding the trade-off between spending more time/resources in sending a horse a long distance in order to compete and, perhaps, the greater chance it might have of winning than at a more local racetrack. Given their access to privileged private information and their familiarity with making such trade-off decisions involving uncertainty, horse race trainers have superior information. The betting public may be aware of this and may use a trainer's decision to send a horse a long distance as a signal of its prospects of winning; this might particularly apply for trainers who the betting public are aware (by their past performances) are particularly good at making these trade-off decisions. The methodology of this paper shows the added benefit that the betting public could generate if they did use distance information to understand trainer's behavior.

3.3.1 Data

Some previous studies on empirical decision-making processes, particularly, the processes used by experts, have been accused of using small and unrepresentative samples

(Kemel and Paraschiv 2013). However, focusing on the decisions of UK horserace trainers affords us the opportunity of examining the decisions of nearly 500 trainers. Further, a competitive market place operates on each race so that information from a variety of sources is incorporated into market prices, including past decisions made by trainers. The odds in the betting market represent the betting public's aggregated view as to the probability of success of each horse. Horseracing also affords the advantage that there is a point in time (completion of the race) when all uncertainty is removed (a winner is declared). This affords us the opportunity of using the results of a large number of races to measure the success of decisions of individual trainers in sending their horses different distances to race. A trainer's decision to send a horse to a specific race depends on a number of factors including day of the race, days since last race, race entry costs and distance to race. While it proves problematic to capture all the variables integral to the trainer's decision-making process, the transport costs represent the second highest variable cost after cost of race entry (data for which is unavailable to the public). [8] Therefore, the study tests to confirm the assumption that distance is a significant variable in the horse entry decision by showing how the distance information provides nuanced meaning to explain the decisions made by trainers. The paper does not try to capture the myriad of factors that may affect a trainer's ultimate decision but rather epitomise the added benefit from distance information.

It also enables a comparison of the ex-ante subjective probability estimates of the crowd (the betting public) with the ex-post, realised probabilities of success of particular groups of horses (e.g., those of a particular trainer which have travelled a certain distance to race). To test empirically whether the distance a horse is required to travel to race by a

---

[8] https://www.roa.co.uk/en/owners-resources/becoming-an-owner/annual-training-costs.cfm [Accessed 10/02/2018]

particular trainer provides information concerning its chance of winning and to what extent the betting public account for this information in their betting decisions.

Data was collected on the performances in 82,703 races of all 84,939 different horses of all 495 English and Welsh horserace trainers, throughout the period 1998-2010 (inclusive). The population of trainers remained relatively fixed throughout the whole period. Using postcode level data to calculate the distance from each training establishment to each racetrack venue.

In particular, the distance to each racetrack from each trainers' establishment is measured using Google Maps Distance Matrix API.[9] This service provides routing information which is often used by the trainers themselves and allows us to calculate the distance based on the recommended shortest route to each of the 38 horserace venues.[10] Table 1 presents summary statistics for the 495 trainers and shows that the prize money won and distance travelled by trainers vary greatly. [11]

Table 3.1 Summary statistics for the 495 active trainers between 1998-2010

	Races Entered	Win Rate (%)	No. of horses	Total Prize Money (£)	Average Prize Money Per Race (£)	Average Distance Travelled per race (km)
Min	1	0	1	0	0.00	0
Mean	1,160	6	143	1,180,120	1017	97
Max	15,214	27	1881	33,961,601	853,600	451
Standard Deviation	1771	4	225	3,256,235	5935	64

9 https://developers.google.com/maps/documentation/distance-matrix/intro

10 Distance from A to B measured 'as the crow flies' was also tested in the analysis and provided similar results. However, the travelling distance was used because this gives a more realistic measure based on the road infrastructure.

11 A map showing the locations of both the racetracks and the trainers' locations is included in Appendix 1.

If all trainers had only sent their horses to their nearest racetrack when their horses ran, then they would have collectively sent their horses 10,153,979km in the period 1998-2010 rather than the 61,866,824 km they actually travelled. There is clearly greater cost/time involved in travelling further and horses run greater chance of injury. However, previous studies have shown that experts often accept greater risk in order to reap greater reward (Petropoulos, Fildes, & Goodwin, 2016; Basu & Nair, 2015). Consequently, the extra travel costs/risks incurred by a trainer when sending a horse a given distance may reveal something concerning their view as to the horse's probability of success.

3.3.2 Distance related decision-making

Distance to a racetrack is important since travelling longer distances will incur higher effort and costs (Zipf 1949). The longer distances travelled by horses will incur higher fuel costs, longer travelling time for the trainers' staff and greater stress and potential injury risks to the horse, potentially reducing the horse's chance of winning. Consequently, ceteris paribus, trainers are likely to minimize the distance travelled (in meters) by horse i to race j at venue v (DT_{ijv}). Ceteris paribus, horses sent further will be considered by the trainer more likely to win.

Consequently, DT_{ijv} provides the basis for testing whether expert decision-makers incorporate distance information. In addition, the distance that a trainer sent horse i in its last race, j-1, $(DT_{\mathrm{i(j-1))v}})$ may be instructive of its performance in the current race since this provides some measure of the trainer's view of the current well-being/form of the horse. Also horses that travel further on average across all their preceding races have a higher chance of winning today's race, since this provides evidence that the trainer has thought highly enough of the horse throughout its career to incur the costs/risks associated with longer distance travel. Therefore, the average distance travelled for horse i ($DTavg_i$)

It is important to take into account the prize money offered in races since this is the primary source of financial reward for most horse owners. The total distance that trainer, *T,* has sent his horses (i =*1*...n_t where n_t is the total number of horses trained by T) in past races is divided by the prize money earned, *PM,* to examine the relative effects of distance and prize money. Those trainers who make effective distance/reward trade-offs and whose decisions to send a horse a longer distance might be more indicative of their greater chances of success, are likely to be associated with lower values of this ratio. These variables are defined in Table 2.

Table 3.2 Distance related variables: definitions and formula

Variable	Definition	Full specification	
DT	Distance in meters from trainer *T's* establishment to racetrack venue *v,* for horse *i* sent to race *j*.	$= DT_{ijv}$	(3.1)
DT^{-1}	Distance in meters that the trainer sent $horse_i$ for its previous race	$= DT_{\mathrm{i(j-1))v}}$	(3.2)
DTavg	Average distance travelled by horse i in its past races j = *1,2*...n_j, where n_{ij} is the total number of past races for horse *i*. This variable only takes into account the distance for past races for a specific horse.	$= \mathrm{E(DT_i)}$	(3.3)
DM	Ratio of total distance (meters) trainer *T* has sent all horses, i = *1*...*n* to race j = 1…n_j relative to prize money (£*PM*) won in race j = 1…n_j. Therefore, this variable is measured across all horses of a given trainer and estimates the success of their decisions with respect to a horse's prospects of winning and the prize money at stake in that race.	$= \frac{SUM(DT_{ijv})}{SUM(PM_{ij})}$	(3.4)

3.3.3 Eliciting subjective beliefs based on distance-related behavior

I employed the preference variable technique (Benter,1994) to capture how well a horse performs in a race based on how far the horse has travelled to that race. Specifically, tocalculate the relationship between the distance travelled and performance for a horse, by regressing the distance travelled by horse i (DT_{ijv}) against its normalized finishing positions (*NFP*) in all its races. The normalized finishing position of horse i in race j, NFP_{ij} is defined as:

$$NFP_{ij} = 0.5 - \frac{Ordinal\ Finishing\ Position_j - 1}{Number\ of\ runners_{ij} - 1} \quad (3.5)$$

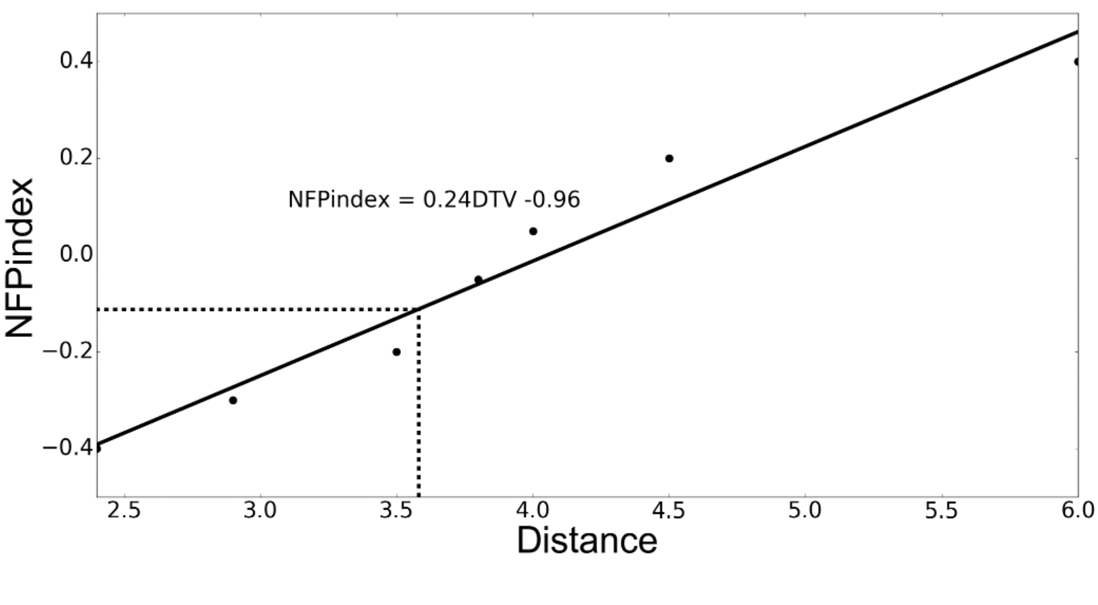
The *NFP* takes a value over the interval [-0.5, 0.5] where the first and last placed horses take values of 0.5 and -0.5 respectively (Brecher 1980). The *NFP* converts the finishing position onto a continuous scale that calculates how well the horse performed relative to the number of competitors in the race. Linear regression is conducted for each trainer to explore the relationship between the distance he/she sent his or her horses and how well they performed. For each trainer *T,* the combined performance of all their horses is taken into account. For trainer *T's* horses, $1 \ldots n_i$ where n_i is the total number of horses for this trainer, the performance of horse i in race j (NFP_{ij}) is subtracted from the horse's mean NFP across all the horse's previous races, defined as:

$$\text{Mean NFP} = \text{SUM}\left(\frac{NFP_{i,j-1}}{n_{i,j-1}}\right) \quad (3.6)$$

where $n_{i,j-1}$ is the number of races run by horse i up to but not including race j. This provides a measure of horse *i's* relative performance in race *j* compared to its average performance. The relative performance of each horse is regressed against the distance travelled to race j, DT_{ijv}. The resulting regression line $NFPindex_{ij}$ is then used to predict how well a trainer's horse will perform, taking into account to what extent the trainer has taken account of travelling distance when sending his horses to race in the past.

If a given horse performs particularly well when it is sent a long distance then the regression line will have a relatively steep positive slope. Similarly, a positive correlation between the combined performances of all the horses of a particular trainer and their NFP suggests that the further this trainer sends his/her horses, ceteris paribus, the greater will be their chances of success. Such a trainer is referred to as having a positive distance travelled preference. Figure 1 shows an example assuming a linear relationship between the distance travelled and a horse's $NFPindex_{ij}$. In this case, the further the horse is sent by the trainer the greater is the horse's chance of performing well relative to its average performance; suggesting that the trainer has the ability to make effective distance/performance trade-off decisions. In the example shown in figure 1, if the distance travelled to a race is 3.58, then the horse's $NFPindex_{ij}$ = -0.11 (i.e. it is likely to perform slightly worse than it does on average).

Figure 3.1 Example of eliciting trainer's subjective beliefs related to the distance for a given trainer



3.3.4 Method

The conditional logit (CL) model (McFadden 1974) is employed to test the hypotheses. The CL model has been widely used in discrete choice modelling, including location choice (McFadden 1978) and in various betting market studies (Benter 1994; Bolton and Chapman 1986). The CL model can be used to compute the probability of a horse winning a race based on a range of variables which might influence its prospects of success, whilst taking into account competition between horses in that race. Specifically, the CL model is used to predict the probability of horse i winning race j, P_{ij}^{e} based on a winningness index, W_{i}^{j}, defined as:

$$W_{ij} = \sum_{l=1}^{m} \beta(l) x_{ij}(l) + \varepsilon_{ij} \tag{3.7}$$

where $\beta(l)$ (for $l = 1,2...m$) is a vector of coefficients which measure the relative importance of the input variables $x_{ij}(l)$ and ε_{ij} is an independent error term. Assuming that the error term follows the double exponential distribution (which has been shown to be a sensible assumption for horseraces: Benter, 1994), the probability of horse i winning race j is given by:

$$p_{ij}^{e} = \frac{\exp(\sum_{l=1}^{m} \beta(l) x_{ij}(l))}{\sum_{i=1}^{n_j} \exp(\sum_{l=1}^{m} \beta(l) x_{ij}(l))} \tag{3.8}$$

where n_j is the number of runners in race j. The vector of coefficients $\beta(l)$ for input variables $x_{ij}(l)$ are estimated by maximizing the joint probability of observing the results in all races in the dataset using maximum likelihood procedures. The variables related to the distance a horse travels to the racetrack (discussed above) are used in the CL model to determine the probability of a horse winning a given race.

To test whether these variables add predictive power over a random model where each horse has an equal chance of winning, the standard normal test statistic $(l) = \frac{\beta(l)}{S.E\,[\beta(l)]}$ is used.

The greater the value, the greater the importance of the variable in predicting winning probability. The predictive accuracy of the model is given by McFadden's pseudo-R^2:

$$R^2 = 1 - \frac{ln\, L(M_{Full})}{\ln L\ (M_{Intercept})} \quad (3.9)$$

where, M_{Full} is the model containing all the predictors, $M_{Intercept}$ is the random choice model where each horse has equal probability of success and L is the estimated Likelihood. The maximum log-likelihood (*Ln*) of the full model is given by

$$Ln \text{ L (Full)} = \sum_{j=1}^{N} \sum_{i=1}^{n_j} y_{ij} \ln p_{ij} \quad (3.10)$$

where y_{ij} = *1* if horse *i* won race *j* and 0 otherwise, *N* is the total number of races in the dataset. The maximum likelihood procedure estimates the parameters of a statistical model by maximizing the probability of observing the particular results of the races in the dataset.

3.3.5 Testing hypotheses

CL model (8) tests H1: Experts make informed distance-based trade-off choices that can only be fully discerned by processing their distance-related trade-offsto understand their preferences and elicited beliefs. This model contains the independent variables, i.e. *DT, DT^{-1}, DTavg*, *DM* and *NFPindex* (referred to hereafter as the 'Distance model'), and the standard normal test statistic (l) shows whetherthe coefficients is significantly different from zero. The test statistic (*l*) shows whether the distance a trainer sends horse *i* for race *j* is useful for predicting horse *i*'s objective probability of winning race *j*. In particular, each of the coefficients *β(l)* shows whether the corresponding independent variables are significantly different from zero. The variables related to distance travelled (*DT, DT^{-1}, DTavg* and *NFPindex)* are arranged in ascending order of complexity. If the distance related decisions of trainers could be fully discerned from only simplistic data analysis then only *β(DT)* to be significant. However, if, trainers' preferences and beliefs can only be fully discerned by sophisticated data analysis, then the coefficients of the remaining variables are also likely to be significant. If trainers do make informed distance-based decisions then the results should show that they are prepared to send their horses further to race when they think their chances of success are greater. If this is the case, then $\beta(DT)$, $\beta(DT^{-1})$, $\beta(DTavg)$ and $\beta(NFPindex)$ will all be significantly different from zero and positive. If trainers make good trade-offs between distance travelled and prize money likely to be earned then DM will have a lower value. a negative$\beta(DM)$ value suggests that horses associated with lower values for DM will have a greater chance of success.

To test H2: Privately held information related to expert's distance-based trade-off choices provides improved forecasting accuracy compared to the crowd's best estimates. Analysis explores whether the betting public recognise that racehorse trainers often send their horses further to take part in races if they believe the horses have a greater chance of success. CL model (8) is estimated with the following independent variables: DT, DT^{-1}, $DTavg$, DM, $NFPindex$ *and* log(*Odds*) (referred to hereafter as the 'Odds & Distance model'). The odds for horse i in race j provide the public's estimate of the horse's probability of winning. Consequently, by including odds in the model to test which of these travelling distance-related variables the crowd discounts. If the public fully discount the travelling distance-related variables then the coefficients of DT, DT^{-1}, $DTavg$, DM and $NFPindex$, should not be significantly different from zero.

Whether the crowd fail to incorporate all valuable information that can be discerned from the distance variables and the behavior of experts is tested by conducting Log-Likliehood tests. The Log-Likelihood Ratio (LLR), allows us to test if the 'Odds & Distance model' model better fits the winning probability data than the 'Odds' model, a model with just log(*Odds*), and is calculated as follows:

$$LLR = -2\ln\left(\frac{log-likelihood\ for\ Distance\ \&\ Predicted\ NFP\ model}{log-likelihood\ for\ Odds\ model)}\right) \quad (3.10)$$

This provides a test statistic which is approximately $\chi 2$ distributed with degrees of freedom equal to the difference in number of parameters. This provides a statistical test of whether the crowd do incorporate all the valuable information that can be discerned from the behavior of experts and distance or otherwise.

3.3.6 Sample windows

To test the hypotheses and how distance information is incorporated over time, a rolling test sample window is adopted; a similar methodology to Lim and Brooks (2011). The first three-years of races in the dataset, i.e. 1999-2001 (inclusive) are used as a training sample to estimate the parameters for the CL models and then test the hypotheses. Then the training sample is more forward one year and the next three years, i.e. 2000-2002 (inclusive) are used to test the hypotheses. This process is continued for the whole dataset from 1999-2010. Whereas a static dataset would allow for one-time dependant observation, the rolling test windows allow us to explore how the use of travelling distance-related information changes over time. This allows us to examine how the experts (trainers) and the public learn to use distance information through time.

It proves difficult to reliably recognise and isolate a set of factors relating to distance because of the turbulence, diversity and complexity of horse racing. This difficulty arises both from isolating distance as a predictive factor of the chances of a horse winning a race and also, isolating distance within the myriad of factors that underpins the aggregate decision-making of betting markets. First, there are a wide range of factors that can be used in making accurate winning probability estimates for each horse that among other factors include lengths beaten in previous races, preferences for track and going, jockey, draw bias and weigh carried (Sung et al. 2016). Rather than including such factors, the analysis seeks to explore how distance can be used as a feature that could, in further research, be included in such a comprehensive model to predict accurate winning probabilities. The focus here is on identifying the predictive value of distance information. Second, the betting markets themselves are influence by a different set of factors that influence the overall accuracy of the crowd's best estimates which includes the overround, race-day, geographic location and last race effects in different markets (Bruce et al. 2014). Further, the composition of the betting markets itself varies and the concentration of experts and novices depends on factors such as the class of race and time of race (weekend) (Sung et al. 2012). Again, the purpose of this paper is not to include all the relevant variables but rather test the added benefit of using distance information from the perspective of the crowd to test the extent to which this information provides superior information.

3.4 Model results

3.4.1 Experts and distance information

First travelling distance-related information is significant for predicting winning probabilities as shown by the 'Distance model' for each of the three-year rolling sample periods. The results are reported in Table 3.

Table 3.3 Distance model estimated for three-year sample periods ending in years 2001-2010

Three year period ending:	Variables in Distance model (Eq.8)									
	DT		DT^{-1}		$DTavg$		DM		$NFPindex$	
	β	P>|z|	β	P>|z|	β	P>|z|	β	P>|z|	β	P>|z|
2001	0.0010	**0.0000****	0.0007	**0.0006****	0.0016	**0.0000****	-0.8823	**0.0000****	2.965	**0.0000****
2002	0.0013	**0.0000****	0.0008	**0.0001****	0.0016	**0.0000****	-0.8081	**0.0000****	3.055	**0.0000****
2003	0.0013	**0.0000****	0.0007	**0.0002****	0.0019	**0.0000****	-0.7877	**0.0000****	3.279	**0.0000****
2004	0.0011	**0.0000****	0.0006	**0.0015****	0.0025	**0.0000****	-0.7633	**0.0000****	3.595	**0.0000****
2005	0.0010	**0.0000****	0.0003	**0.0432***	0.0028	**0.0000****	-0.8156	**0.0000****	3.157	**0.0000****
2006	0.0006	**0.0001****	0.0003	**0.0225***	0.0025	**0.0000****	-0.7763	**0.0000****	3.384	**0.0000****
2007	0.0007	**0.0000****	0.0003	**0.0465***	0.0025	**0.0000****	-0.8946	**0.0000****	3.082	**0.0000****
2008	0.0004	**0.0092****	0.0003	**0.0315***	0.0020	**0.0000****	-0.8034	**0.0000****	3.159	**0.0000****
2009	0.0005	**0.0010****	0.0004	**0.0328***	0.0021	**0.0000****	-0.7271	**0.0000****	3.433	**0.0000****
2010	0.0008	**0.0000****	0.0004	**0.0128***	0.0019	**0.0000****	-0.9649	**0.0000****	3.031	**0.0000****

**,* Significant at 1% and 5% levels in a 2- tailed test respectively.

The coefficients of each of the travelling distance-related variables *(DT, DT^{-1}, DTavg*s and *DM*) are significantly different from 0 for all of the three year sample periods (all p< 0.0000). These coefficients are positive for *DT,* (mean = 0.0008), DT^{-1}(mean = 0.0005) and *DTavg*s (mean = 0.0021) and negative for *DM* (mean= -0.8223). These results indicate, respectively, that the further a horse travelled to its current race, the further a horse was sent to race in its previous outing, and the greater the average distance travelled by a horse across all its previous races, the greater was its probability of winning its current race. Equally, the negative coefficient for *DM* suggest that horses trained by individuals who make effective distance/reward trade-offs have a higher probability of winning when travelling longer distances.

The coefficient for *NFPindex* is similarly positive and significant for all the three year samples ($P < 0.0000$), with mean = 3.214 showing that the elicited beliefs of experts is significantly predictive of a horse's chances of winning. The fact that the β values is higher for *NFPindex* shows that this variable has a stronger impact of the chances of winning compared to the other distance-related variables, highlighting that in order to account for the full value of distance-related trade-offs, it is necessary to account for the elicited beliefs of experts.

Since all the variables in the Distance model are significant, experts make informed distance-based trade-off choices. In addition, the fact that even the more complex variables (e.g., DTavgs and for NFPindex) are significant, suggests that to fully discern trainers' preferences and elicited beliefs regarding their distance-related trade-offs requires sophisticated analysis. Taken together, these results confirm H1.

3.4.2 The crowd and distance information

The results of estimating the 'Odds & Distance' model for each of the three year rolling sample periods is shown in Table 4.

Table 3.4 Odds & Distance model estimated for three-year sample periods ending in years 2001-2010

Three year period ending	Variables in Odds & Distance model											
	DT		***DT⁻¹***		***DTavg***		***DM***		***NFPindex***		**Odds**	
	β	**P>|z|**	β	**P>|z|**	β	**P>|z|**	β	**P>|z|**	β	**P>|z|**	β	**P>|z|**
2001	0.0001	0.7609	0.0004	0.0641	-0.0001	0.9309	-0.083	**0.0352***	0.1974	0.1685	1.1935	**0.0000****
2002	0.0002	0.2004	0.0004	**0.0304***	-0.0001	0.9042	-0.0972	**0.0073****	0.2193	0.0976	1.2072	**0.0000****
2003	0.0003	0.0988	0.0004	**0.0215***	-0.0001	0.8027	-0.0283	0.4192	0.2794	**0.0414***	1.2101	**0.0000****
2004	0.0001	0.5853	0.0004	**0.0359***	0.0004	0.1661	-0.0416	0.2085	0.5903	**0.0000****	1.1790	**0.0000****
2005	-0.0001	0.7046	0.0001	0.4301	0.0005	0.0729	-0.1083	**0.0043****	0.4073	**0.0011****	1.1574	**0.0000****
2006	-0.0001	0.3666	0.0001	0.5781	0.0001	0.6615	-0.0707	0.0500	0.3649	**0.0025****	1.1422	**0.0000****
2007	-0.0003	0.0614	0.0001	0.4653	-0.0001	0.8836	-0.0641	0.0827	0.2461	0.0548	1.1396	**0.0000****
2008	-0.0003	0.0555	0.0002	0.3400	-0.0003	0.2170	0.0161	0.6460	0.3295	**0.0076****	1.1425	**0.0000****
2009	-0.0001	0.4998	0.0002	0.3245	-0.0001	0.6157	-0.0236	0.4863	0.2595	**0.0369***	1.1384	**0.0000****
2010	0.0002	0.2250	0.0002	0.3363	-0.0003	0.3753	-0.0765	0.0580	0.1878	0.1252	1.1212	**0.0000****

The coefficients for the four travelling distance-related variables *(DT, DT^{-1}, DTavg*s and *DM*) are not significantly different from 0 for the majority of sample periods ending in 2001-2010 suggesting that the betting public did fully account for these travelling distance-related variables in the sample period. While some of these travelling distance-related variables are significant before 2005, from 2006 onwards the public fully account for all information related to these distance related variables, suggesting that distance based information no longer improves forecasting accuracy. Placing these results in the contextual events, the release of the free and easily accessible geographical information in Google maps, 8th February 2005, could have improved the information environment by providing the means for gathering and interpreting distance information. It is far-reaching to suggest that bettors factor in Google maps info into their decisions, the results in Table 4 show that distance information could be used to improve forecasting accuracy up until 2005, after which this information no longer provided added value.

By contrast, the coefficient for relatively complex $NFPindex$ was significant and positive for the majority of the sample periods (in seven of the ten sample periods), suggesting that the elicited beliefs of experts and more complex information that requires careful consideration provides improve forecasting accuracy. Clearly, these trade-offs are more complex to discern than the distance-related information contained in the other four variables. Taken together, the results support H2; privately held information related to expert's distance-based trade-off choices provides improved forecasting accuracy compared to the crowd's best estimates. However, the results suggest that the more complex trade-offs which could help discern some of the private information held by experts, are less readily learned by the public.

Finally, results of conducting Likelihood tests compares the 'Odds & Distance' model and a CL model incorporating only odds ('CL (odds')) in Table 5. These tests are designed examine whether the Odds & Distance model better predicts winning probabilities than the CL (odds) model.

Table 3.5 Odds & Distance (‘O & D’) and CL(*odds*) models estimated for three year sample periods ending in years 2001-2010

Three	Model R^2		Model Log Likelihood		LLR Test	
year	**CL*(odds)***	**O&D**	**CL*(odds)***	**O&D**	Value	**P>|z|**
2001	0.1657	0.1659	-23008.16	-23001.572	13.18	**0.0104***
2002	0.1728	0.1731	-24062.578	-24051.47	22.22	**0.0002****
2003	0.1723	0.1726	-25162.893	-25154.787	16.21	**0.0027***
2004	0.1772	0.1777	-26065.946	-26052.536	26.82	**0.0000***
2005	0.1724	0.1728	-27803.215	-27791.427	23.58	**0.0001***
2006	0.1675	0.1676	-29492.581	-29488.401	8.36	0.0792
2007	0.1610	0.1611	-31033.159	-31028.583	9.15	0.0574
2008	0.1618	0.1619	-31247.179	-31245.107	4.14	0.3870
2009	0.1640	0.1640	-31115.784	-31114.596	2.38	0.6669
2010	0.1621	0.1623	-31489.116	-31484.652	8.93	0.0629

**,* Significant at 1% and 5% levels in a 2- tailed test respectively

The Model R^2 values show that the Odds & Distance model better predicts winning probabilities of horses for the majority of the sample datasets. The LLR test results show that between 2001 and 2005 inclusive, the Odds & Distance model provides statistically superior information compared to odds alone (2001: P = 0.0104, 2002: P = 0.0002, 2003: P = 0.0027, 2004: P = 0.0000, 2005: P = 0.0001). Consequently, between 2001 and 2005 the public fail to fully account for all information related to distance. However, post 2005 the Odds & Distance model does not provide any statistically superior information. Therefore, these results confirm H2, privately held information related to expert’s distance-based trade-off choices provides improved forecasting accuracy compared to the crowd’s best estimates. The added benefit of distance information actually deteriorates over time, as complex transformations of data designed to fully capture the experts’ private information and elicited beliefs provide no improvements in forecasting accuracy. By comparing how distance information provides information beyond odds for forecasting accuracy, the results show that distance related information and expert elicited beliefs provide improved forecasting accuracy for a limited amount of time. It could be speculated that the limited forecasting accuracy is related to the crowd's ability to learn but this paper does not have privileged information on how the crowd wagered and the reasons for doing so, only the improved accuracy that results from using distance information.

3.5. Discussion

The results provide strong support for the use of distance information in the horse racing context to understand the behavior of experts and show how distance information could have been used to improve crowd based forecasts. In particular, experts do make informed distance-related choices but the crowd learn to incorporate a surprisingly large proportion of the experts privately held information.

This is not the first paper to refer to betting markets as crowds. Suroweicki (2005) explains that betting and financial markets existed long before the internet and the idea 'wisdom of crowds', however the ubiquity of internet access and information technology has made collective wisdom more apparent. The mechanism of financial markets, whereby the diverse opinion of independent individuals are aggregated to produce market estimates (prices) from the perceived probabilities reflect collective wisdom in action in betting markets. Indeed, the betting crowd effectively aggregate information and the sports betting market provides a 'ready-made laboratory' to explore information efficiency and the wisdom of crowds (Surowiecki 2005, p. 14). Furthermore, 'the crowd is composed of informed and un-informed members just like a betting market' showing that the crowd contains similar sub-groups of bettors to real financial markets (Leary 2017, p. 717). Exploring the betting crowd as a wise crowd can therefore help draw theory on how experts and novices use information that may be typical of informed and uninformed participants in financial markets.

This study examines and compares expert and crowd decision-making related to distance. A growing need to understand decision-making of both experts and 'the crowd' in the real world has emerged in decision-making because ultimately, a better understanding improves efforts to solve complex real world decision problems (Becker 2015; White 2016). The methodology herein to elicit experts' subjective beliefs based on their distance-related choices demonstrates how expert knowledge elicitation can be undertaken in a real world empirical setting, extending results which have largely been developed in laboratory settings (Cerroni et al. 2012). Results show that the crowd learn, through time, to discount information derived from the behavior of experts to improve their own decision-making. The results, therefore, provide an insight into one of the ways in which the 'wisdom of the crowd' develops.

Consequently, the crowd are good at discerning and utilizing distance-based information. However, recent research has shown that the crowd are less good at incorporating time based factors and fail to fully account for duration information in speculative markets (Ma et al. 2016). This suggests that the crowd are better at discerning and using certain types of information and further research is needed to examine the reasons for this and to distinguish the features of information which crowds may be better able to employ.

These results demonstrate to some extent that the wisdom of the crowd is, at least in part, discerned from the behavior of experts and that crowds are good at learning to discount elicited expert information. Coussement et al. (2015), Chen & Zeng (2016) and O'Keefe (2016) observed that more work was needed to examine the impact of integrating expert opinions into the decision-making process and, understanding how experts and the wider public differ in their use of information. Since expert advice is frequently sought after to help make decisions in uncertain environments, the ability to discern forecast relevant information from experts is an important process. New information is released on the internet and not always in a ready-to-use format meaning that experts are increasingly needed to guide novices on how best to use data to obtain useful insight. This study, for example, uses the Google Maps Distance Matrix API, a service that allows individuals to programmatically request and received the desired information. The level of expertise to programmatically access the API is high so that the barriers to learning and using this information may be sufficient to prevent the layperson. However, new tools such as Google Maps provide a simply to use interface enabling anyone to gather and implement distance based information in their decision-making. In this way, experts have removed the technical difficulties to obtaining the geographic information and provided the public with an easily accessible tool to aid their decision-making.

A recent study suggested that it is important to undertake research to understand the extent to which elicitation of experts' views can be undertaken in real world systems (Werner et al. 2017). In such natural experiments it is fundamental to understanding expert elicitation using active modellers (O'Keefe 2016). However, the majority of studies in this area have employed student subjects. Decisions made by horse race trainers exemplify expert behavior in the real world and provide valuable insight for improving probability forecasts. Experts introduce specialized knowledge that is not captured by statistical models and it has been suggested that understanding expert knowledge elicitation is important since it can lead to improved forecasts (Alvarado-Valencia et al. 2017). This has led to the call for more research on the "design of forecasting systems that combine statistical forecasts and the judgement of experts" (Leitner and Leopold-Wildburger 2011, p. 466). Such models that integrate expert and statistical forecasts are likely to provide higher accuracy forecasts and shed light on how experts can provide superior information not captured by traditional modelling approaches.

Previous research in this area have faced two important difficulties: First, the specific criteria of what constitutes an expert are ill-defined (Alvarado-Valencia et al. 2017). As a result, previous studies have often attempted to identify experts in specific domains, hoping that they are experts. Indeed, identifying experts is a non-trivial task. Studies have identified experts using a wide range of criteria, including experience, performance and even self-identified expertise. This inability to determine the quality of the expertise of subjects has been questioned by Van Wesep, (2016) and could be the reason for the contradictory results found in the literature on expert performance. Second, previous research has typically involved "laboratory experiments and do not include actual forecasters" or experts (Legerstee and Franses 2014). As Katsikopoulos (2013) observed, biases and heuristics that affect real world decision-making may not be evident in a controlled laboratory setting. Consequently, the applicability of the findings from the laboratory to real world settings may be questioned. To overcome such limitations this paper explores decisions made in a real world environment by professional experts whose reputation and livelihood depends, to some extent, on making correct distance-based decisions under uncertainty.

Information derived from the behavior of experts can be combined with statistical models to produce improved subjective probability estimates. The finding that the public account for the behavior of experts when making their judgements supports the results of laboratory experiments which have shown that both experts and novices are aware of each other's abilities and, novices even think that experts will perform better (McKenzie et al. 2008; Torngren and Montgomery 2004). The crowd appears to observe expert behavior and is able to discern from this valuable information related to the experts' own beliefs that they use to improve their own judgements. The crowd is able to discount the beliefs and private information held by experts. The results contrast with the finding that decision-makers in some professional contexts ignore the advice of experts, leading to inaccurate long-term judgements and poor decisions (Ecken & Pibernik, 2016). The contrasting results could arise from the fact that Ecken & Pibernik (2016) explore forecasting where there is no incentive offered for accurate estimates, whereas this study deals with real economic gain and loss. In the betting market context, the economic incentive for accurate forecasts may increase the desire to make better-informed decisions and as such, depend on expertise rather than personal judgement.

Properly incorporating private information from those with an information advantage will produce significantly more accurate forecasts (Lessmann and Voß 2017). As this paper shows, the superior information is not necessarily private, rather innovative techniques can be employed using online information to contextualise the observed behavior of experts.

It is important to observe the actions of experts since they tend to make better probability judgements (cf. naïve respondents) when there is constant and clear feedback on their judgements (Andersson and Nilsson 2015). Another study examining betting markets used an elicitation technique to leverage the beliefs of the crowd (Chen & Zeng, 2016). Two basic strategies have emerged for expert elicitation, the first is sampling from a large number of less skilled individuals and gathering expertise from the 'wisdom of crowds' and the second is sampling from a smaller number of highly skilled individuals (Bolger and Wright 2017). Expertise exists in a number of different formats and various methods can be applied to recognise how different sub-groups can exhibit expertise. This paper has highlighted both strategies: the small number of trainers represent highlight skilled individuals and the horse race betting market are similar to the 'wisdom of crowds'. It is important to recognise the structural properties inherent to both groups and the varying objectives of each group, because different knowledge elicitation methods are needed depending on the sub-group population and their structure.

Betting markets provide an ideal environment to test how well information is employed, as it incorporates both expert's decisions (e.g. the distance trainers decide to send their horses to compete) and decisions by the crowd (e.g., betting decisions which lead to odds formation). Importantly, at a point in time (once the race is finished) all uncertainty is resolved and the result is announced, providing the means of assessing the quality of decisions made by the market. Betting markets thus provide an ideal real-word environment to test the hypotheses related to distance-related decision-making and belief elicitation.

A key aspect of real world experiments which may not be found in the laboratory is the fact that these systems are dynamic (Hamalainen et al. 2013). Economic feedback in these environments affects expert behavior and leads to more accurate forecasts (Petropoulos et al. 2016). In the study, trainer income depends on them making accurate judgements and correct decisions when dealing with uncertainty. Horserace trainers therefore, provide very good examples of experts as their livelihood depends on success in dealing with uncertainty on a day-to-day basis. At the same time, the dynamic nature of betting markets and the horseracing makes it both an opportunity and a challenge: new information can be used in event studies to understand how the relationship between markets and the information environment evolves, but also isolating specific information proves a challenge considering the multitude of dynamically changing factors.

Clearly, in laboratory settings, it is important to explore how the behavior of experts will change if they know they are being observed (Hamalainen et al. 2013). Indeed, finding non-intrusive ways to elicit information from experts is a challenge (Werner et al. 2017). By observing experts' actions in the real world, research design is able to find consistency between beliefs and actions while bypassing the fear that the process of belief elicitation itself changes the behavior of subjects (Schotter and Trevino 2014). Indeed, when respondents know that they are the subject of laboratory experiments this can lead to certain biases that can blur the findings. Recognising the human elements of the decision-making process and individuals' heuristics that are prescriptive of behavior will lead to more accurate decision-making systems in forecasting.

Expert knowledge elicitation, the process of eliciting and interpreting estimates of uncertain quantities from a single expert or group of experts has emerged in the forecasting literature as a key technique to develop theory on judgemental forecasting (Bolger and Wright 2017). Compared to judgemental forecasting, expert knowledge elicitation deals with actual experts and involves the whole process of expert selection, screening and aggregation of expertise to determine statistical forecasts and the evaluation of such forecasts (Bolger and Wright 2017). Research focusing on incorporating expert judgement in forecasting has progressed over the years and there are some marked differences between judgemental forecasting and knowledge elicitation. To oversimplify, judgemental forecasting is the process of incorporating the expertise and judgement of a single individual to forecast an uncertain event outcome where there is little or no available data (Lawrence et al. 2006).

On the other hand, expert knowledge elicitation differs in that is it characterized by empirical (rather than theoretical) methods to extract information from a group of experts (rather than an individual) (Bolger and Wright 2017). Ultimately, 'expert knowledge elicitation lies at the core of judgemental forecasting' and both are integral methods to extract expertise for forecasting, but expert knowledge elicitation marks a more granular approach of extracting exactly the proponent of knowledge that makes experts worthy decision-makers (Alvarado-Valencia et al. 2017, p. 298). Moreover, the shift towards more empirical analytical methods to extract expertise from a group of experts (rather than individuals) is better suited to the data rich environment and especially with the popularity of betting markets, which represent a ready-made laboratory to study aggregated predictions and their outcome.

Understanding how distance can improve forecasting accuracy is important in the wider context of the web where individuals frequently use mapping services and must evaluate multidimensional information. Spatial distance is a key factor in many decisions and the results show that experts travel further when there is a greater likelihood of reward, confirming previous findings in information search and internet commerce literature, that physical proximity is influences actual behavior (Ghose et al. 2013). In mobile advertisement targeting it has been shown that physical proximity is associated with consumer engagement and purchasing (Cachon 2014). The significant set of distance factors in this study confirm the important role that distance plays in decision-making; highlighting the fact that individuals are more likely to travel further if the probability of success (however this may be defined) is greater (cf. Zipf 1949).

3.6. Conclusion

Decision-makers have access to an increasing quantity and variety of decision tools that is changing how they ultimately behave (Habjan et al. 2014). This study highlights how information from online decision-making tools, such as Google maps, enhances understanding of how distance information can be used to understand expert and crowd based decision-making. The methodology shows how distance information can support elicitation techniques from experts and provide contextual evidence to their behavior. Furthermore, harnessing distance information and the elicited knowledge of experts can improve the forecasting accuracy of the betting crowd for a short amount of time, before the crowd produce aggregate estimates that account for any information advantage from distance information.

Previous research has largely compared the performance accuracy in uncertain decision-making of both experts and novices in laboratory settings and has often used students as supposed experts. By contrast, this study analyses the decisions of acknowledged experts (whose livelihood depends on successful decision-making under uncertainty on a daily basis) and those without direct access to the private information held by those experts in a real world environment. Further research on expert knowledge elicitation that elicits information from real experts with non-obtrusive methods will help complement the laboratory-based methods that suffer from lack of expertise, responses that conflict with actual behavior and, the very means of observation itself that influences the behavior of respondents.

Chapter 4 The long and short of how online sentiment signals impacts financial trading behavior and performance

Abstract

This paper shows how sentiment in the news affects individual trader behavior and performance. Combining news sentiment data from Thomson Reuters News Analytics on the FTSE 100 in the period 1st April to September 30th 2012 with data related to 190,363 FTSE 100 trades placed by 3583 individual traders during that period shows how individual traders respond to sentiment information. Results show news sentiment significantly affects the volume of trading and that individuals behave in a sentiment-contrarian fashion, buying the market around negative sentiment and selling the market around positive sentiment. When trading long, the returns do not appear to be impacted whether this follows the release of positive or negative market sentiment. However, trading short following negative market sentiment is a profitable strategy whilst trading short following positive market sentiment has a negative effect on returns. This is the first paper to highlight the effect of sentiment on long (buy) and short (sell) individual behavior and performance at intra-day level and the implications of the sentiment contrarian behavior for regulators of financial markets.

Keywords: Individual Trading, buying (longing), selling (shorting), Sentiment, Sentiment Contrarian behavior, Returns, Risk

The long and short of how sentiment signals in online news information impacts financial trading behavior and performance

4.1. Introduction

Online news plays an important role in informing financial markets. Markets react to public news conveying, for example, company-specific, macroeconomic and political information (Groß-Klußmann and Hautsch 2011). Based on the efficient market hypothesis, much of the current literature assumes that market prices respond directly to news information. However, to fully understand the relationship between online news and financial markets it is necessary to recognise that it is human reaction to news information that drives price changes (Kauter et al. 2015). There is recent evidence that public news information is interpreted by individuals and that their subsequent reactions lead to changes in market prices (Peress 2014). Indeed, Peress (2014) shows that newspaper strikes affect individual investor behavior. In particular, on days when newspapers go on strike, the market experiences lower turnover as well as lower market volatility and returns.

In summary, a number of studies explore how news affects market prices, but little research has examined how online news affects individuals' trading behavior (Lillo et al. 2015). By exploring the impact of online news information on individual trading behavior and performance, this paper offers new insights into the behavioral factors that result in online news affecting financial markets (Kauter et al. 2015). Understanding the effects of news at the individual level is particularly important because it is the process by which news diffuses through investors and their consequent trading that affects price formation (Peress 2014). Previous research has shown that the amount of news affects market prices but it is necessary to decompose the effect of news information by examining the content of the news separately from the volume (Engelberg and Parsons 2011). That is, rather than simply showing how the volume of news is associated with trading and lack of information limits trading, this paper highlights how sentiment affects individuals' trading activity.

The sentiment related to a particular piece of news (i.e. whether information is perceived as positive or negative) is likely to impact traders' behavior. Although interpretation is highly subjective, recent studies have shown that the TRNA sentiment is a reliable measure for sentiment (Smales 2014; Yang et al. 2017). In fact, the literature has suggested that negative (cf. positive) information is more salient. For example, on an investment-related social media message board, negative words were shown to negatively predict stock returns (Chen et al. 2014). Similarly investors have been shown to react more strongly to negative than positive information contained in analysts' reports (Huang et al. 2014) and in the gold futures market there is a stronger response to negative (cf. positive) news (Smales 2014). Equally, stock prices for Standard & Poor 500 firms have been shown to respond to negative words in news information (Tetlock et al. 2008) and the degree of abnormal trading has been shown to be greater after extreme negative earnings surprises (Hirshleifer et al. 2008). Despite these findings concerning aggregate market level effects, there has been little attempt to explain *why* investors are affected more by negative information and specifically, *how* sentiment affects individuals' trading behavior.

One possible explanation for the observed impact of negative information is 'negativity bias'. In particular, negative information has a stronger impact than equivalent positive information on individuals because individuals give greater weight to potential costs than gains when making decisions under risk. Consequently, negative information is more heavily weighted in the formation of opinions and behavior (Peeters and Czapinski 1990). In addition, negative information has been shown to be more readily *objectively* identified and causes a 'fight or flight' response. This arises because humans readily respond to prevent any harmful situations/environments while positive information is more *subjectively* identified and leads to less of a behavioral response (Peeters and Czapinski 1990). Negative information lends itself to objective processing as negative cues are accessible through common thought processes whereas positive information is subjectively interpreted (and, therefore, interpreted differently by individuals) (Akhtar et al. 2013).

Theory has suggested that sentiment of news affects both attention and timeliness of decisions. Media coverage focuses on negative news information because it increases viewer ratings and sales (Akhtar et al. 2013). High attention is usually given to negative articles and both the attention and the negativity cause greater stock reactions than that given to positive news, which is often given less attention by the media (J. L. Zhang et al. 2016). Regarding the timeliness of decisions, positive news, has been shown to affect immediate decisions and to increase impulsive buying behavior while negative news has been shown to have more impact on long-term decisions and can increase resistance to buying (Sul et al. 2016). As a result, sentiment asymmetrically affects individual behavior.

Research to date suggests that positive and negative sentiment may well affect buying and selling decisions differently. However, this important aspect of the influence of news sentiment and its asymmetric effects has not yet been explored at the individual trading level.

The advent of the internet has provided investors with access to constantly evolving, up-to-date information and has emphasized the importance of speed and immediacy of information (Barber and Odean 2003). Consequently, the internet has transformed the information environment and is also likely to influence trading behavior (Sabherwal et al. 2011). There has been a call for studies that examine how online, real-time news information diffuses among the investor population. In particular, there has been a call to examine the intra-day effects, in order to explore the speed at which information diffuses (Peress 2014). Understanding how individuals respond to sentiment is of fundamental importance when investors have constant access to information online.

The only previous study to explore how online news sentiment impacts individual behavior found that sentiment was weakly and positively correlated with individual behavior: positive sentiment was associated with individuals buying the market and negative news was associated with them selling the market (Lillo et al. 2015). The authors conceded that their sentiment indicator was limited since they only counted the number of positive and negative words related to one stock (Nokia). Further, they explored the effect of sentiment at the inter-day level, which fails to reflect the instantaneous nature of the modern market place. Indeed, today, information diffuses rapidly and “returns should appear auto correlated over shorter horizons, for example, within the day instead of between days” (Peress (2014, p. 2041)). However, no study to date has reflected this.

The difficulty in exploring individual level trading is the lack of sufficiently detailed data (Lillo et al. (2015), Kaniel et al. (2012)). In particular, having such data would enable “news response coefficients” to be decomposed into “media effects” and “content effects” (Engelberg and Parsons 2011), thus allowing a more detailed understanding of the manner in which individual traders respond to different types of news.

By securing access to individual trade-level data from a large spread-trading broker research can overcome the issues of data availability. Individuals in spread trading markets can buy (long) or sell (short) an index (e.g., FTSE 100). Therefore, if an individual thinks the market price will increase/decrease then they can “long”/“short” the market. Combining this individual level trading data with data from the Thomson Reuters News Analytics (TRNA) service provides real-time sentiment information for online news information. By combining these datasets, analysis explores how online news sentiment affects individuals. Spread traders typically trade several times in a day and this provides the opportunity to explore the diffusion of information over short time periods.

This paper makes three important contributions: First, news sentiment is useful for explaining individual trading behavior. Compared to the majority of literature which explores how sentiment affects markets prices (Li et al. 2014), this paper focuses on the relationship between sentiment and individual behavior. In particular, the underlying order flow (total buy stakes minus sell stakes) at the individual level shows that individuals act in a sentiment-contrarian fashion, buying the market in relation to negative sentiment and selling the market in relation to positive sentiment. This finding is robust after controlling for price-change behavior which is the tendency for individuals to buy when the price decreases and sell when the price increases (Lillo et al. 2015). Consequently, the results highlight how sentiment has a contrarian effect on individual behavior.

Second, the effect of news sentiment on individuals' behavior at the intra-day level, exploring the effect of news in finer-grained periods than has previously been undertaken. Specifically, how news sentiment in the hour prior to the trading affects trading behavior. For example, sentiment in the hour period prior to a trade (e.g. 9-9.47am where 9.47am represents the time of the last trade in that hour) helps explain the variation in trading volume during that hour. In addition, analysis considers how trading performance is related to news sentiment released in a range of different times. For example, how sentiment in the prior 15 minutes is related to subsequent returns. Akhtar et al. (2013) point out that a short period between measuring market sentiment and the trading response is necessary to isolate the impact of sentiment on trading, as, *ceteris paribus,* the trading behavior can be attributable to that sentiment. Importantly, only considering the sentiment prior to a trade as a means to isolate the effect of sentiment prior to trading. This is important because looking at aggregate daily sentiment and trading behavior does not account for the fact that sentiment released after is unrelated to trading decisions unless individuals can pre-empt news announcements.[12] Based on the extant literature, this is the finest-grained treatment of sentiment at individual-level to date (Yang et al. 2017).

Third, news sentiment is useful for explaining individuals' trading performance. Most previous studies have examined how news and earnings announcements affect market returns (e.g. Frazzini (2006); Kaniel et al. (2012)) and not the returns of individual investors. The sentiment prior to an individual's trade is related to the returns and variability of the individual's returns which follow. In particular, for individuals who predict the market price will decrease (i.e. execute short trades after the release of sentiment), increased positive sentiment prior to the trade is associated with decreased returns and negative sentiment prior to the trade is associated with increased returns. For individuals who predict the market price will increase there is no significant correlation between sentiment and average returns. However, the variability of their returns is significantly impacted by sentiment: positive/negative sentiment prior to the long trading is associated with decreased/increased variability in returns. These results highlight how trading short around prevailing negative market sentiment can lead to profitable returns whilst prevailing positive sentiment has a negative effect on short trading returns.

The reminder of the paper is structured as follows: In section 2, a review of the literature on online news information and financial markets, showing the importance of understanding the effects at the individual level. In section 3, the methodology and define the hypotheses related to individuals' reactions to news sentiment. In section 4, the data used to

[12] While institutional trading has been shown to predict news announcements (Hendershott et al. 2015) it has been shown that individuals are not able to predict news announcements and it is the moment when news is released that informed traders gain any advantage (Engelberg et al. 2012).

test how individuals behave and perform in relation to online news information signals. In section 5, the variables related to sentiment by exploring the positive and negative attributes of news in specific intra-day intervals. The list of control variables related to individual and market characteristics is detailed. The statistical procedures employed to test the hypotheses are described in section 6. The results are presented in section 7 and conclusions drawn in section 8.

4.2 Research background

4.2.1 Markets and news information

News plays a fundamental role in financial markets. Indeed it has been suggested that "news [is] driving financial markets" (J. L. Zhang et al. 2016, p. 3). However, there are diverging views of how markets actually respond to news information (Kauter et al. 2015). The most dominant view is that news drives market price fluctuations. It has been shown that web information can be used to predict prices and that trading strategies using financial news articles can produce profitable returns (Li et al. 2014). Similarly, it has been shown that automated news reading systems can be used to predict stock market prices (Hagenau et al. 2013).

The majority of previous research has assumed that online news information impacts market prices directly (Allen et al. 2017). However, individuals respond to the news and it is their trading decisions that drive prices. Consequently, the role that online news information plays in affecting investor behavior is even more important, with individuals increasingly using the internet to gather information (Rubin and Rubin 2010). Indeed, the "evidence suggests that public information diffuses gradually through the investor population and that this gradual diffusion affects prices" (Peress 2014, p. 2007). Therefore, exploring how individuals behave in relation to online sentiment signals and how sentiment affects their trading behavior, will help identify the underlying mechanisms that lead to market price adjustments.

The early studies focussed on the effects of news information at the aggregate level, exploring for example, to what extent sentiment signals in online news predicts prices (Li et al. 2014). Previous research on the impact of news on individual traders has been limited to exploring the impact of news contained in newspapers in the pre-internet era. These studies found that the content of newspapers affects both individual level behavior (Engelberg and Parsons 2011) and return performance (Yuan 2015).

However, as Kearney and Liu (2014, p. 172) observe, "it is […] unclear how investors interpret textual sentiment" and how this sentiment drives their behavior. Consequently, to further understand the effect of online news information on investor behavior, it is important to examine its impact on individual investors (Zhang and Zhang 2015; Baker and Wurgler 2007). This should provide novel insight regarding the effects of sentiment signals, which previous research have taken as exogenous (Baker and Wurgler 2007).

A new stream of research is now emerging concerning how individuals respond to online information. For example, Lillo et al. (2015) investigate how different categories of traders (financial, insurance, government, non-profit and households) react to online news information. They focus on the differences between these categories, and how they behave in relation to Nokia news information. The authors find that all categories of trader react in a price-contrarian fashion, buying/selling Nokia stock when the price goes down/up. They found that companies and households are only affected by positive sentiment in online news; buying stock around positive news. However, the relationship between their sentiment score and behavior may be weak because the authors conceded that their sentiment score fails to correctly identify the sentiment of news articles. In addition, their study focussed on behavior in relation to only one stock and they examined the impact of news on returns at the intra-day level (and not variability of returns).

Clearly, a full undertstanding of the effect of news sentiment on individual investor behavior will only be forthcoming by overcoming these limitations. Importantly, given how quickly online news information disseminates (Groß-Klußmann and Hautsch (2011) it is important to examine impacts at much shorter than daily intervals.

Most research has assumed that markets respond directly to news. However, it is trading which moves markets and highlighting how individuals respond to news is key to understanding how sentiment drives markets.

4.2.2 Sentiment

Research has suggested that markets are impacted differently by the sentiment (positive or negative) of online news information. There is overwhelming evidence that negative news is more salient. For example, it has been shown that in the gold futures market, negative (cf. positive) sentiment has a greater effect on returns (Smales 2014) and negative words in social media have been found to predict stock returns (Chen et al. 2014). In addition, negative words in stock analysts' forecasts have been shown to convey information beyond historical accounting data for S&P 500 companies (Tetlock et al. 2008) and investors have been found to react more strongly to negative news in analysts' reports (Huang et al. 2014). Equally, it has been shown that the amount of abnormal trading is greater after extreme negative (cf. positive) earnings surprises (Hirshleifer et al. 2008).

Whilst the overwhelming majority of research suggests that negative information affects trader more than positive sentiment, two studies have found that positive news is more salient. Good news in twitter messages related to S&P 500 companies has been shown to be better (cf. negative) for estimating returns (Sprenger et al. 2014) and positive sentiment has been shown to be positively correlated with index returns in 11 international markets (Zhang et al. (2016)).

Negative news appears to affect markets more than positive news but further research is needed to understand to what extent negative news is more salient at the individual-level and how negative sentiment affects their behavior differently to positive news.

4.2.3 Speed of diffusion

Previous research examining the impact of information from newspapers on financial markets is of less relevance in today's markets driven by electronic information flows (Engelberg and Parsons 2011). Newspapers are released on a daily basis, but traders now have access to real-time information via the internet. This has changed investors' focus to speed and immediacy (Barber and Odean 2003). This culture of speed among traders necessitates the use of the most current information. Consequently, there have been calls for studies to examine the impact of sentiment in real time, to understand the behavioral processes associated with traders who are more reliant on internet-based news and to examine how trading volume is impacted by contemporaneous news (Peress 2014).

Clearly, markets and news are intrinsically linked. However, there is considerable debate as to whether markets lead news or vice versa (Nardo et al. 2015). Some studies have suggested that institutions can predict news announcements by observing trading patterns (Hendershott et al. 2015), but other studies argue that it is far more likely that trading follows the release of news (Engelberg et al. 2012). There is clearly a need to shed further light on the relationship between news and trading activity and the study sets out to achieve this.

The efficient market hypothesis would suggest that online news information would be incorporated into markets almost immediately. However, research at intra-day level, rather than inter-day level, is needed to understand the actual rate at which news information diffuses amongst individuals. Although high frequency trading analysis explores automated trading at the millisecond level, it is unlikely that humans can interpret news articles at such speed and exploring how humans respond at the intra-day level provides the level of detail needed to understand how individuals respond to sentiment. Since it takes humans a number of minutes to read an article, minute-by-minute level analysis would be unreflective of the rate at which traders interpret information. News information and sentiment are constantly released on the internet and individuals are free to trade on this information in real time. Few studies have explored the effect of sentiment on security prices in the short term (Akhtar et al. 2013).

As Akhtar et al. (2013) point out, examining short periods between sentiment release and individual response, *ceteris paribus*, means that any market movements around the moment sentiment is released can be attributable to that sentiment. Therefore, exploring the speed of diffusion at the intra-day level minimizes the effects of endogenous information.

4.3 Methodology

4.3.1 Hypotheses

In order to understand how sentiment in online information affects trading behavior and performance it is necessary explore to what extent, and in what manner, positive and negative information affects both long and short trades (J. L. Zhang et al. 2016). Positive and negative sentiment may be viewed differently depending on the situation of the individual. When an individual makes a long/short trade, positive sentiment may be perceived as precursor or indication that the price will rise and they may earn a profit/loss from their position. On the other hand, negative sentiment may be perceived as a sign that the price will decrease and that they will make a loss/profit from their position. The setting of spread trading is ideal for showing how prevailing positive and negative news sentiment differently affects those individuals who trade on the expectation that price will increase and those who trade on the expectation that price will decrease.

It is widely suggested that media content is a proxy for new information about fundamentals and that this causes market volatility as new information drives price changes (Tetlock 2007). Research has found a significant and positive relationship between internet postings on message boards and subsequent trading volume (Antweiler and Frank 2004). Equally, it has been shown that high and low levels of pessimism in the media predict high market trading volume (Tetlock 2007). These studies have focused on the effect of sentiment on overall volumes but they have failed to examine the differential effects of sentiment on long and short trading.

Previous research examining the impact of news sentiment on trading volume has produced mixed results. It has been found that individual traders are net buyers of attention-grabbing stocks and that news volume is positively related to higher trading volume (Barber and Odean 2008). It has been suggested that increasing amounts of news information cause individuals to become more active in processing information and stimulating long trading in attention-grabbing stocks. However, it has also been suggested that individuals sell following high market-attention on a given stock because they are subject to the disposition effect (the tendency to sell winners too early and hold losers too long) and they seek to rebalance their portfolio to a desired set of weights (Yuan 2015).

It has also been found that individuals are net buyers after both extreme positive and negative earnings surprises (Hirshleifer et al. 2008) and that individuals act in a news-contrarian fashion to earnings announcements; buying/selling the market after negative/positive earnings news (Kaniel et al. 2012). While previous research has often focused on exploring earnings surprises, this paper examines how news sentiment (positive/negative) affects buying/selling activity. The literature reports mixed findings in relation to the way markets respond to news. The only other paper to explicitly explore the effect of sentiment found that individual's buy the market after positive sentiment (Lillo et al. 2015). Since this is the only previous paper to explore how sentiment is related to individual trading, Lillo et al.'s conclusion is the basis for the following hypothesis:

Hypothesis 1 (H1): Individuals' decisions to open long/short trades are positively correlated with positive/negative news sentiment.

This hypothesis is tested looking at the effect of sentiment on both the absolute volume of trades, and the relative cash volume of long versus short trades, developing a clearer view of how sentiment and trading behavior are related considering relative and absolute levels.

Previous studies have explored the effects of news and sentiment on aggregate market returns. For example, Peress (2014) found that newspaper strikes affected trading volume (lower volumes on strike days) but market returns were unaffected. Sul, Dennis, and Yuan (2016) showed that sentiment in tweets had a significant impact on stock returns for specific companies up to 20 days after the tweet.

The data enables examination of the impact of news sentiment at the individual trader level and the impact on both short and long trading returns. Understanding how sentiment impacts these aspects of individual trader returns will help illustrate how individuals should respond to sentiment.

Little research has examined how sentiment might affect risk taking; in particular, whether positive and/or negative sentiment is associated with risk-seeking/averse behavior. Schwager and Rothermund (2013) show that attention bias helps explain why people tend to avoid risky options in a gain frame but tend to choose risky options in a loss frame. Their argument is summarized as follows: a positivity/negativity bias in the loss/gain domain will lead to risk-seeking/averse behavior because of the increased salience of the positive/negative outcome of the risky option.

J. L. Zhang et al. (2016) found that positive and negative sentiment had an asymmetric effect on the levels of attention that individuals gave to news, with individuals focusing more on negative sentiment than positive sentiment. Since sentiment affects individual attention differently, it may also be the case that individual returns are affected differently by positive and negative sentiment. Indeed, individuals paying more attention to negative sentiment might generate higher returns. Further, individuals respond to negative news much more readily than positive news (Stieglitz and Dang-Xuan 2013) and so their returns might also be affected by sentiment differently depending on whether individuals make long or short trading decisions.

Finally, Individuals who read the same news information are likely to interpret it differently. In fact, Kaniel et al. (2012) found that informed traders have superior skills in interpreting public information and it has been suggested that the ability to properly interpret information is a key ingredient in distinguishing informed from less informed traders. Short sellers are generally regarded as informed traders and it has been shown that they are able to anticipate price falls (Boehmer et al. 2008). They appear to gain an information advantage because of their ability to analyse publicly available information, particularly negative information, at the moment it emerges (Engelberg et al. 2012). Whilst previous studies speculate on these relationships, no studies have been able to demonstrate how short trading is related to sentiment and profitability:

Hypothesis 2 (H2): individual trader returns and risk are asymmetrically affected by sentiment, and when individuals execute short trades, they achieve superior performance (compared to long trades) as evidenced by higher returns

In sum, to help understand how news sentiment affects traders, this paper explores how sentiment affects behavior and trading volume (H1), and differences in how sentiment affects the performance of short and long trades, and whether short trades achieve superior returns (H2).

4.4 Data

In order to test these hypotheses, individual trading data supplied by a large UK spread-trading broker is used. The majority of studies examining the relationship between online news information and financial markets have used stock market data. However, data from the spread trading market is valuable for studying the effects of news sentiment for a number of reasons: First, spread traders can speculate on the movement of underlying securities and can readily choose to either buy or sell a security. Traders can take a 'long' position and gain (lose) if the market rises (falls) or take a 'short' position and gain (lose) if the market falls (rises). Via their trading activity, individual spread traders clearly signify whether they think the market will rise or fall.

Second, spread trading data offers the prospect of examining large numbers of trades from a significant number of individuals. The data includes details of 190,363 trades from 3,583 individuals over a 6 month period, from 1st April to September 30th 2012. The data allows examination of a large number of short trades (52.38% of the trades are short trades) and this is often not possible using data of individual stock market investors.

Third, the spread trading data contains individuals' transactions so their idiosyncratic behavior and performance can be determined. In addition, individual spread traders may be less constrained than institutional traders by diversification requirements (Kaniel et al. 2012). Spread traders are more concerned with whether the price will go up or down relative to current levels. Consequently, providing more detail about individual trading preferences and behavior.

Fourth, spread traders are more likely to be affected by real time news information since they tend to be short-term investors, trading more frequently and holding trades for shorter periods (the average trade time in the dataset is 6 minutes). The time period from information release to an individual trading is relatively short and, consequently, confident that a particular news item is driving a particular trade. In addition, spread traders can leverage their position with small deposits, meaning that they are better poised to act on information contained in news and take positions that could yield high returns without being limited by their wealth.

The FTSE 100 is one of the most popular markets and is well suited for analysis on the behavior of spread trading. Spread-trading data is matched with Thomson Reuters News Analytics (TRNA) data covering the same period. Overall, there were 514,645 pieces of news related to the FTSE 100 during the sample period. This service provides instantaneous access to financial news and is used by investors and traders to support their trading (Gregoriou 2015). The service provides almost real time access to news stories online with metadata on each news article; including the timestamp of creation, the ticker symbol for the company to which the news relates, and the probability that this piece of news is positive or negative. Sentiment news scores are provided by SIRCA (The Securities Industry Research Center of the Asia Pacific). For example, one record in the dataset corresponds to the timestamp of the news release ('19 APR 2012 06:31:00.000'), the stock ticker symbol for the company to which the news is relevant ('AAL' for Anglo American PLC), the location of the exchange ('.L' for London stock exchange), the positive ('0.407582') and negative ('0.305649') sentiment score, and the title of the news announcement ('Anglo American says iron-ore production up 17%').

Employing TRNA provides an immediacy advantage over the data employed in previous studies examining how individuals analyse publicly available information. These studies focused on news reports in daily newspapers (see Engelberg, Reed, and Ringgenberg (2012)). The TRNA allows exploration of the short-term effects of online news sentiment on the market. The TRNA service has been used in studies examining how different classes of investors (governmental, households and companies) respond to information related to one stock (Lillo et al. 2015), how institutions are informed about news (Hendershott et al. 2015) and how news sentiment is linked to returns in the gold future market (Smales 2014). However, none of these studies has focused on the asymmetric impact of sentiment signals on individual traders.

In sum, combining individual level spread-trading transactions on the FTSE 100 with online news information related to FTSE 100 companies will provide a clear lens through which to explore the relationship between online news sentiment signals and individual behavior. The FTSE 100 provides a highly salient stock, which is frequently discussed in the media. This avoids the problems of exploring an individual stock, which may be subject to idiosyncratic price changes or company specific announcements.

4.5 Variables

4.5.1 Online news information signals

In order to calculate the sentiment signals from online news information a similar method to Smales (2014) and Hendershott et al. (2015) is used to aggregate the sentiment measures for the news articles from the TRNA dataset. The FTSE 100 is made up of constituent companies and for each day (d) the weightings (W) of the constituents (c) change:W_{Cd}. Following previous studies, W_{Cd} is used to normalize the relative sentiment attributable to each firm by using the firm's market capitalization (Hendershott et al. 2015).

The relative positivity and negativity measures of news related to the FTSE 100 prior to trade k are defined, respectively, as:

$$newsPos = \sum_{a \in A_p} SentPos_a * W_{Cd} \tag{4.1}$$

$$newsNeg = \sum_{a \in A_p} SentNeg_a * W_{Cd} \tag{4.2}$$

where $SentPos_a$ and $SentNeg_a$ are the positive and negative sentiment scores produced by the TRNA system for news article *a*. *newsPos* and *newsNeg* are calculated for all the news articles released in different time periods, *Ap,* prior to trade *k*. Research has shown that 60 minutes is an appropriate duration to explore the speed of diffusion (Nassirtoussi et al. 2014). However, to ensure the robustness of the results, a range of other time intervals prior to trade *k* (15, 30, 60, 120 and 300) was included. These variables are labelled *newsPos or newsNeg* {15, 30, 60, 120 and 300} *Min*, respectively. This technique is similar to that employed in other studies which have explored how quickly information disseminates (Antweiler and Frank 2004). Whilst previous research has examined the total sentiment on a day and how it relates to trading behavior (Lillo et al. 2015), isolating the sentiment prior to the trade better identifies the sentiment that was likely to impact trading decisions.

Trading at each hour interval are used to explore how sentiment is related to the volume of trades made and the relative cash volume. So considering the calendar hour 9-10am, the number of trades and the relative cash invested by an individual up until their last trade was calculated. So if an individual makes 5 trades, with the last being at 9.50am then the sentiment up until this time was calculated (i.e. the 50 minutes from 9:00am to 9:50am). These variables are denoted $newsPosHour_h$ and $newsNegHour_h$ for trade *k* by summing the positive (*newsPos)* and negative news (*newsNeg)* in the calendar hour prior period. Whilst previous research has explored aggregate day-level information, exploring at the calendar hour-level provides more detail. Also by considering information up until the last trade in this way, the analysis refines the focus on the information prior to the trade in an effort to better explore sentiment's effects.

News information that is released whilst a trade is live (i.e. after the trade is opened and before the trade is closed) is accounted for. Analysis examines whether there are any differences in the manner and extent to which the realised performance is affected by the information released prior to or during an active trade separately.

The relative positivity and negativity measures of news related to the FTSE 100 which is released while trade *k* is live are defined, respectively, as follows:

$$inTPos = \sum_{a \in A_{in}} SentPos_a * W_{Cd} \tag{4.3}$$

$$inTNeg = \sum_{a \in A_{in}} SentNeg_a * W_{Cd} \tag{4.4}$$

where A_{in} is the set of relevant news articles released while the trade was open.

4.5.2 Measuring behavior

In order to capture the behavior of individual traders, number of long and short positions trader *i* opens in trading hour *h* are defined as follows:

$$noTrades_{short\ i\ h} = \sum k \in K_{ih} \tag{4.5}$$

$$noTrades_{long\ i\ h} = \sum k \in K_{ih} \tag{4.6}$$

where K_{ih} is the set of trades placed by trader *i* in hour *h*.

Barber and Odean's (2008) provides a suitable approach to measure the degree of abnormal trading. Specifically, the number of long or short trades in hour *h* by trader *i*, divided by that trader's average number of trades in the same hour using historical information (*abTrades*):

$$abTrades_{short\ i\ h} = \frac{\sum k \in K_{ih}}{avgTradesPerHour} \tag{4.7}$$

$$abTrades_{long\ i\ h} = \frac{\sum k \in K_{ih}}{avgTradesPerHour} \tag{4.8}$$

where $avgTradesPerHour$ is the average number of short or long trades made by trader i in hour h throughout their trading history.

The relative propensity for traders to take long vs short positions in a given hour h, is subtracted the sum of GBP invested in buy and sell trades, as follows:

$$netBuySell_h = \sum_{k \in K_d} GBP_{buy} - \sum_{k \in K_d} GBP_{sell} \tag{4.9}$$

4.5.3 Measuring performance

In order to measure individual traders' performance the daily average rate of return for trader i in trading day d is constructed separately for long and short trades as follows:

$$ROR_{id} = \frac{\sum_{k \in K_{id}} PL_{ik}}{\sum_{k \in K_{id}} Stake_{ik}} \tag{4.10}$$

where PL_{ik} is the total profit/loss for trader i on day d when the position is closed and $Stake_{ik}$ is the GBP stake of trade k on day d.

The variability of returns for trader i on day d from their first trade to their k^{th} trade is determined as follows:

$$Variability\ of\ Returns_{id} = Std\ Deviation\ (PL_{ik}), k \in K_{id} \tag{4.11}$$

Variability of Returns for a trader's long and shorts trades is calculated separately. Importantly, overnight trades are included in the analysis and the relative ROR or variability of returns at the time of the closing of the trade is considered as the actual day in which the profit was realized. Having controlled for the sentiment during the trade, any overnight trading is unlikely to influence the results as the sentiment prior to the trade is still being examined. For the sake of robustness, the results were crosschecked by considering the time of the opening of the trade as the actual day the profit was realized and the results were similar.

4.5.4 Control variables

In order to isolate the effects of news sentiment, various market/demographic factors that have been shown to impact individuals' behavior are included. For example, young male traders have been found to take on more risk and hold more volatile stocks (Barber and Odean 2001). Research has shown that there is a strong relationship between message volume and volatility (Antweiler et al. 2004) and also changes in sentiment are positively related to changes in stock prices (Das and Chen 2007). One of the reasons for this is that new information is contained in news releases and investors then act on this information, driving prices up or down. Consequently, market volatility variables in the model control for the known effects of sentiment on volatility. Market volatility in the FTSE 100 in the period throughout which trade k is open ($MarketVol_k$) and volatility in market price in the hour prior to the opening of trade k (*volatilityHour*) are defined as follows:

$$MarketVol_k = \frac{S.D.(price \;\; p \in P_k)}{mean\,(price \;\; p \in P_k)} \tag{4.12}$$

$$volatilityHour_h = \frac{S.D.(price \;\; p \in P_h)}{mean\,(price \;\; p \in P_h)} \tag{4.13}$$

where $p \in P_k$ is the set of market prices from the opening of trade k to close and $p \in P_h$ is the set of market prices in the hour prior to trade k being opened. S.D. is the standard deviation.

Individuals have been shown to exhibit price-contrarian behavior, buying/selling the market when the price of a security decreases/increases (Kaniel et al. 2012) . Consequently, to account for overall market sentiment and to control for behavior related to changes in price, the price change of the FTSE 100 index relative to its opening price for different time periods prior to trade k [i.e. 15, 30, 60, 120 and 300 Minutes] are included. These variables are labelled *priceX* {15, 30, 60, 120 and 300 Minutes} *Min*, respectively, where *priceX* is defined as:

$$priceX = \frac{(opening\ Price_k - opening\ Price_p)}{opening\ Price_k} \quad (4.14)$$

where $openPrice_k$ is the price of the FTSE 100 when the trade is opened and $openPrice_p$ is the price *p* minutes prior to the opening of trade $k\{15, 30, 60, 120 \text{ and } 300\}$. Therefore, *priceX* is the relative change in price at the time trade *k* is opened. In a similar fashion, the relative price change of the FTSE 100 index for the hour before trade *k* is opened (*priceXHour*) is calculated.

It has been shown that individual wealth has a significant impact on information acquisition, portfolio choice and amount of risk taken. For example, wealthier individuals have been shown to take on more risk, hold riskier assets for longer and achieve higher returns (Barber and Odean 2001). Furthermore, Peress (2004) suggested that wealthier investors process news and make trading decisions faster. A proxy for the wealth of trader *i* is included to help control for these effects. Specifically, his/her average stake per trade (*stakeAvg*) and his/her opening balance at time of trade *k* (*openBalance*). The average number of trades an individual makes in the hour trade *k* is opened (*tradesHourAvg*) is also included.

In order to control for times when markets may be bullish/bearish (most individuals think the market price will increase/decrease) the *buySellRatio* (Bhattacharya et al. 2012) is defined as the total number of long trades relative to the total number of short trades on day *d* (*buySellRatio*):

$$buySellRatio_{\,d} = \frac{\sum_{k \in K_d} k_{long}}{\sum_{k \in K_d} k_{short}} \quad (4.15)$$

Previous studies have found that age and gender affect the levels of risk taken and the returns achieved (Barber and Odean 2001, 2002). To control for these factors a dummy variable for female investors (*female*) is included and by incorporating the age of the trader (years) at the time trade *k* was opened *(age).*

Those who trade using smart devices compared to other channels (e.g. telephone) are more likely to be connected to the internet and hence more affected by online news information. Consequently, to control for smart-device initiated trades by using a dummy variable (1 if the trade was opened using a smart device and 0 otherwise (*smart*)). The number of trades opened by trader *i* in hour *h* using smart devices as a ratio of all trades that hour (*smartHour)* is included. Finally, the hour of the day that the trade is made is included in the model by using hour dummy variables (*hourDummy).*

4.6. Testing hypotheses

Linear Mixed Models (LMM) account for trader heterogeneity (Engelberg and Parsons 2011). Since previous research suggests that positive and negative sentiment have asymmetric effects on individuals, analysis assesses to what extent sentiment affects buying and selling behavior differently. Consequently, interaction terms between positive and negative sentiment with a short trade dummy, *S*, which takes the value 1 if a trade is a short trade a 0 if it is a long trade are used so that differing effects on long and short trades can be isolated.

In order to test **H1**, that sentiment affects the amount of trading behavior, analyses explores how sentiment impacts variation in the trading volume: that is, how the volume of long and short trading in a calendar hour (e.g. 9-10am) is related to the aggregate positive and negative sentiment during that hour (up until the last trade). This hypothesis was also tested by examining to what extent sentiment appears to influence an individual to undertake an unusually high or low volume of long or short trades.

To examine the view that individuals' decisions to open long/short trades are positively correlated with positive/negative news respectively, the analysis shows whether and to what extent the online sentiment variables are positively correlated with the number of long and short positions opened. LMMtest whether the variables related to sentiment (*newsPos and newsNeg*) are useful for predicting the variation in the hourly volume of both long and short trades ($noTrades_{\text{long}}$ and $noTrades_{\text{short}}$) respectively, for trader i, by estimating the following:

$$\text{noTrades}_{long\ or\ short\ i\ h} = \alpha_i + \beta_1\ newsPosHour + \beta_2\ newsNegHour + \beta_3 newsPosHour{*}S^{13} + \beta_4\ newsNegHour{*}S + Control + \varepsilon_i \quad (4.16)$$

where the intercepts of the model (α) vary by trader and ε_i is the error term of the model. The control variables discussed in Section 5.4 are included in the model (indicated by *Control* in (16)).

Testing whether the β coefficients calculated for the news sentiment (*newsPosHour* and *newsNegHour*) are significantly different from zero explains whether and to what extent news sentiment impacts the volume of long trades in an hour, $noTrades_{long_{ih}}$. A series of planned contrasts examine the significance of the sentiment variable on short trades. For example, the sum of the coefficients for *newsNeg*S* and *newsNeg* and their significance test whether negative sentiment during the calendar hour (9-10am) has a significant impact on the volume of short trades in that calendar hour (e.g. 9-10am). The sum of the coefficients for *newsPos*S* and *newsPos* and their significance test shows whether positive sentiment has a significant impact on the volume of short trades. The planned contrasts test whether sentiment has a different effect on the volume of long and short trading.

Testing whether sentiment signals affect trading volume and, if a trader will trade long (or short) more or less than normal, by estimating the following LMM model:

---

[13] Shorting is abbreviated to S for the rest of the paper.

$$\text{abTrades}_{long\ or\ short\ i\ h} = \alpha_i + \beta_5\ newsPosHour + \beta_6\ newsNegHour + \beta_7 newsPosHour{*}S + \beta_8 newsNegHour{*}S + Control + \varepsilon_i \quad (4.17)$$

β coefficients assess to what extent and in what manner news sentiment explains an individual's abnormal trading volumes. For example, by exploring the significance of β_5 to assess the impact of positive sentiment released in the calendar hour prior to trading on the abnormal number of long trades during that hour. By assessing the significance of the coefficient formed by the sum of *newsNeg* + *newsNeg*S* to test to what extent negative news sentiment has on the abnormal volume of short trades placed by an individual.

To test **H1** in terms of relative cash volume and if individuals buy/sell the market after positive/negative news, analysis explores the relationship between sentiment and order flow (total buy stake minus total sell stake at the individual level). That is, how online news information relates to the bullish (buy) and bearish (sell) nature of an individual's trades. To achieve this the following LMM model is developed:

$$netBuySell_{ih} = \alpha_i + \beta_9\ newsPosHour + \beta_{10}\ newsNegHour + \beta_{11} priceXHour + Control + \varepsilon_i \quad (4.18)$$

This model captures how the variation in an individual's buy-sell imbalances can be explained by online news sentiment at each hour. To achieve this the positive (*newsPosHour*) and negative (*newsNegHour*) sum of sentiment in the online news information in a calendar hour prior to a trade is calculated. If β_9 and β_{10} are both significant and positive and negative, respectively, this would support **H1**. Hour-wise price changes (*priceXHour)* are accounted for by controlling for the relative change of price in the calendar hour (i.e. the price at 9am relative to the price at the last trade during the calendar hour 9-10am) so that any effects of price change on an individual's decision to trade are included. Lillo et al. (2015) show that individuals behave in a price-contrarian fashion, buying after the price decreases and selling after the price increases. Consequently, by including the price change during the same period the effects of price change found by Lillo et al. are included whilst isolating the effects of sentiment from those caused by price changes. The change in market price that hour and market volatility (*volatilityHour)* are also controlled for.

To test **H2,** individual trader's returns and risk are asymmetrically affected by sentiment, and when individuals execute short trades, they achieve superior performance (compared to long trades) as evidenced by higher returns, analysis explores to what extent news sentiment signals are correlated with an individual's short and long trading performance. In a similar manner to Zhang et al. (2016), returns and variability of returns are regressed against sentiment variables. In contrast to Zhang et al. (2016) trader heterogeneity is accounted by using LMM andit is assumed that the intercepts of the models α varies by trader. As in (16) and (17) the separate effects of sentiment on long/short trades are isolated using interaction terms between the dummy variable, *S* which takes the value 1 if the trade was a short trade and 0 for long trades, and each of the sentiment variables. Consequently, the following LMMs are estimated:

$$ROR_{long\ or\ short\ id} = \alpha_i + \beta_{12} k\ newsPos + \beta_{13}\ k\ newsNeg + \beta_{14}\ k\ newsPos*S + \beta_{15}\ k\ newsNeg*S + \beta_{16}\ Shorting + Control + \varepsilon_i \quad (4.19)$$

$$Variability\ of\ Returns_{long\ or\ short\ id} = \alpha_i + \beta_{17}\ k\ newsPos + \beta_{18}\ k\ newsNeg + \beta_{19}\ k\ newsPos*S + \beta_{20}\ k\ newsNeg*S + \beta_{21}\ Shorting + Control + \varepsilon_i \quad (4.20)$$

These LMMs explore an individual's returns and return variability for both short and long trades separately by assessing the sign and significance of the sentiment coefficients. For example, the sign and significance β_{12} is explored to determine the impact of positive news sentiment on the rate of return of long trades. In addition, planned contrasts are conducted to assess whether negative and positive news sentiment affect the rate of return of short trades differently (i.e. by exploring the significance of *newsNeg* + *newsNeg**S and newsPos* + *newsPos**S*, respectively*)*.

To test that short trading achieves superior performance compared to long trading as evidenced by higher returns, the dummy variable, *Shorting*, which takes the value 1 if the trade was a short trade and 0 for long trades in equation 19. If the coefficient for β_{16} is significant and positive then this will confirm that short trades are able to generate superior returns. A dummy variable, *Shorting,* is also included in equation 20 to test for any difference between long and short trading in terms of risk-taking.

The impact of news information on performance over time is shown by exploring the impact of the news sentiment released in different periods prior to trade *k* (i.e. 15, 30, 60, 120 and 300 Minutes). That is, the average of the sentiment in the period *k* (15, 30, 60, 120 and 300 Minutes) prior to each of the trades during the day to explore how this average sentiment is related to the performance.

To test H1, hour-long intervals capture how sentiment is related to the volume and relative cash value associated with long or short trades. To explore the relative cash value an interval with multiple trades is need to estimate the buy-sell imbalance. To test H2, the sentiment prior to each trade and average of this information is used to see how it correlates with an individual's daily performance. Individual hourly performance was also tested in the analysis but contained too much variance to show any conclusive results.

4.7 Results

4.7.1 Trader behavior

The results of estimating LMM (16) are shown in Table 1. The coefficients of the *newsPosHour* and *newsNegHour* variables demonstrate the effects of sentiment on the long trades only and the coefficients of *newsPosHour* + *S*newsPosHour* and *newsNegHour* + *S*newsNegHour* demonstrate the effects of sentiment on short trading only.

Table 4.1 Results of estimating LMM (16) to examine the relationship between sentiment and the volume of trades an individual makes each hour

Variable	**coefficient**	**S.E**	**Z**	**Sig**
(Intercept)	-0.1146	0.0490	-2.3382	**0.0194***
tradesHourAvg	0.3137	0.0042	75.2633	**0.0000****
openBalance	0.0089	0.0045	1.9791	**0.0478***
buySellRatio	0.0139	0.0026	5.3790	**0.0000****
smartHour	0.0168	0.0028	5.9615	**0.0000****
age	-0.0069	0.0041	-1.6880	0.0915
female	0.0063	0.0138	0.4555	0.6488
volatilityHour	0.0004	0.0028	0.1439	0.8856
priceXHour	0.1122	0.0030	37.0513	**0.0000****
newsPosHour	0.0580	0.0067	8.6398	**0.0000****
newsNegHour	0.0466	0.0066	7.0118	**0.0000****
Planned Contrasts				
*newsPosHour + S*newsPosHour*	0.1033	0.0063	16.3129	**0.0000****
*newsNegHour + S*newsNegHour*	-0.0022	0.0061	-0.3640	0.7159

**, * represent significant at 1% and 5% levels respectively in a 2- tailed test.

Sentiment has a significant impact on an individual's volume of long and short trading. The volume of long trading is positively related to the hourly positive sentiment (*β*(*newsPosHour*)

= 0.0580, z = 8.6398, p = 0.0000) and positively related to the hourly negative sentiment (β(*newsNegHour*) = 0.0466, z = 7.0118, p = 0.0000). Importantly, β(*newsPosHour*) is greater in magnitude than β(*newsNegHour*) ($|0.0580| > |0.0466|$) showing that hourly positive sentiment has a greater effect. Cross validation was also conducted using a sample of 80% of the dataset to determine the probability of achieving a β(*newsNegHour*) greater than or equal to 0.0580. This resulted in a z score of -1.702, suggesting that β(*newsNegHour*) is significantly different from β(*newsPosHour*) at the $p < 0.05$ level.

Previous research has found that individuals are net buyers after both extreme positive and negative earnings surprises (Hirshleifer et al. 2008). In a similar fashion, individuals make significantly more long trades in the presence of positive sentiment than negative sentiment, but long trading increases in the presence of both positive and negative sentiment.

The volume of short trades is also positively related to hourly positive sentiment as shown by the planned contrast (β(*newsPosHour* + *S**newsPosHour*) = 0.1033, z = 16.3129, p = 0.0000) but there is no correlation between short trading volume and negative sentiment. This finding suggests that while individuals make significantly more long trades in regards to sentiment (positive and negative), short trading is associated with sentiment-contrarian behavior, with an increase in positive sentiment being associated with an increase in short trades. This finding is supported in the recent literature where there is ‘a contrarian effect at 15 min time scale, meaning positive news would trigger negative trading decisions’ (Yang et al. 2017, p. 14). However, the results at 15 minute intervals on long trading does not show a similar contrarian effect and this is raised in the discussion.

Having explored how sentiment affects normal trading behavior, next analysis shows how sentiment affects abnormal trading behavior. The results in Table 2 show that sentiment has a significant impact on the degree of abnormally high or low volume of short or long trading which an individual undertakes. Abnormal volumes of long trading are positively related to the hourly positive sentiment (β(*newsPosHour*) = 0.0800, z = 10.8596, p = 0.0000) and positively related to the hourly negative sentiment (β(*newsNegHour*) = 0.0457, z = 6.2657, p = 0.0000). Abnormal volumes of short trading are positively related to the hourly positive sentiment (β(*newsPosHour* + *S*newsPosHour*) = 0.1276, z = 18.3627, p = 0.0000). Once again, the amount of abnormal short trading is unrelated to negative sentiment.

Table 4.2 Results from estimating the LMM (Eq.17) to examine the relationship between sentiment and the abnormal low or high trades an individual makes each hour, using interaction variables to control for the effects of sentiment on long and short trades separately

Variable	**coefficient**	**S.E**	**Z**	**Sig**
(Intercept)	-0.1105	0.0542	-2.0403	0.0413
tradesHourAvg	-0.2896	0.0052	-55.2544	**0.0000****
openBalance	0.0140	0.0063	2.2129	**0.0269***
buySellRatio	0.0194	0.0028	6.8366	**0.0000****
smartHour	0.0345	0.0032	10.7336	**0.0000****
age	-0.0355	0.0062	-5.7731	**0.0000****
female	0.0162	0.0216	0.7518	0.4522
volatilityHour	0.0035	0.0031	1.1563	0.2476
priceXHour	0.1564	0.0033	46.9027	**0.0000****
newsPosHour	0.0800	0.0074	10.8596	**0.0000****
newsNegHour	0.0457	0.0073	6.2657	**0.0000****
Planned Contrasts				
newsPosHour + *S*newsPosHour*	0.1276	0.0070	18.3627	**0.0000****
newsNegHour + *S*newsNegHour*	-0.0057	0.0067	-0.8504	0.3951

**, * represent significant at 1% and 5% levels respectively in a 2- tailed test.

These results go some way to supporting H1, in that sentiment has a significant impact on individual level trading volume, with positive sentiment being positively correlated with an individual's degree of long trading. However, an individual's degree of short trading is significantly correlated with positive sentiment. This indicates, perhaps surprisingly, that increases in positive sentiment are associated with more short trading. The signs of the coefficients and significant variables are consistent across periods of normal and abnormal hourly trading by an individual.[14]

4.7.2 Bullish/Bearish market

Table 4.3 Results from estimating the LMM (Eq.18) to examine the relationship between sentiment and the degree of bullish/bearish behavior of individual traders

Variable	**coefficient**	**S.E**	**Z**	**Sig**
(Intercept)	0.1452	0.1796	0.8090	0.4186
tradesHourAvg	0.0170	0.0054	3.1540	**0.0016****
openBalance	0.0116	0.0053	2.1840	**0.0290****
smartHour	-0.0085	0.0054	-1.5760	0.1150
age	-0.0043	0.0053	-0.8140	0.4154
female	-0.1074	0.0157	-6.8310	**0.0000****
volatilityHour	0.0011	0.0055	0.2110	0.8331
priceXHour	0.0117	0.0059	1.9920	**0.0464***
newsPosHour	-0.0363	0.0087	-4.1870	**0.0000****
newsNegHour	0.0240	0.0083	2.8890	**0.0039****

**, * represent significant at 1% and 5% levels respectively in a 2- tailed test.

To explore how the relative cash volume associated with long and short trading is related to sentiment LMM(18) is estimated; i.e. whether individuals are net buyers or sellers following positive and negative sentiment. The results displayed in Table 3 show that

[14] The hypothesis was tested at the day level as well and found the same results but only included the results for hour level since this provides more detailed analysis.

individual-level bull/bearish behavior is significantly related to sentiment. An individual selling the market is negatively correlated with hourly positive sentiment (β(*newsPosHour*) = -0.0363, z = -4.1870, p = 0.0000) and an individual buying the market is positively correlated with hourly negative sentiment (β(*newsNegHour*) = 0.0240, z = 2.8890, p = 0.0039). Consequently, individual traders act in a sentiment-contrarian fashion, buying/selling the market after negative/positive sentiment news.

My results directly contrast with Lillo et al. (2015) and Akhtar et al. (2013), who found individuals buy the market after positive sentiment. However, the finding of sentiment-contrarian behavior is supported by the news-contrarian behavior found in other studies (Kaniel et al. 2012).

Lillo et al. (2015) found that individuals behave in a price contrarian fashion, buying/selling when the price goes down/up. After controlling for the effects of price change on the bullish/bearish behavior (β(*priceXHour*) = 0.0117, z = 1.9920, p = 0.0464), individuals also act in a sentiment contrarian fashion buying/selling on negative/positive news sentiment. This suggests that the way in which individuals respond to the news depends upon both price and sentiment and both need to be considered when seeking to fully understand how individuals respond to sentiment. These results show that in terms of volume of trades, H1 is supported in that positive information is more positively correlated (cf. negative) with the number of long trades. However, in the terms of the relative cash volume of trades, the results show sentiment-contrarian behavior: positive/negative sentiment is associated with individuals selling/buying the market.

4.7.3 Performance

4.7.3.1 Returns

The results of estimating LMM (Eq. 19) to examine the relationship between online news sentiment and the rate of return (ROR) achieved by an individual are presented in table 4. Eq (19) incorporates interaction terms between the shorting dummy (*S*) and the positive and negative news variables (*newsPos* and *newsNeg*, respectively). In order to assess the effects of positive and negative sentiment on the returns of short trades, respectively, planned contrasts identify the sign and level of significance of the combined coefficients of *newsNeg* + *S*newsNeg* and *newsPos* + *S*newsPos* . In a similar manner, the impact of news sentiment released during the period the trade is held open is controlled for (via variables: *inTPos* and *inTNeg*).

Table 4.4 Results of estimating the LMM (Eq.19) to examine the relationship between online news sentiment and an individual's average daily rate of Return (ROR) for short and long trading respectively.

	Time interval prior to a trade during which the positive/negative nature of news sentiment is assessed				
Variable	**15 min**	**30 min**	**60 min**	**120 min**	**300 min**
newsPos	0.0113	-0.0042	-0.0067	-0.0242	-0.0274
newsNeg	-0.0197	-0.0131	-0.0190	-0.0109	-0.0154
inTPos	**-0.1880***	**-0.1850***	**-0.1914***	**-0.1872***	**-0.1957****
inTNeg	0.1093	0.1060	0.1125	0.1081	0.1166
PriceX	**-0.0437****	**-0.0327****	-0.0217	-0.0106	0.0018
Shorting	**0.0260***	**0.0262***	**0.0262***	**0.0262***	**0.0263***
Planned Contrasts					
*newsPos + S*newsPos*	**-0.0587****	**-0.0732****	**-0.0937****	**-0.1048****	**-0.1254****
*newsNeg + S*newsNeg*	0.0314	0.0373	**0.0475***	**0.0491***	**0.0627****
*inTPos + S*inTPos*	**0.2486****	**0.2464****	**0.2493****	**0.2468****	**0.2471****
*inTNeg + S*inTNeg*	**-0.1801****	**-0.1780****	**-0.1809****	**-0.1785****	**-0.1789****

**, * represent significant at 1% and 5% levels respectively in a 2- tailed test.

The results presented in Table 4 demonstrates how individuals' daily returns relate to the average sentiment of news released across different time intervals prior to each trade. For short trades, an individual's average daily returns are significantly and positively related to negative sentiment (considering sentiment at time intervals greater than 60 minute prior to the trade) and significantly and negatively related to positive sentiment, across all time intervals examined (determined via the coefficients of *newsNeg* + *S*newsNeg* and *newsPos* + *S*newsPos*, respectively). These results suggest that in relation to short trading, a momentum type strategy following news in at least the last 60 minutes prior to the trade, appears to lead to increased profits (i.e. short trading following negative sentiment). These results are similar to Akhtar et al. (2013) who found that in the Dow Jones futures index market individuals exhibited negativity bias and negative sentiment was followed by price falls. Consequently, individuals are rewarded by correctly perceiving negative sentiment and trading on the basis that the price will decrease.

It is interesting that individual's short trading returns are immediately significant in relation to positive sentiment but the same returns are only significantly related to negative sentiment after 60mins. This finding is supported in the literature where positive information is more likely to affect immediate decision-making and impulsive buying behavior, whereas negative information affects more long term decision-making (Sul et al. 2016).

The results suggest that for long trades there is no relationship between the positivity/negativity of news sentiment and a trader's returns (coefficients of *newsPos* and *newsNeg*) in Eq. (19) are not significant for any of the time intervals examined). It has been shown that market prices respond much more readily to negative news than positive news (Tetlock et al. 2008). As such, an individual who trades short following prevailing negative market sentiment in at least the last 60 minutes can benefit from the way in which negative sentiment leads to price changes. A similar strategy with long trading may not be as successful because there is no similar correlation between sentiment and individual's returns.

Importantly, whether short trading achieves superior returns compared to long trading is tested by including the dummy *Shorting* in equation 19. *Shorting* is significant across all the time periods showing that short trading achieves superior returns compared to long trading. This finding confirms part of H2.

4.7.3.2 Variability of returns

The results of estimating LMM (Eq.20), to examine the relationship between online news sentiment and the daily variability of individuals' returns (standard deviation in returns), are presented in table 5. For long trades, the variability of returns is negatively correlated with positive sentiment and positively correlated with negative sentiment across all time intervals (*newsNeg / newsPos* coefficients are all significant and negative/positive). Consequently, when individuals trade long against the prevailing negative news sentiment, the variability of their returns increases but when they trade long, in line with the prevailing positive news sentiment, the variability of their returns decreases. The variability of returns could be regarded as a measure of risk taken by traders and consequently, these results suggest that in the face of negative/positive sentiment, long trades take greater/less risk.

Table 4.5 Results of estimating the LMM (Eq.20) to examine the relationship between online news sentiment and the daily variability of Returns (standard deviation in returns) to which individuals expose themselves

	Time interval prior to a trade during which the positive/negative nature of news sentiment is assessed				
Variable	**15 min**	**30 min**	**60 min**	**120 min**	**300 min**
newsPos	**-0.1230****	**-0.1536****	**-0.1596****	**-0.1626****	**-0.1776****
newsNeg	**0.0698****	**0.0879****	**0.0906****	**0.0967****	**0.1083****
inTPos	**1.5235****	**1.5340****	**1.5425****	**1.5424****	**1.5538****
inTNeg	**-1.3884****	**-1.3991****	**-1.4075****	**-1.4074****	**-1.4190****
PriceX	**0.0508****	**0.0639****	**0.0703****	**0.0703****	**0.0783****
Shorting	0.0091	0.0096	0.0091	0.0091	0.0090
Planned Contrasts					
*newsPos + S*newsPos*	-0.0204	-0.0387	**-0.0599***	**-0.0622***	**-0.0698***
*newsNeg + S*newsNeg*	-0.0249	-0.0184	-0.0041	-0.0015	0.0011
*inTPos + S*inTPos*	**0.4682****	**0.4678****	**0.4734****	**0.4765****	**0.4769****
*inTNeg + S*inTNeg*	**-0.2910****	**-0.2903****	**-0.2954****	**-0.2984****	**-0.2990****

**, * represent significant at 1% and 5% levels respectively in a 2- tailed test.

For short trading, positive sentiment significantly affects variability of returns considering sentiment at least in the last 60 minute prior to trading. In particular, variability of returns is significantly decreased around positive sentiment at 60, 120 and 300 minute periods. Again this is partially explained by the findings in the literature that positive information is more likely to affect immediate decision-making (Sul et al. 2016). These results show that in terms of variability of returns, only positive information affects short trading. Consequently, sentiment asymmetrically affects short and long trading.

Overall, these results confirm H2. In particular, when individuals trade short they achieve significantly higher returns compared to long trades. They can also increase their returns when they trade in line with prevailing negative market sentiment (after 60 minutes) but their returns decrease when prevailing market sentiment is positive. Short trades also achieve significantly lower variability of returns after 60 minutes. This suggests that when individuals trade short their risk taking behavior is less affected by prevailing market sentiment than that of long trades (since for those who trade long, their risk increases/decreases when the prevailing sentiment is negative/positive). However, the risk taking of long trades does appear to be affected by prevailing sentiment and long trades do not appear able to increase their returns by responding in either a contrarian fashion or in line with market sentiment.

4.8 Discussion

It is important to recognise the transformational impact that the internet has had on information dissemination and diffusion amongst individual investors. Information now spreads rapidly at intervals much finer than days (Liu and Ye 2016). Existing research has studied how earnings announcements affect individual behavior but novel research exploring more general items of news, including qualitative information (hard to quantify in earnings announcements) is limited. This study sets out to help understand how an important new source of news, online news, affects financial trading in real-time.

The finding that individuals buy following the release of negative news sentiment and sell after the release of positive news sentiment contradicts the findings of Lillo et al. (2015), that buying activity was associated with positive sentiment. The contrast in these results may arise from the different methods employed to estimate sentiment. Lillo et al. (2015) only counted the number of negative and positive words in the article title and these were not necessarily related to a specific company. By contrast, the sentiment score provided by TRNA is based on the whole article and the sentiment relates to a specific company. The TRNA sentiment scores have been found to be reliable measures in previous studies (Hendershott et al. 2015; Smales 2014). Furthermore, Lillo et al. (2015) only consider trading related to Nokia and their results may therefore be stock specific. In addition, this paper explores the relationship at much finer level of detail – exploring intra-day relationships between sentiment and trading.

The finding that individuals act in a sentiment contrarian fashion could be related to the nature of spread trading markets and the availability bias. Spread trading markets are typified by much shorter trading periods and can even incur extra costs for holding trades overnight. Consequently, spread traders may be more focused on contemporaneous information than investors in stock markets who generally hold stock for longer periods.

It is unclear from the literature if individuals who employ new-contrarian strategies are trading naively or as a profit-taking strategy (Kaniel et al. 2012). However, results show individual's average daily returns are correlated with sentiment and, a momentum type strategy rather than a sentiment-contrarian fashion would result in returns that are more profitable. Therefore, these findings suggest that sentiment-contrarian behavior is not related to a profit enhancing strategy since it is only trading on the basis that the market price will decrease following negative news sentiment that individuals can increase their returns.

Understanding the effects of sentiment-contrarian behavior is important in the context of post-earnings announcement drift (PEAD), where stocks returns have been shown to drift in the direction of earnings surprises over a number of days and even months. Sentiment-contrarian behavior could contribute to PEAD, as individuals' trade in a contrarian manner following price announcements and in a contrarian manner related to sentiment. Hirshleifer et al. (2008, p. 1547) found that individuals did not contribute to PEAD. However, the relative cash volume is oppositely related to sentiment and individuals could contribute to prices slowly adjusting to news information since individuals in fact trade in opposition to sentiment in relative cash terms and the increased volume of short trading following positive sentiment.

By separately exploring the performance of short and long trades this paper contributes to the view in the literature that short sellers are generally more informed by showing that they do indeed, achieve significantly higher returns. It has been suggested that short traders are better able to analyse publicly available information (Engelberg et al. 2012). The results show that short trading does indeed achieve significantly higher returns than long trading, and it is only by trading short following negative news sentiment that returns of traders can increase. Consequently, it may be those individuals who do this who give short traders the reputation of being informed.

Individuals who trade short following the release of negative sentiment achieve higher returns, and those who trade short following the release of positive sentiment achieve lower returns. These results suggest that the advantage reported in the literature for short traders (Boehmer et al. 2008) may come in part from those who are sensitive to negative news sentiment and respond by shorting the market. This could be attributable to the asymmetric effects of sentiment reported in the literature whereby negative sentiment has a stronger impact on behavior than its positive counterpart (Akhtar et al. 2013). This may lead to prices reacting (falling) to a greater degree in the wake of negative (cf. positive) sentiment.

It has been suggested that the superior returns generated by short trading do not arise from public information but result rather from private information (Boehmer et al. 2008). However, this paper shows that superior short trading returns are correlated with publicly available information prior to trading. In particular, those traders who appropriately perceive news sentiment as negative and short the market achieve higher returns. This view is consistent with the wider literature where short traders are regarded as possessing exceptional skills in deriving value-relevant information from information and it is precisely the moments around news release where this skill emerges (Engelberg et al. 2012).

Previous research has found that individual trader's risk preferences are impacted by the way individual's interpret numbers and particularly, by the left digit effect (Fraser-Mackenzie, Sung, and Johnson (2015)). This paper shows that the way in which individual traders interpret sentiment can also influence risk-taking behavior. These results are important, since they suggest that risk-taking behavior associated with long and short trading significantly differ. The role of sentiment in impacting risk-taking is clearly important, with both bubbles and crashes potentially being influenced by the associated positive or negative news sentiment.

Recent studies have found a similar sentiment-contrarian effect with positive news triggering short selling at the 15 minute time scale (Yang et al. 2017). Similarly, these results show that positive sentiment is significantly and positively correlated with short selling but there is no similar effect with negative sentiment and long trading. Although there is no immediate obvious reason for this in the current literature (Yang et al. 2017) the results herein show that short trading following positive sentiment is associated with negative returns while a momentum-type strategy surrounding positive sentiment would be associated with positive returns. Therefore, further research should identify the reasons why individuals short following positive news even though this strategy does not generate profitable returns.

4.9 Conclusion

Sentiment asymmetrically affects individual performance. This paper confirms that short trading does actually achieve significantly higher returns than long trading. A momentum type strategy following prevailing negative sentiment can lead to increased profits. These findings are particularly important for regulators in the contexts of market crashes. Since individuals who recognise the effect of negative sentiment on prices and short the market following prevailing negative market sentiment can profit, regulators can better understand how negative sentiment can directly affect market stability.

Chapter 5 Conclusion

This section concludes the thesis by summarising the main contributions of the three papers and the importance of these contributions in the wider literature. Then discussion shows how the three papers come together to form part of a larger totality that provides new knowledge and understanding of employing web information in decision-making. Finally, some limitations and future research areas are identified.

Recent transformations in web-based information availability have increased the amount of data that can be used in forecasting. With this abundance of data, the growing need to capture and employ web-based information to improve forecasting becomes a prevalent challenge (Zhang et al. 2018). As the quantity of data exponentially increases, so too does the diversity of data that becomes an inhibiting factor to harnessing web information for forecasting as data is produced by diverse sources in multiply types of formats. While some data such as numerical pricing is easier to incorporate directly in forecasting models, more unstructured data relating to geographical and textual information proves more difficult as it has to be structured and processed before being used in models (Agarwal and Dhar 2014). Therefore, the ability to process new unstructured information into models proves how increased access to information drives faster and more effective decisions highlighting the value of increased information availability (Bharadwaj et al. 2013). Since data exists in many different formats, the ability to determine quantitative signals from unstructured data is an important endeavour.

Contributing to the literature on how new web-based information can be used in forecasting, this thesis provides empirical examples of how unstructured geographical data and textual data can be used in forecasting models. The core contribution of paper 1 shows how geospatial information can be utilized to improve the prediction of performances of racing horses compared to odds information. Further, paper 2 shows how new distance information made available online through tools such as Google maps makes it possible to elicit expert knowledge to improve predictions compared to the market. Finally paper 3 shows how online news information helps explain the decision-making behavior and performance of individual traders. These three papers combined show how web information is offering the potential to improve forecasting accuracy as new data adds superior information to further contextualise and increase awareness around decision-making in the real world. Furthermore, increased access to information enables individuals to use diverse types of information that has proven predictive value in decision-making, which ultimately increases the types of information that are used in forecasting and thus, the information which is compounded into market prices.

New online web information has significant impact on market efficiency. Research has shown financial markets work to effectively aggregate information and IT-enabled information transparency actually improves market performance as improved information transparency actually increases trading activity, leading to improved information efficiency (Yang et al. 2015). Therefore, information plays a key role in financial markets and although IT has been transforming markets for decades, research has failed to show the extent to which IT affects financial markets (Zhang and Zhang 2015). Collectively, the three papers show how IT-enabled information can improve forecasting accuracy. While in papers 1 and 2, analysis focused on how IT-enabled geospatial information could improve decision-models, paper 3 showed how sentiment was related to individual trading volume and performance. Paper 3 therefore supports the theory that increased access to new information can improve market efficiency as individuals actually trade based on news sentiment and can achieve profitable returns by trading short following prevailing negative sentiment. Theory on how increases in information availability leads to efficiency has relevance and implications in the wider literature, beyond financial markets, showing how information from diverse online sources such as social media and news websites drives behavior. Individuals are frequently using web information to inform behavior both online and in the real world. For example, checking the weather or the best route to a destination involves using web information to inform behavior.

Online information allows individuals to more accurate decisions that has economic benefits. Paper 1 shows how new information provides brief opportunities for profitable returns to prosper from geospatial information as the market converges. As such, profitable opportunities arise as new information is released and prediction scores may be diluted over time, as the information becomes public knowledge to the market. This finding that there are profitable opportunities as the market converges supports the theory of adaptive market efficiency. The adaptive market hypothesis states that price reflects information from the combination of environmental conditions and the nature of the market participants (Lo 2004). In this sense, markets are adapting to a perpetually evolving IT-enabled environment where technology makes available new information and profitable opportunities arise with trading strategies based on this superior information.

An important regulatory implication raised by this thesis is that access to superior information and IT-enabled technology can offer individuals a significant advantage in competitive markets. Also, the efficient operation of markets depends on the ability to aggregate information. By understanding how IT-enabled information is linked to information transparency, markets can be better regulated by ensuring fair and equal access to information.

Results on the speed at which the market adjusts to the new information as shown in paper 1 negates the efficient market hypothesis and the idea that information is completely and immediately incorporated. Rather, in accordance with adaptive efficiency, efficiency is subject to the changes in behavior and environment, where market participants are in a state of learning and evolving the new practices that are better able to take advantage of technology-mediated information (Lo 2012).

Markets are constantly adapting and this thesis has shown how the information environment changes and evolves, causing markets to adjust to this information change. These three papers contribute to the growing literature on how markets absorb information and the dynamics of how markets constantly reach new equilibriums of information efficiency (Nassirtoussi et al. 2014). Researching how new web-based information becomes discounted in market prices will become important as the amount of information available increases. These three empirical studies, show how geographical and textual data can be harnessed for decision-making, serving as examples of modelling diffusion of web-based information.

A key challenge for economic analysis is quantifying the economic value of accurate forecasts (Lessmann and Voß 2017). One of the benefits of studying the value of information in relation to financial markets is that the economic benefits of using information can be estimated. Indeed, paper 1 shows the simulated profits from wagering strategies that employ geospatial information. Similar to other research, this thesis has adopted the information-based view to study financial markets (Yang et al. 2015). That is, research focuses on information and how it can be used to add to forecasting models for incremental improvements in accuracy. An important consideration though, is that individuals differ in their ability to process information.

Recognising experts and novices exist as distinct sub-populations on the demand side of financial markets is fundamentally important to understand how different groups use information. Advances in technology have influenced information retrieval behavior and a prime example is online mapping services which have transformed individual decision-making capabilities (Constantiou et al. 2014). Online information is constantly available to decision-makers and in order to enhance capabilities, it depends how well individuals use such information. Since individuals differ in their abilities to use information (Chen and Zeng 2016), identifying how novices and experts use online information reveals has a direct link to how much value is created. This thesis has shown that the crowd are able to fully discount expert elicited knowledge and that individual trading behavior is correlated with sentiment.

Researchers have called for more analysis on how experts use information and how their subjective beliefs can be incorporated into forecasting models (Alvarado-Valencia et al. 2017). By understanding how individuals use information available to them and how it influences their behavior, research can highlight how information available on the web is influencing real world decision-making, and even bringing about behavioral changes. However, a recent review exposed the gap in the literature related to eliciting subjective beliefs from expert's behavior and the lack of real experts (not students) in the real world (Werner et al. 2017).

Paper 2 shows how distance information from Google maps can explain decision-making of novices and experts. By observing horse race trainers who represent real experts, the analysis highlights the value of using distance information in eliciting the subjective beliefs that underlie expert decisions. Expect elicitation is an important concern in the forecasting literature (Bolger and Wright 2017) and this paper shows how distance information can be used to tease out subjective beliefs that incrementally improve forecasting. Therefore, real world experts can provide ideal examples individuals who possess superior ability to make forecasts which offers a better perspective to support the widely used student-subjects used in previous research (Werner et al. 2017). Using actual real world experts can then offer examples of experts making accurate predictions with information despite the theory on the performance paradox of experts (Camerer and Johnson 1991).

On the other hand, the betting crowd are able to discount complex distance related information. The results of paper 2 show that the complex beliefs of experts provide statistically superior winning estimates for a limited time, suggesting that the crowd learn to discount any advantage from expert elicited beliefs. Though the lack of data on individual betting behavior makes any analysis difficult to directly prove that individuals learn over time and place bets in accordance with the probabilities derived from the model using complex distance information. The results of testing how distance information provides added (but albeit decreasing) predictive value for a limited time shows how the odds effectively discount information contained in the distance-based variables.

Observing how individuals make decisions taking into account distance will be of benefit to a wide range of applications: when individuals make decisions on where to eat, where to live or even where to go on holiday, distance plays a fundamental role in the decision-making process. Location data will become increasingly useful to target individuals with the right product or search results and as such, combining the statistical results with information related to an individual's historic behavior and personal preferences will enable corporations to maximise revenues.

Information diffusion (that is, how much a financial market changes as new technologies are actually used) has been overlooked (Hall 2004). Since information exists in varying formats, it is necessary to understand the factors influencing the diffusion of different forms of information within different markets. As Rogers (1995) points out, the diffusion of information depends on four primary factors including the innovation itself, which channels are used for communication, time itself and the participants. Diffusion then becomes a complicated interplay between a multitude of different factors. This thesis has not offered an extensive comparison of different types of information and the rate at which these different types diffuse.

However, comparing the results of paper 1 where information diffused over a number of years, with paper 3 where information diffused in a matter of minutes shows that these two types of information diffuse differently. Diffusion takes place with different types of information but also within different markets, at different time-periods and in different channels (Google maps vs financial news). Therefore, this thesis cannot make any claims on how fast information will diffuse – only highlight the important primary factors that must be controlled for to make any generalizable claims of information diffusion rates of different types. Similar to other literature, this thesis shows that different types of information diffuse at different rates (Sundararajan et al. 2013). Finally, since it has been shown that the speed of diffusion affects the value of information (Manela 2014), betting markets provide an ideal experiment to develop theory on diffusion rates and value. Indeed, new information offers individuals significant advantages and opportunities for positive returns meaning they will trade away these information advantages until the market aligns. Market convergence studies offer one potential to study how markets adapt over time to new information, highlighting how the adaptive market hypothesis aptly captures how markets adjust to new environments.

Extending the theme of web information in decision-making, paper 3 identifies how individual's behavior and performance are related to sentiment information. Using an even more fine-grained level of analysis than the first two papers, this research identifies the effects of information at the individual level. Previous research has assumed that markets respond directly to news information but in fact, news information diffuses among the investor population and individuals drive price fluctuations through trading (Peress 2014). Similar to papers 1 and 2, research has shown how information is related to prices at the aggregate level and insufficient data has enabled investigation into how individuals drive price changes. Overcoming some of the shortfalls of papers 1 and 2 related to not being able to show individual betting patterns, paper 3 examines how sentiment is correlated with individual trading behavior.

Thus far, the literature has explored how information from newspapers or online resources affects trader behavior looking at day or even month long diffusion time intervals (Lillo et al. 2015; Yuan 2015). However, news information diffuses among individual investors at intervals much finer than days (Liu and Ye 2016) and research needs to understand how online news information affects within day trading (Peress 2014).

Paper 3 explores how sentiment affects individual decision-making at unprecedented levels. The study shows that information diffuses among the investor population at the intra-day level, providing the most detailed understanding of how online information is used by individuals. In real world trading, individuals have constant access to up-to-date information and research must account for how the internet has facilitated the speed of modern financial trading.

As one of the first studies to identify sentiment-contrarian behavior and the first study at the individual level, paper 3 contributes to the financial literature theory on how individuals are affected by sentiment (See also Yang et al. 2017). Analysis shows that the individual buy-sell imbalances can be explained by sentiment and individuals act in a sentiment-contrarian fashion. This contrarian behavior is important in the context of price drifting and helps to explain why market prices slowly adjusts to news. Important regulatory concerns are raised and policy makers should be aware of the manner in which individuals act in a contrarian manner, even though (as shown by paper 3's results on returns) it can be unprofitable.

Overall, the thesis has identified how information derived from IT can incrementally improve forecasting at three different levels: market, expert/novice and individual.

Structurally, the analysis has shown how information diffuses through a market, how individuals learn to use information and how information affects individual's behavior and performance differently. Having covered the two extreme levels of the spectrum – information at the aggregate market level and information at the individual level – this thesis makes significant contribution to providing a holistic view of how web information improves decision-making and forecasting at various levels in the real world.

For many individuals, the web has become essential for everyday information gathering and decision-making (Roscoe et al. 2016). The way information is presented on the web has changed over time and this thesis has shown how decision-making has duly been impacted in three empirical studies. The contributions made by this thesis relate beyond financial decision-making and have implications for a broad range of literature where decision-making is a central tenet.

Individuals frequently use the internet to discover information related to health, current events, consumer goods purchasing and et cetera (Roscoe et al. 2016). As technology, and particularly the web, becomes more immersed in our everyday lives, it is important to recognise the impact that technology has on decision-making. This thesis offers significant contributions to understanding how web information has influenced real world decision-making and, as the web becomes more influential and immersed in our decision-making processes, the importance of this area of research will only proliferate.

5.1 Limitations and future research

Unlike laboratory studies, empirical analysis cannot fully control for specific factors that may be important in a study. The complexity and abundance of information that can be used to forecast winning estimates makes it difficult for any model to encompass all such information (Schotter and Trevino 2014). Furthermore, betting markets themselves are dynamic and research has identified various biases and factors that impact the underlying efficiency of markets (Sung et al. 2012). The inability to control and observe for all the relevant variables is a limitation that is not specific to the three empirical studies herein, but known more widely to be a limitation to economic studies in general. While empirical studies are not able to control and manipulate various factors as research in laboratories can, empirical data relates to actual behavior in complex real world situations.

As datasets from a variety of real world contexts become available to researchers, the naturalistic setting will become the new behavioral research lab (Chang et al. 2014). Moreover, future research would benefit from exploring to what extent the findings of this thesis are also found in the lab. For example, laboratory research on sentiment-contrarian behavior will be able to control for factors which have been excluded in this study, such as what kinds of information an individual is exposed to, where the information is from and what format the information is presented in.

Ultimately, recognising the issue of bounded rationality becomes increasingly important as information volume and diversity escalates. Since individuals are limited by cognitive resources and are only able to process a certain amount of information, they make satisfying, rather than optimal decisions (Lo 2004). Employing web information in financial decision-making depends largely on the skills of individuals to discover, collect and properly interpret information. Individuals vary in how well they are able to use information from various sources and how much information they can process. Whether individuals know about all the information available to them on the web is dubious and the amount of information on the web may be too much for any one individual to process at one time. For example, although Wikipedia is well known, it would be near impossible for one person to digest and use all the information available to them. Future research should explore how much information individuals are able to use and, at what point, information overload occurs.

This thesis has focused on how individuals could use and interpret information. It has to be conceded that individuals vary in their skills and abilities. Exploring experts and the betting crowd separately offers one way to study how distinct sub-groups differ in their information processing abilities.

An individual's ability to interpret information, and the speed with which information diffuses will likely be impacted by how easy information is to understand. Information exists in varying types, for example, text, video, image and sound require varying processes to gather value from such information. For example, the diffusion rate seen in paper 1 compared to paper 3 might be attributable to the difference between geospatial versus textual information. Further research should focus on how data is presented and how quickly individuals are able to use information in varying formats.

There is an abundance of information available on the web and not all of it is factually correct. Some information is even created for the purpose of misleading individuals. How trust-worthy the source of the information is will also influence how likely individuals are to use information from IT. This thesis has focused on how individuals use information that is shown to be significantly useful to predict probabilities. Further research, understanding how well individuals detect and react to malicious information will expand knowledge of employing web information in financial decision-making.

The particular context of horserace betting and spread trading markets make the contributions specific to the data employed and just like any empirical study, the observed results do not necessarily generalize to other settings (Lessmann and Voß 2017). As mentioned in the introduction, betting markets offer a lens through which researchers can identify theory and ideas that relate to wide financial markets.

Replicating the studies with new and different data would provide valuable support extending the contributions found in this thesis on how different information diffuse through the market at varying speeds and how valuable different information is for improving forecasting accuracy. Development of tracking technologies and GPS data enables new digital sensing that can provide highly mobile, contextually relevant location data to improve decision-making (Chen et al. 2012). GPS technology is being used to track a horse's position on the track and this newfound data offers a parallel study to Papers 1 and 2, to explore how such data can improve forecasting.[15] Also, Google search is frequently used by individuals to discover information and therefore, serves as a proxy to explore interest in stock that can be used to forecast stock prices (Yu et al. 2018). Further research on how Google search data related to the FTSE 100 would complement paper 3 and provide rich understanding of how online information availability, sentiment and information search are all linked.

[15] http://www.turftrax.co.uk/tracking-technologies.html [Accessed 17/02/2018]

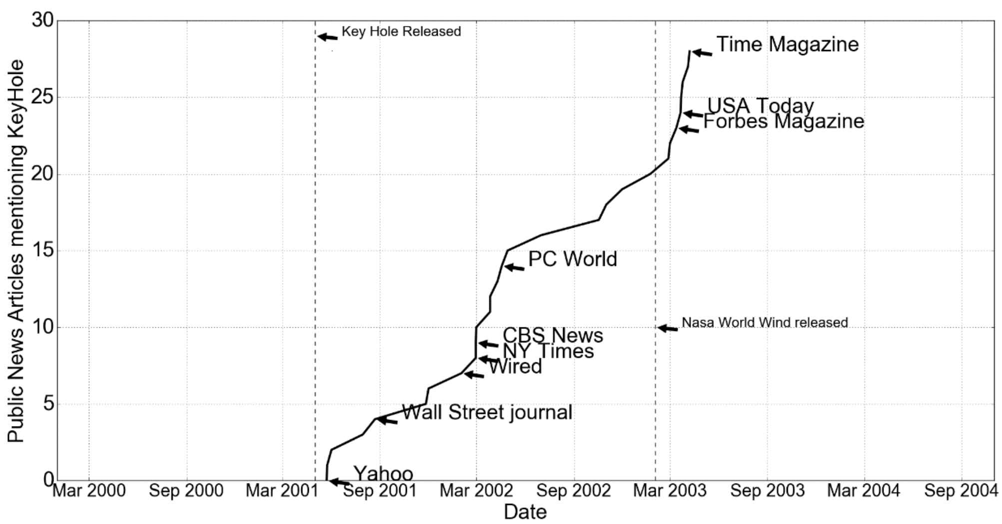
Chapter 6 Appendices

Appendix A: Brief introduction to Virtual Globes

The concept of a digital earth, where users could navigate the earth in 3-Dimensions on a computer screen was first conceptualised on the internet by Keyhole, Inc., with their software EarthViewer 3D in 2001. Google bought Keyhole Inc. in 2004 and Google Earth was released in 2005. Although there were competing virtual globe technologies (e.g., NASA's World Wind, ArcGIS Explorer), Google Earth has risen to prominence and provided data for studies in a number of high-ranking journals. It has been widely used in a multitude of fields such as Aerospace engineering, urban planning and business (Goodchild et al. 2012).

Since geospatial data has become freely available to the public, mass media channels have been influential in raising awareness of the technological innovation; exemplifying the role of media as communicators in the technology adoption process. As can be seen in Figure 5, KeyHole rose in prominence from 2001. Specifically, articles in popular media such as Time, Forbes Magazine and the Wall Street Journal were influential in raising the profile of VGs.

Figure 6.1 Media coverage of KeyHole in popular press



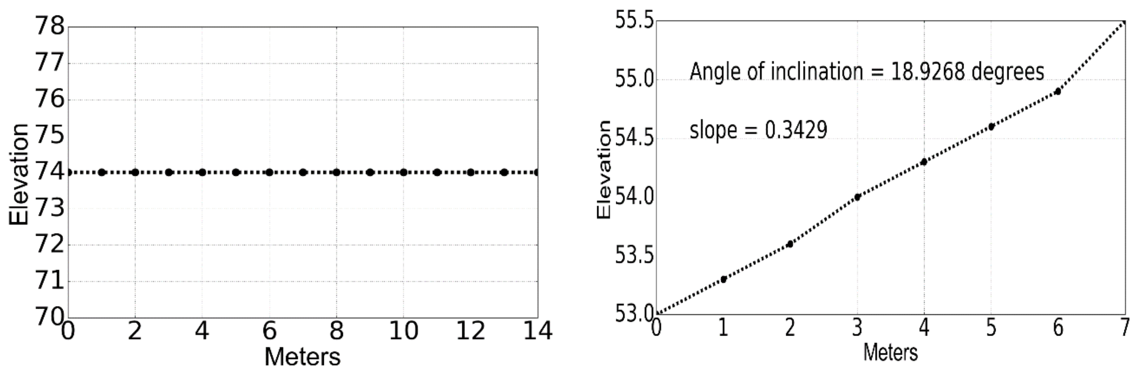
Appendix B: Description of Geospatial variables

Variable: *CAMBERS*

Definition: The number of MPs with flat cambers as a ratio of all MPs in the final quarter of the course.

Concept: Hobbs, Licka, & Polman (2011) indicate that horses move at different speeds on flat and banked surfaces. It is likely that horses with different physiques may be more/less inconvenienced by different cambers (i.e. the slope towards the centre for the course). Consequently, for each course the number of 'flat' and 'steep' cambers were counted. Following Hobbs, Licka, & Polman (2011), a steep camber was defined by a slope of +/- 10°. Figure 6 below illustrates a flat and steep camber, taken as cross-sections of the course at different MPs.

Figure 6.2 Example of flat and steep Camber



Creation: at 50 meter intervals from the finish line the camber was measured (i.e. slope of the by cross section of the track) by taking the heights above sea level at 1 meter intervals across the track. A regression line for each cross-section of track was determined, based on the observed heights above sea level across the track at this point. In order to determine whether the camber at this MP was 'flat' or 'steep,' the slope of the regression line (coefficient β) at

MP_f was used to calculate the angle of inclination, as follows:

Angle of inclination at MP_f on the course = Tan^{-1} (β_f), where

$$\beta_f = \frac{\sum_{fx=1}^{n_x}(fx - f\bar{x})(fy - f\bar{y})}{\sum_{fx=1}^{n_x}(fx - f\bar{x})^2} \tag{6.1}$$

Where f = 1,2,3…n_f is the total number of MPs, fx represents the horizontal distance from the inside of the track and fy is the elevation above sea level (vertical distance) for the corresponding point at MP_f. fx = 1,2,3…n_x corresponds to readings at 1 meter intervals across the track, where n_x is the total number readings. The number of MPs in the final quarter with flat cambers, where the slope corresponds to angles less than 10° and greater than -10° ($\beta_f > -0.176$ and $\beta_f < 0.176$ based on Hobbs et al. (2011)), at each course were counted. This was used to calculate the variable, '*CAMBERS*' as follows:

$$CAMBERS_k = \frac{\sum_{f=1}^{\left(\frac{n_f}{4}\right)} \begin{cases} 1 \text{ if } \delta_f < 10° \text{ and } \delta_f > -10° \\ 0 \text{ else.} \end{cases}}{\frac{n_f}{4}} \tag{6.2}$$

Where 1,2…n_f is the total number of MPs at race course k and $\frac{n_f}{4}$ gives the number of MPs in the last quarter, where n_1 is closest to the finish line and n_f is furthest. δ_f is the angle of incline (decline) at MP_f. The number of MPs in the last quarter of each track ($\frac{n_f}{4}$) is the divisor, so that the length of each course is accounted for. This ensures that long and short tracks are treated equally, to nullify effects of distance within the variables.

Example: at one course with a total of 28 cambers, the β_f for the first 7 MPs closest to the finish line (= $\frac{n_f}{4}$) is measured and the angle of inclination is calculated by Tan^{-1} (βf) for MP_f = 1,2,3…7. For MP_1 for example,

fx = 1,2,3,4,5,6,7,8,9,10,11,12,13,14

fy = 86.58, 86.68, 86.79, 86.89, 87.0, 87.10, 87.20, 87.31, 87.41, 87.51, 87.62, 87.72, 87.83, 87.93

Therefore, $\beta f\ MP_1$ = 0.10 which corresponds to an angle of inclination of 5.71°.

Since $\beta f\ camber$ 1 is < 0.176 and > -0.176, $\beta_{f_{n_1}}$ is a flat camber.

The same process is performed for MP_f= 2,3…7.

This gives β_f for cambers 1,2,3….7 as [0.1, 0.09, 0.07, 0.14, 0.18,0.19,0.19]. There are 3 flat cambers out of a total of 7. Consequently:

$$CAMBERS_k = \frac{3}{7} = 0.4286$$

Variable: *DOWNSLOPE*

Definition: Cumulative drop in height above sea level for the final quarter of the course.

Concept: While the speed of human athletes increases on downhill sections, horse speed decreases. Self et al. (2012) show that a horse's top speed is slower on declines. This may be attributable to the anatomical simplicity of a horse's front legs which limit weight support and stability. Furthermore, differences in the length and muscle mass of a horse's front forelimbs will impose varying limits on the top speed that can be achieved by individual horses and will therefore be useful for identifying horses that can gallop faster on declines (Self et al. 2012). In addition, the rate at which the decline affects a horse's top speed is about -0.45 ms^{-1} .1% $gradient^{-1}$ (Self et al. 2012), meaning every 1% change in gradient slows the horse by -.045 $ms^{-1.}$

Creation: The final quarter of the course was divided into 50-meter intervals at which measurements were taken. At the each of these measuring points (MPs) the (vertical) elevation above sea level was recorded, at a point nearest to the inside of the track railing. Cumulative decline is measured by reduction in height from the one MP to the next MP. This variable is measured along the length of the track. By measuring the total decline in meters, the variable is able to capture whether there is a decline and its magnitude. The effect of distance at different courses (e.g., between a 2 mile race and a 6 furlong race) is accounted for by dividing by the total number of MPs in the final quarter of the course ($\frac{n_f}{4}$) , to create:

$$DOWNSLOPE_k = \frac{\sum_{f=1}^{\left(\frac{n_f}{4}\right)}(Df - Df_{+1})}{\frac{n_f}{4}} \tag{6.3}$$

where $1,2\ldots n_f$ is the total number of MPs at race venue k and $\frac{n_f}{4}$ gives the number of MPs in the last quarter (where n_1 is closest to the finish line and n_f is furthest from it). Df is the elevation above sea level at the inside rail of MP_f, and Df_{+1} is the height at MP_{f+1}.

Example: At one course with a distance of 1,250 meters there are a total of 24 MPs. The *DOWNSLOPE* is measured in the last quarter of the race. Therefore, since the readings start at the finish line, the difference in heights between successive MPs is calculated to work out the cumulative negative drop: decline from MP_1 to MP_2, from MP_2 to MP_3 ……$MP_{\frac{n_f}{4}-1}$ to $MP_{\frac{n_f}{4}}$.

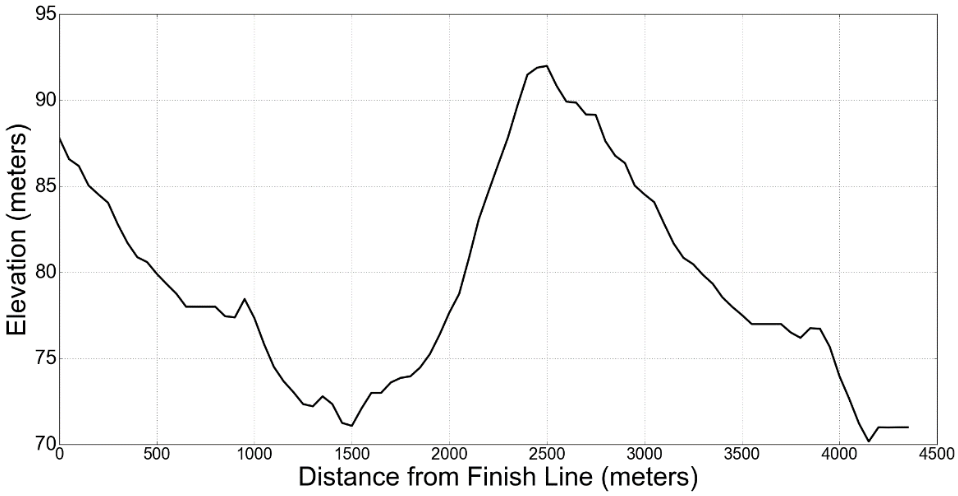
The associated Df for $MP1,2,3\ldots6$ is [12.17, 11.06, 10.56, 10.40, 9.18, 7.6]. This is equal to a drop of 1.11, 0.50, 0.15, 1.23 and 1.58 between the MPs, giving a total drop of 4.57 meters between the MPs in the last quarter of the course. Finally, divide by 6 (= $\frac{n_f}{4}$) to calculate the decline per MP, giving *DOWNSLOPE* = 0.76. As indicated above, this controls for race distance, otherwise a longer race (e.g. 2 miles) would have more MP_f readings than a short race (e.g. 5 furlong) and therefore would be susceptible to higher values of *DOWNSLOPE.* By dividing through by the number of MPs in the last quarter of varying length courses, the effects of distance are neutralized.

Variable: *UNDULATION*

Definition: A measure of the degree of undulation in the final furlong of a course.

Concept: Although all races are considered "flat" there is considerable change in the vertical height along the course, as shown in Figure 7 for one particular course. It is well known that it is easier to run on a flat (cf. undulating) surface (Minetti et al. 2002; Self et al. 2012). How much a track undulates will therefore have an effect on competitor performance, with some horses at an advantage, as they can reach higher top speeds on flat surfaces than other horses.

Figure 6.3 Illustration of a 'Flat' Course



Creation: Since most competitors use the inside of the track to minimize the total distance run in a race, the inside rail at each 50 meter interval provides a useful point at which to measure the elevation above sea level. The standard deviation of elevations above sea level at these MPs in the final furlong were calculated, as follows:

$$UNDULATION_k = \sqrt{\frac{1}{4}\sum_{f=1}^{4}(Df - \mu)^2}\ , \qquad (6.4)$$

$$\text{where } \mu = \frac{1}{4}\sum_{f=1}^{4} Df$$

where Df is the elevation above sea level on the inside rail at MP_f. Df at $MP_f, f = 1,2,3,4$ since the final furlong covers 201.68 meters, corresponding to 4 intervals of 50 meters.

Example: If the corresponding Df for the last 4 individual 50 meter stretches of a particular course is given by [17.30,16.57, 16.68, 16.33] then μ is equal to 16.72, giving $\sum_{f=1}^{4}(Df - \mu)^2 = 0.54$ and

$$UNDULATION_k = \sqrt{\frac{1}{4}(0.54)}$$

$$UNDULATION_k = 0.36$$

Variable: *WIDTH*

Definition: Average width of the track in the final furlong of the course.

Concept: The width of the track will have a significant effect on how close together the horses run. When horses are running side by side the effects of pack running might be heightened and the horses that are physically stronger may have an advantage over weaker ones. As Spence et al. (2012) discuss, there are a multitude of racing strategies that can be played in the running of a race that may suit a horse's individual characteristics, including 'front-runner', 'mid-pack' or 'chaser' roles. While these strategies have their own particular styles (e.g., a front-runner will run ahead of the pack), their ability to do so might be affected by the number of horses in a race and the track width as they try to assert themselves ahead of the pack. Equally, chasers may find it more difficult to overtake the pack when the track is narrow and the pack is dense. The width of the track then may affect performances of horses differentially.

Creation: *WIDTH* is defined as the average width in the final furlong of the course.

$$Width_k = \frac{\sum_{f=1}^{4} Width\ at\ MP_f}{4} \tag{6.5}$$

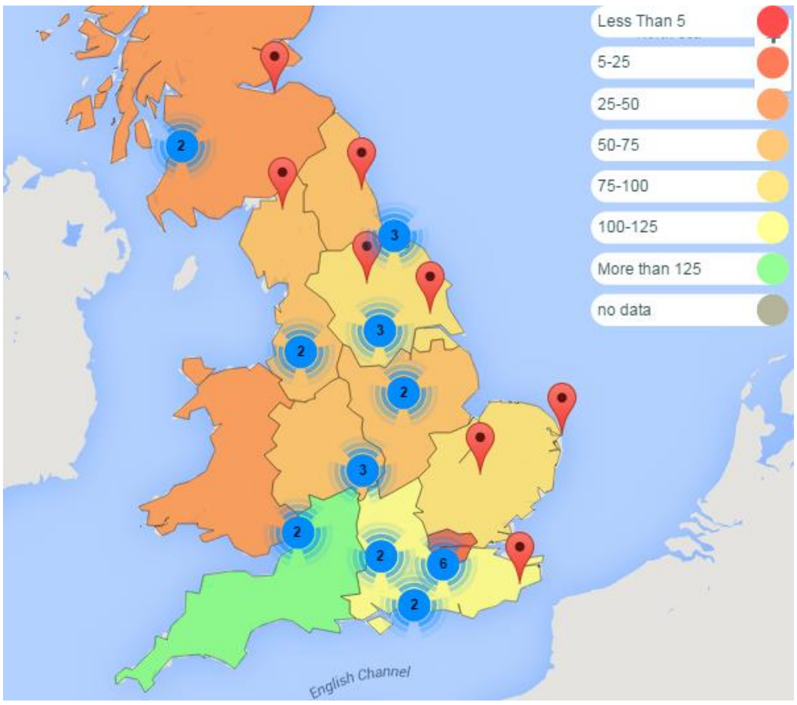
Example: Assuming that at a particular track the *Width* at each MP in the final furlong is given by [24, 23,23,25], then the average width in the final furlong is,

$$Width_k = \frac{95}{4}$$

Appendix C: Map of UK with locations of trainers and race courses

Figure 6.4 below shows a map of England and Wales colour coded according to how many trainers live in each region. The locations of the racecourses are also displayed on the map with markers, red markers showing one race course, and blue markers showing a cluster of courses with the number of venues noted in the blue circle. The map shows that the majority of trainers are located in the south and the majority of venues are in the middle and South. The average distance between trainers and the venues is 163.59 kilometres (KM). As such, the implications of this map are that trainers must travel a considerable distance to enter a venue and must take into account that races are on at a particular date and be strategic on which races to enter while weighing up the travelling costs.

Figure 6.4 Map of England and Wales with regions colour code according to number of trainers. Racetrack Venues are also marked with pointers on the map, with red markets signifying one venue and blue markers signifying a cluster with the number of venues.



Chapter 7 References

Abdellaoui, M., and Kemel, E. 2014. "Eliciting Prospect Thery When Consequences Are Measured in Time Units: 'Time Is Not Money,'" *Management Science* (60:7), pp. 1844–1859.

Agarwal, R., and Dhar, V. 2014. "Big Data, Data Science, and Analytics: The Opportunity and Challenge for IS Research," *Information Systems Research* (25:3), pp. 443–448.

Aggarwal, R., Kryscynski, D., Midha, V., Singh, H., Aggarwal, R., and Kryscynski, D. 2015. "Early to Adopt and Early to Discontinue: The Impact of Self-Perceived and Actual IT Knowledge on Technology Use Behaviors of End Users," *Information Systems Research* (26:1), pp. 127–144.

Akhtar, S., Faff, R., Oliver, B., and Subrahmanyam, A. 2013. "Stock salience and the asymmetric market effect of consumer sentiment news," *Journal of Banking and Finance* (37:11), Elsevier B.V., pp. 4488–4500 (doi: 10.1016/j.jbankfin.2013.07.032).

Alevy, J. E., Haigh, M. S., and List, J. a. 2007. "Information Cascades: Evidence from a Field," *Journal of Finance* (62:1), pp. 151–181.

Allen, D. E., McAleer, M., and Singh, A. K. 2017. "An entropy-based analysis of the relationship between the DOW JONES Index and the TRNA Sentiment series," *Applied Economics* (49:7), Routledge, pp. 677–692 (doi: 10.1080/00036846.2016.1203067).

Alvarado-Valencia, J., Barrero, L. H., Önkal, D., and Dennerlein, J. T. 2017. "Expertise, credibility of system forecasts and integration methods in judgmental demand forecasting," *International Journal of Forecasting* (33), pp. 298–313 (doi: 10.1016/j.ijforecast.2015.12.010).

Andersen, S., Fountain, J., Harrison, G. W., and Rutström, E. E. 2014. "Estimating subjective probabilities," *Journal of Risk and Uncertainty* (48:3), pp. 207–229 (doi: 10.1007/s11166-014-9194-z).

Andersson, P., Edman, J., and Ekman, M. 2005. "Predicting the World Cup 2002 in soccer: Performance and confidence of experts and non-experts," *International Journal of Forecasting* (21:3), pp. 565–576 (doi: 10.1016/j.ijforecast.2005.03.004).

Andersson, P., and Nilsson, H. 2015. "Do Bettors Correctly Perceive Odds? Three Studies of How Bettors Interpret Betting Odds as Probabilistic Information," *Journal of Behavioral Decision Making* (28:4), pp. 331–346 (doi: 10.1002/bdm.1851).

Antweiler, W., and Frank, M. Z. 2004. "Is all that talk just noise? The information content of internet stock message boards.," *The Journal of Finance* (59:3), pp. 1259–1294.

Antweiler, Werner, Frank, and Z., M. 2004. "Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards," *Journal of Finance* (59:3), pp. 1259–1294.

Baker, M., and Wurgler, J. 2007. "Investor Sentiment in the Stock Market," *Journal of Economic Perspectives* (21:2), pp. 129–151 (doi: 10.1257/jep.21.2.129).

Barber, B. M., and Odean, T. 2001. "Boys will be Boys: Gender , Overconfidence , and Common Stock Investment," *Quarterly journal of Economics*, pp. 261–292.

Barber, B. M., and Odean, T. 2002. "Online Investors: Do the Slow Die First?," *Review of Financial Studies* (15:2), pp. 455–487 (doi: 10.2139/ssrn.219242).

Barber, B. M., and Odean, T. 2008. "All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors," *Review of Financial Studies* (21:2), pp. 785–818 (doi: 10.1093/rfs/hhm079).

Barber, B., and Odean, T. 2003. "The Internet and the Investor," *Journal of Economic Perspectives* (15:1), pp. 41–54.

Bass, F. M. 1969. "A new product growth for model consumer durables," *Management Science*, pp. 215–227 (doi: 10.1287/mnsc.15.5.215).
Basu, P., and Nair, S. K. 2015. "Analyzing operational risk-reward trade-offs for start-ups," *European Journal of Operational Research* (247:2), Elsevier Ltd., pp. 596–609 (doi: 10.1016/j.ejor.2015.06.003).
Becker, K. H. 2015. "An Outlook on Behavioural OR – Three tasks, three pitfalls, one definition," *European Journal of Operational Research* (249:3), pp. 806–815 (doi: 10.1016/j.ejor.2015.09.055).
Bekiros, S., Nguyen, D. K., Sandoval Junior, L., and Uddin, G. S. 2017. "Information diffusion, cluster formation and entropy-based network dynamics in equity and commodity markets," *European Journal of Operational Research* (256:3), Elsevier B.V., pp. 945–961 (doi: 10.1016/j.ejor.2016.06.052).
Benbasat, I., and Zmud, R. W. 2003. "The Identity Crises within the IS Discipline: Defining and Communicating the Discipline's Core Properties," *MIS Quarterly* (27:2), pp. 533–556.
Benter, W. 1994. "Computer Based Horse Race Handicapping and Wagering Systems: A Report," in *Efficiency of Racetrack Betting Markets*, London: Academic Press, pp. 183–198.
Bernasco, W., Block, R., and Ruiter, S. 2013. "Go where the money is: Modeling street robbers' location choices," *Journal of Economic Geography* (13:1), pp. 119–143 (doi: 10.1093/jeg/lbs005).
Bharadwaj, A., El Sawy, O., Pavlou, P., and Venkatraman, N. 2013. "Digital Business Strategy: Toward a Next Generation of Insights," *MIS Quarterly* (37:2), pp. 471–482.
Bhattacharya, U., Holden, C. W., and Jacobsen, S. 2012. "Penny Wise, Dollar Foolish: Buy-Sell Imbalances On and Around Round Numbers," *Management science* (58:2), pp. 413–431 (doi: 10.1287/mnsc.1110.1364).
Boehmer, E., Jones, C. M., and Zhang, X. 2008. "Which shorts are informed?," *Journal of Finance* (63:2), pp. 491–527 (doi: 10.1111/j.1540-6261.2008.01324.x).
Bolger, F., and Wright, G. 2017. "Use of expert knowledge to anticipate the future: Issues, analysis and directions," *International Journal of Forecasting* (33:1), Elsevier B.V., pp. 230–243 (doi: 10.1016/j.ijforecast.2016.11.001).
Bolton, R. N., and Chapman, R. G. 1986. "Searching for Positive Returns at the Track: a Multinomial Logit Model for Handicapping Horse Races*," *Management Science* (32:8), pp. 1040–1060.
Brecher, S. L. 1980. *Beating the Races with a Computer*, Long Beach: Software Supply.
Bruce, A. C., Johnson, J. E. V, and Peirson, J. 2012. "Recreational versus professional bettors : Performance differences and efficiency implications," *Economics Letters* (114:2), Elsevier B.V., pp. 172–174 (doi: 10.1016/j.econlet.2011.10.014).
Bruce, A. C., Johnson, J., Peirson, J., and Yu, J. 2009. "An Examination of the Determinants of Biased Behaviour in a Market for State Contingent Claims," *Economica* (76), pp. 282–303 (doi: 10.1111/j.1468-0335.2008.00741.x).
Bruce, A., Marginson, D., Bruce, A., and Marginson, D. 2014. "Power , Not Fear : A Collusion-Based Account of Betting Market Inefficiency," *International Journal of the Economics of Business* (21:1), pp. 77–87 (doi: 10.1080/13571516.2013.782982).
Budescu, D. V., and Chen, E. 2014. "Identifying Expertise to Extract the Wisdom of Crowds," *Management Science* (61:2), pp. 267–280 (doi: 10.1287/mnsc.2014.1909).
Bulchand-Gidumal, J., and Melián-González, S. 2011. "Maximizing the positive influence of IT for improving organizational performance," *Journal of Strategic Information Systems* (20:4), pp. 461–478 (doi: 10.1016/j.jsis.2011.09.004).

Cachon, E. G. P. 2014. "Mobile Targeting," *Management Science* (60:7), pp. 1738–1756.
Camerer, C., and Johnson, E. 1991. "The process-performance paradox in expert judgment: How can experts know so much and predict so badly?," in *Toward a General Theory of Expertise: Prospects and Limits*A. Ericsson and J. Smith (eds.), Cambridge, UK: Cambridge University Press, pp. 195–217.
Cerroni, S., Notaro, S., and Shaw, W. D. 2012. "Eliciting and estimating valid subjective probabilities: An experimental investigation of the exchangeability method," *Journal of Economic Behavior and Organization* (84:1), pp. 201–215 (doi: 10.1016/j.jebo.2012.08.001).
Chae, H., Chang E. Koh, and Prybutok, V. R. 2014. "Information Technology Capability and Firm Performance: Contradictory Findings and Their Possible Causes," *MIS Quarterly* (38:1), pp. 305–326.
Chakrabarti, A., and Mitchell, W. 2013. "The Persistent Effect of Geographic Distance in Acquisition Target Selection," *Organization Science* (24:6), pp. 1805–1826.
Chang, R. M., Kauffman, R. J., and Kwon, Y. 2014. "Understanding the paradigm shift to computational social science in the presence of big data," *Decision Support Systems* (63), pp. 67–80 (doi: 10.1016/j.dss.2013.08.008).
Charles, A., Darné, O., and Kim, J. H. 2012. "Exchange-rate return predictability and the adaptive markets hypothesis: Evidence from major foreign exchange rates," *Journal of International Money and Finance* (31:6), pp. 1607–1626 (doi: 10.1016/j.jimonfin.2012.03.003).
Chen, H., Chiang, R., and Storey, V. C. 2012. "Business Intelligence and Analytics: From Big Data to Big Impact," *MIS Quarterly* (36:4), pp. 1–24.
Chen, H., De, P., Hu, Y. Y., and Hwang, B.-H. 2014. "Wisdom of Crowds: The Value of Stock Opinions Transmitted Through Social Media," *Review of Financial Studies* (27:5), pp. 1367–1403 (doi: 10.1093/rfs/hhu001).
Chen, W., and Zeng, D. 2016. "Modelling Fixed Odds Betting for Future Event Prediction," *MIS Quarterly* (Forthcomin), pp. 1–54.
Chordia, T., Roll, R., and Subrahmanyam, A. 2005. "Evidence on the speed of convergence to market efficiency," *Journal of Financial Economics* (76:2), pp. 271–292 (doi: 10.1016/j.jfineco.2004.06.004).
Clemons, E. K., Sawy, O. A. El, Weber, B., and Lucas, H. 2013. "Impactful Research on Transformational Information Technology: An Opportunity to Inform New Audiences," *MIS Quarterly* (37:2), pp. 371–382.
Comin, D., Hobijn, B., and Rovito, E. 2006. "Five Facts You Need to Know About Technology Diffusion," No. w11928, *National Bureau of Economic Research* (doi: 10.3386/w11928).
Constantiou, I. D., Lehrer, C., and Hess, T. 2014. "Changing information retrieval behaviours: an empirical investigation of users' cognitive processes in the choice of location-based services," *European Journal of Information Systems* (23:5), pp. 513–528 (doi: 10.1057/ejis.2014.12).
Costa-gomes, M., and Weizsäcker, G. 2008. "Stated Beliefs and Play in Normal Form Games," *Review of Economic Studies* (75:3), pp. 729–762 (doi: 10.1111/j.1467-937X.2008.00498.x).
Coussement, K., Benoit, D. F., and Antioco, M. 2015. "A Bayesian approach for incorporating expert opinions into decision support systems: A case study of online consumer-satisfaction detection," *Decision Support Systems* (79), pp. 24–32 (doi: 10.1016/j.dss.2015.07.006).
Das, S. R., and Chen, M. Y. 2007. "Yahoo! for Amazon: Sentiment Extraction from Small

Talk on the Web," *Management Science* (53:9), pp. 1375–1388 (doi: 10.1287/mnsc.1070.0704).
Dennis, A. R., and Carte, T. A. 1998. "Using Geographical Information Systems for Decision Making : Extending Cognitive Fit Theory to Map-Based Presentations," *Information Systems Research* (9:2), pp. 194–203.
Devaraj, S., and Kohli, R. 2003. "Performance Impacts of Information Technology: Is Actual Usage the Missing Link?," *Management Science* (49:3), pp. 273–289 (doi: 10.1287/mnsc.49.3.273.12736).
Drnevich, P. L., and Croson, D. C. 2013. "Information technology and business-level strategy: toward an integrated theoretical perspective," *MIS Quarterly* (37:2), pp. 483–509 (doi: 10.1016/j.sbspro.2013.06.099).
Ebert, J. E. J., and Prelec, D. 2007. "The Fragility of Time: Time-Insensitivity and Valuation of the Near and Far Future," *Management Science* (53:9), pp. 1423–1438 (doi: 10.1287/mnsc.1060.0671).
Ecken, P., and Pibernik, R. 2016. "Hit or Miss: What Leads Experts to Take Advice for Long-Term Judgments ? Hit or Miss : What Leads Experts to Take Advice for," *Management Science* (62:7), pp. 2002–2021.
Eickhoff, M., and Muntermann, J. 2016. "Stock analysts vs. the crowd: Mutual prediction and the drivers of crowd wisdom," *Information & Management* (doi: 10.1016/j.im.2016.03.008).
Engelberg, J. E., and Parsons, C. A. 2011. "The Causal Impact of Media in Financial Markets," *Journal of Finance* (66:1), pp. 67–97 (doi: 10.1111/j.1540-6261.2010.01626.x).
Engelberg, J. E., Reed, A. V., and Ringgenberg, M. C. 2012. "How are shorts informed? Short sellers, news, and information processing," *Journal of Financial Economics* (105:2), pp. 260–278 (doi: 10.1016/j.jfineco.2012.03.001).
Fama, E. F. 1970. "Efficient Capital Markets : A Review of Theory and Empirical Work," *The Journal of Finance* (25:2), pp. 383–417.
Fama, E. F., Fisher, L., Jensen, M. C., and Roll, R. 1969. "The Adjustment of Stock Prices to New Information," *International Economic Review* (10:1), pp. 1–21.
Faria, A., Fenn, P., and Bruce, A. 2003. "A Count Data Model of Technology Adoption," *Journal of Technology Transfer* (28:1), p. 63.
Festjens, A., Bruyneel, S., Diecidue, E., and Dewitte, S. 2015. "Time-based versus money-based decision making under risk: An experimental investigation," *Journal of Economic Psychology* (50), pp. 52–72 (doi: 10.1016/j.joep.2015.07.003).
Flaxman, M., and Vargas-Moreno, J. C. 2012. "Introduction: Science, Technology, and Engineering (Tools and Methods)," in *Restoring Lands - Coordinating Science, Politics and Action: Complexities of Climate and Governance*H. Karl, L. Scarlett, J. C. Vargas-Moreno, and M. Flaxman (eds.), London: Springer, pp. 21–27.
Fotheringham, S. 1988. "Consumer Store Choice and Choice Set Definition," *Marketing Science* (7:3), pp. 299–310.
Fox, C. R., and Tversky, A. 1998. "A Belief-Based Account of Decision Under Uncertainty," *Management Science* (44:7), pp. 879–895 (doi: 10.1287/mnsc.44.7.879).
Fraser-Mackenzie, P., Sung, M., and Johnson, J. E. V. 2015. "The prospect of a perfect ending: Loss aversion and the round-number bias," *Organizational Behavior and Human Decision Processes* (131:v), Elsevier Inc., pp. 67–80 (doi: 10.1016/j.obhdp.2015.08.004).
Frazzini, A. 2006. "The disposition effect and underreaction to news," *Journal of Finance* (61:4), pp. 2017–2046 (doi: 10.1111/j.1540-6261.2006.00896.x).

Friedman, M., and Savage, L. 1948. "The utility analysis of choice involving risk," *The journal of Political Economy* (56:4), pp. 279–304.
Fuentelsaz, L., Fuentelsaz, L., Gomez, J., Gomez, J., Polo, Y., and Polo, Y. 2003. "Intrafirm diffusion of new technologies: an empirical application," *Research Policy* (32), pp. 533–551 (doi: 10.1016/S0048-7333(02)00081-1).
Fuentelsaz, L., Gomez, J., and Palomas, S. 2012. "Production technologies and financial performance: The effect of uneven diffusion among competitors," *Research Policy* (41:2), pp. 401–413 (doi: 10.1016/j.respol.2011.09.006).
Ghose, A., Goldfarb, A., and Han, S. P. 2013. "How is the mobile internet different? Search costs and local activities," *Information Systems Research* (24:3), pp. 613–631 (doi: 10.1287/isre.1120.0453).
Goodchild, M. F., Guo, H., Annoni, A., Bian, L., de Bie, K., Campbell, F., Craglia, M., Ehlers, M., van Genderen, J., Jackson, D., Lewis, A. J., Pesaresi, M., Remetey-Fülöpp, G., Simpson, R., Skidmore, A., Wang, C., and Woodgate, P. 2012. "Next-generation Digital Earth.," *Proceedings of the National Academy of Sciences of the United States of America* (109:28), pp. 11088–94 (doi: 10.1073/pnas.1202383109).
Gottschlich, J., and Hinz, O. 2014. "A decision support system for stock investment recommendations using collective wisdom," *Decision Support Systems* (59:1), pp. 52–62 (doi: 10.1016/j.dss.2013.10.005).
Gregoriou, G. N. 2015. *The Handbook of High Frequency Trading*, Academic Press.
Groß-Klußmann, A., and Hautsch, N. 2011. "When machines read the news: Using automated text analytics to quantify high frequency news-implied market reactions," *Journal of Empirical Finance* (18:2), pp. 321–340 (doi: 10.1016/j.jempfin.2010.11.009).
Habjan, A., Andriopoulos, C., and Gotsi, M. 2014. "The role of GPS-enabled information in transforming operational decision making: an exploratory study," *European Journal of Information Systems* (23:4), pp. 481–502 (doi: 10.1057/ejis.2014.2).
Hagenau, M., Liebmann, M., and Neumann, D. 2013. "Automated news reading: Stock price prediction based on financial news using context-capturing features," *Decision Support Systems* (55:3), pp. 685–697 (doi: 10.1016/j.dss.2013.02.006).
Hall, B. H. 2004. "Innovation and Diffusion," No. w10212, *National Bureau of Economic Research*.
Hamalainen, R. P., Luoma, J., and Saarinen, E. 2013. "On the importance of behavioral operational research: The case of understanding and communicating about dynamic systems," *European Journal of Operational Research* (228:3), pp. 623–634 (doi: 10.1016/j.ejor.2013.02.001).
Hao, L., and Houser, D. 2011. "Belief elicitation in the presence of naïve respondents: An experimental study," *Journal of Risk and Uncertainty* (44:2), pp. 161–180 (doi: 10.1007/s11166-011-9133-1).
Hendershott, T., Livdan, D., and Schurhoff, N. 2015. "Are institutions informed about news?," *Journal of Financial Economics* (117:2), pp. 249–287 (doi: 10.1016/j.jfineco.2015.03.007).
Hirshleifer, D. A., Myers, J. N., Myers, L. A., and Teoh, S. H. 2008. "Do individual investors cause post-earnings announcement drift? Direct evidence from personal trades," *The Accounting Review* (83:6), pp. 1521–1550 (doi: 10.2308/accr.2008.83.6.1521).
Ho, H.-P., Chang, C.-T., and Ku, C.-Y. 2015. "House selection via the internet by considering homebuyers' risk attitudes with S-shaped utility functions," *European Journal of Operational Research* (241:1), pp. 188–201 (doi: 10.1016/j.ejor.2014.08.009).
Hobbs, S. J., Licka, T., and Polman, R. 2011. "The difference in kinematics of horses walking, trotting and cantering on a flat and banked 10 m circle," *Equine veterinary*

journal (43:6), pp. 686–94 (doi: 10.1111/j.2042-3306.2010.00334.x).
Hong, H., and Stein, J. C. 1999. “A Unified Theory of Underreaction , Momentum Trading , and Overreaction in Asset Markets,” *Journal of Finance* (54:6), pp. 2143–2184.
Huang, A., Zang, A., and Zheng, R. 2014. “Evidence on the Information Content of Text in Analyst Reports,” *The Accounting Review* (89:6), pp. 2151–2180 (doi: 10.3386/w19846).
Huberty, M. 2015. “Can we vote with our tweet? On the perennial difficulty of election forecasting with social media,” *International Journal of Forecasting* (31:3), Elsevier B.V., pp. 992–1007 (doi: 10.1016/j.ijforecast.2014.08.005).
Jarupathirun, S., and Zahedi, F. M. 2007. “Exploring the influence of perceptual factors in the success of web-based spatial DSS,” *Decision Support Systems* (43:3), pp. 933–951 (doi: 10.1016/j.dss.2005.05.024).
Johnson, J., Bruce, A., Yu, J., Johnson, J., Bruce, A., and Yu, J. 2010. “The ordinal efficiency of betting markets : an exploded logit approach The ordinal efficiency of betting markets : an exploded logit approach,” *Applied Economics* (42:29), pp. 3703–3709 (doi: 10.1080/00036840802314622).
Johnson, J. E. V., and Bruce, A. C. 2001. “Calibration of Subjective Probability Judgments in a Naturalistic Setting.,” *Organizational behavior and human decision processes* (85:2), pp. 265–290 (doi: 10.1006/obhd.2000.2949).
Johnson, J. E. V, and Bruce, A. C. 2001. “Calibration of Subjective Probability Judgments in a Naturalistic Setting,” (85:2), pp. 265–290 (doi: 10.1006/obhd.2000.2949).
Johnstone, D. 2016. “The Effect of Information on Uncertainty and the Cost of Capital,” *Contemporary Accounting Research* (33:2), pp. 752–774 (doi: 10.1111/1911-3846.12165).
Joseph, K., Babajide Wintoki, M., and Zhang, Z. 2011. “Forecasting abnormal stock returns and trading volume using investor sentiment: Evidence from online search,” *International Journal of Forecasting* (27:4), Elsevier B.V., pp. 1116–1127 (doi: 10.1016/j.ijforecast.2010.11.001).
Jun, S. P., Yoo, H. S., and Choi, S. 2017. “Ten years of research change using Google Trends: From the perspective of big data utilizations and applications,” *Technological Forecasting and Social Change* (Forthcomin), Elsevier (doi: 10.1016/j.techfore.2017.11.009).
Kahneman, D. 2003. “Maps of Bounded Rationality : Economicst Psychology for Behavioral,” *The American Economic Review* (93:5), pp. 1449–1475 (doi: 10.1257/000282803322655392).
Kaniel, R., Liu, S., Saar, G., and Titman, S. 2012. “Individual Investor Trading and Return Patterns around Earnings Announcements,” *Journal of Finance* (67:2), pp. 639–680 (doi: 10.1111/j.1540-6261.2012.01727.x).
Katsikopoulos, K. V. 2013. “Behavioral operations management: a blind spot and a research program,” *Journal of Supply Chain Management* (49:1), pp. 3–7.
Kauter, M. Van de, Breesch, D., and Hoste, V. 2015. “Fine-Grained Analysis of Explicit and Implicit Sentiment in Financial News Articles,” *Expert Systems with Applications* (42:11), pp. 4999–5010 (doi: 10.1016/j.eswa.2015.02.007).
Kearney, C., and Liu, S. 2014. “Textual sentiment in finance: A survey of methods and models,” *International Review of Financial Analysis* (33), pp. 171–185 (doi: 10.1016/j.irfa.2014.02.006).
Kelly, J. L. 1956. “A New Interpretation of Information Rate,” *Information Theory, IRE Transactions on 2.3*, pp. 185–189.
Kemel, E., and Paraschiv, C. 2013. “Prospect Theory for joint time and money consequences

in risk and ambiguity," *Transportation Research Part B: Methodological* (56), pp. 81–95 (doi: 10.1016/j.trb.2013.07.007).
Keyhole. 2003. "Keyhole: Industry Solutions," *Web.archive.org*, (available at http://web.archive.org/web/20030801173507/http://www.keyhole.com/industry_solutions/index.html; retrieved 19 July, 2013).
Kim, T., Hong, J., and Kang, P. 2015. "Box office forecasting using machine learning algorithms based on SNS data," *International Journal of Forecasting* (31:2), Elsevier B.V., pp. 364–390 (doi: 10.1016/j.ijforecast.2014.05.006).
Kocher, M. G., Pahlke, J., and Trautmann, S. T. 2013. "Tempus Fugit: Time Pressure in Risky Decisions.," *Management Science* (59:10), pp. 2380–2391 (doi: 10.1287/mnsc.2013.1711).
Kocher, M. G., and Sutter, M. 2006. "Time is money-Time pressure, incentives, and the quality of decision-making," *Journal of Economic Behavior and Organization* (61:3), pp. 375–392 (doi: 10.1016/j.jebo.2004.11.013).
Lawrence, M., Goodwin, P., O'Connor, M., and Önkal, D. 2006. "Judgmental forecasting: A review of progress over the last 25 years," *International Journal of Forecasting* (22:3), pp. 493–518 (doi: 10.1016/j.ijforecast.2006.03.007).
Leary, D. E. O. 2017. "Crowd performance in prediction of the World Cup 2014," *European Journal of Operational Research* (260:2), Elsevier B.V., pp. 715–724 (doi: 10.1016/j.ejor.2016.12.043).
Leclerc, F., Schmitt, B. H., and Dube, L. 1995. "Waiting Time and Decision Making: Is Time like Money?," *Journal of Consumer Research* (22:1), p. 110 (doi: 10.1086/209439).
Legerstee, R., and Franses, P. H. 2014. "Do experts' SKU forecasts improve after feedback?," *Journal of Forecasting* (33:1), pp. 69–79 (doi: 10.1002/for.2274).
Leitner, J., and Leopold-Wildburger, U. 2011. "Experiments on forecasting behavior with several sources of information - A review of the literature," *European Journal of Operational Research* (213:3), pp. 459–469 (doi: 10.1016/j.ejor.2011.01.006).
Leitner, J., and Schmidt, R. 2006. "A systematic comparison of professional exchange rate forecasts with the judgemental forecasts of novices," *Central European Journal of Operations Research* (14:1), pp. 87–102.
Lessmann, S., Sung, M.-C., and Johnson, J. E. V. 2011. "Towards a methodology for measuring the true degree of efficiency in a speculative market," *Journal of the Operational Research Society* (62:12), pp. 2120–2132 (doi: 10.1057/jors.2010.192).
Lessmann, S., Sung, M.-C., Johnson, J. E. V., and Ma, T. 2012. "A new methodology for generating and combining statistical forecasting models to enhance competitive event prediction," *European Journal of Operational Research* (218:1), pp. 163–174 (doi: 10.1016/j.ejor.2011.10.032).
Lessmann, S., and Voß, S. 2017. "Car resale price forecasting: The impact of regression method, private information, and heterogeneity on forecast accuracy," *International Journal of Forecasting* (33:4), Elsevier B.V., pp. 864–877 (doi: 10.1016/j.ijforecast.2017.04.003).
Li, Q., Wang, T., Gong, Q., Chen, Y., Lin, Z., and Song, S. 2014. "Media-aware quantitative trading based on public Web information," *Decision Support Systems* (61), pp. 93–105 (doi: 10.1016/j.dss.2014.01.013).
Lillo, F., Miccichè, S., Tumminello, M., Piilo, J., and Mantegna, R. N. 2015. "How News Affects the Trading Behavior of Different Categories of Investors in a Financial Market," *Quantitative Finance* (15:2), pp. 213–229 (doi: 10.2139/ssrn.2109337).
Lim, K.-P., and Brooks, R. 2011. "The Evolution of Stock Market Efficiency Over Time: a Survey of the Empirical Literature," *Journal of Economic Surveys* (25:1), pp. 69–108

(doi: 10.1111/j.1467-6419.2009.00611.x).

Liu, X., and Ye, Q. 2016. "The different impacts of news-driven and self-initiated search volume on stock prices," *Information & Management* (53:8), pp. 997–1005 (doi: 10.1016/j.im.2016.05.009).

Liu, Y., Lee, Y., and Chen, A. N. K. 2011. "Evaluating the effects of task-individual-technology fit in multi-DSS models context: A two-phase view," *Decision Support Systems* (51:3), pp. 688–700 (doi: 10.1016/j.dss.2011.03.009).

Lo, A. 2004. "The adaptive markets hypothesis: Market efficiency from an evolutionary perspective," *Journal of Portfolio Management* (30:5), pp. 15–29.

Lo, A. 2005. "Reconciling efficient markets with behavioral finance: the adaptive markets hypothesis," *Journal of Investment Consulting* (7:2), pp. 21–44.

Lo, A. 2012. "Adaptive Markets and the New World Order," *Financial Analysts Journal* (68:2), pp. 18–29.

Ma, T., Tang, L., McGroarty, F., Sung, M.-C., and Johnson, J. E. . 2016. "Time is money: Costing the impact of duration misperception in market prices," *European Journal of Operational Research*, Elsevier B.V. (doi: 10.1016/j.ejor.2016.04.044).

Maclean, L. C., Thorp, E. O., and Ziemba, W. T. 2010. "Long-term capital growth: the good and bad properties of the Kelly and fractional Kelly capital growth criteria," *Quantitative Finance* (10:7), pp. 681–687 (doi: 10.1080/14697688.2010.506108).

Manela, A. 2014. "The value of diffusing information," *Journal of Financial Economics* (111:1), Elsevier, pp. 181–199 (doi: 10.1016/j.jfineco.2013.10.007).

Mannes, A. 2009. "Are We Wise About the Wisdom of Crowds? The Use of Group Judgments in Belief Revision," *Management Science* (55:8), pp. 1267–1279 (doi: 10.1287/mnsc.1090.1031).

Manski, C. F. 2004. "Measuring Expectations," *Econometrica* (72:5), pp. 1329–1376.

Martens, D., and Provost, F. 2014. "Explaining Data-Driven Document Classifications," *MIS Quarterly* (38:1), pp. 73–99.

McFadden, D. L. 1974. "Conditional Logit Analysis of Qualitative Choice Behavior," in *Frontiers in Econometrics*P. Zarembka (ed.), New York: Academic Press, pp. 105–142.

McFadden, D. L. 1978. "Modeling The Choice of Residental Location," *Spatial Interaction Theory and Planning Models*, pp. 75–96 (available at http://eml.berkeley.edu/reprints/mcfadden/location.pdf).

McKenzie, C. R. M., Liersch, M. J., and Yaniv, I. 2008. "Overconfidence in interval estimates: What does expertise buy you?," *Organizational Behavior and Human Decision Processes* (107:2), pp. 179–191 (doi: 10.1016/j.obhdp.2008.02.007).

McKinsey Global Institute. 2011. "Big data: The next frontier for innovation, competition, and productivity," (available at http://www.mckinsey.com/~/media/mckinsey/dotcom/insights and pubs/mgi/research/technology and innovation/big data/mgi_big_data_full_report.ashx; retrieved December 7, 2015).

Meade, N., and Islam, T. 2006. "Modelling and forecasting the diffusion of innovation – A 25-year review," *International Journal of Forecasting* (22:3), pp. 519–545 (doi: 10.1016/j.ijforecast.2006.01.005).

Meeks, W. L., and Dasgupta, S. 2004. "Geospatial information utility: an estimation of the relevance of geospatial information to users," *Decision Support Systems* (38:1), pp. 47–63 (doi: 10.1016/S0167-9236(03)00076-9).

Melville, N., Gurbaxani, V., and Kraemer, K. 2007. "The productivity impact of information technology across competitive regimes: The role of industry concentration and dynamism," *Decision Support Systems* (43:1), pp. 229–242 (doi:

10.1016/j.dss.2006.09.009).
Mennecke, B. E., Crossland, M. D., and Killingsworth, B. L. 2000. "Is a Map More than a Picture? The Role of SDSS Technology, Subject Characteristics, and Problem Complexity on Map Reading and Problem Solving," *MIS Quarterly* (24:4), pp. 601–629.
Minetti, A. E., Moia, C., Roi, G. S., Susta, D., and Ferretti, G. 2002. "Energy cost of walking and running at extreme uphill and downhill slopes.," *Journal of applied physiology* (93:3), pp. 1039–46 (doi: 10.1152/japplphysiol.01177.2001).
Mollick, E. R., and Nanda, R. 2016. "Wisdom or Madness? Comparing Crowds with Expert Evaluation in Funding the Arts," *Management science* (62:6), pp. 1533–1553 (doi: 10.2139/ssrn.2443114).
Nadkarni, A., and Vesset, D. 2015. "Worldwide Big Data Technology and Services Forecast, 2015–2019 - 259532," *IDC*, pp. 1–11 (available at http://www.idc.com/getdoc.jsp?containerId=259532; retrieved October 5, 2016).
Nagarajan, M., Shaw, D., and Albores, P. 2012. "Disseminating a warning message to evacuate: A simulation study of the behaviour of neighbours," *European Journal of Operational Research* (220:3), Elsevier B.V., pp. 810–819 (doi: 10.1016/j.ejor.2012.02.026).
Nardo, M., Petracco-Giudici, M., and Naltsidis, M. 2015. "Walking down wall street with a tablet: A survey of stock market predictions using the web," *Journal of Economic Surveys* (30:2), pp. 356–369 (doi: 10.1111/joes.12102).
Nassirtoussi, A. K., Wah, T. Y., Aghabozorgi, S. R., and Ling, D. N. C. 2014. "Text Mining for Market Prediction: A Systematic Review," *Expert Systems with Applications* (41:16), pp. 7653–7670 (doi: 10.1016/j.eswa.2014.06.009).
O'Keefe, R. M. 2016. "Experimental behavioural research in operational research: What we know and what we might come to know," *European Journal of Operational Research* (249:3), pp. 899–907 (doi: 10.1016/j.ejor.2015.09.027).
Oikonomidis, A., Bruce, A. C., and Johnson, J. E. V. 2015. "Does transparency imply efficiency ? The case of the European soccer betting market," *Economics Letters* (128), Elsevier B.V., pp. 59–61 (doi: 10.1016/j.econlet.2015.01.015).
Önkal, D., Yates, J. F., Simga-Mugan, C., and Öztin, Ş. 2003. "Professional vs. amateur judgment accuracy: The case of foreign exchange rates," *Organizational Behavior and Human Decision Processes* (91:2), pp. 169–185 (doi: 10.1016/S0749-5978(03)00058-X).
Papagiannidis, S., Gebka, B., Gertner, D., and Stahl, F. 2015. "Diffusion of web technologies and practices: A longitudinal study," *Technological Forecasting and Social Change* (96), pp. 308–321 (doi: 10.1016/j.techfore.2015.04.011).
Peeters, G., and Czapinski, J. 1990. "Positive-Negative Asymmetry in Evaluations: The Distinction Between Affective and Informational Negativity Effects," *European Review of Social Psychology* (1:1), pp. 33–60 (doi: 10.1080/14792779108401856).
Peress, J. 2004. "Wealth, information acquisition, and portfolio choice," *Review of Financial Studies* (17:3), pp. 879–914 (doi: 10.1093/rfs/hhg056).
Peress, J. 2014. "The Media and the Diffusion of Information in Financial Markets: Evidence from Newspaper Strikes," *Journal of Finance* (69:5), pp. 2007–2043 (doi: 10.1111/jofi.12179).
Petropoulos, F., Fildes, R., and Goodwin, P. 2016. "Do 'big losses' in judgmental adjustments to statistical forecasts affect experts' behaviour?," *European Journal of Operational Research* (249:3), pp. 842–852 (doi: 10.1016/j.ejor.2015.06.002).
Pick, J. B., Turetken, O., Deokar, A. V., and Sarkar, A. 2017. "Location analytics and decision support: Reflections on recent advancements, a research framework, and the path

ahead," *Decision Support Systems* (99), pp. 1–8 (doi: 10.1016/j.dss.2017.05.016).
Rabus, B., Eineder, M., Roth, A., and Bamler, R. 2003. "The shuttle radar topography mission - A new class of digital elevation models acquired by spaceborne radar," *ISPRS Journal of Photogrammetry and Remote Sensing* (57:4), pp. 241–262 (doi: 10.1016/S0924-2716(02)00124-7).
Reboredo, J. C., Rivera-Castro, M. a., Miranda, J. G. V., and García-Rubio, R. 2013. "How fast do stock prices adjust to market efficiency? Evidence from a detrended fluctuation analysis," *Physica A: Statistical Mechanics and its Applications* (392:7), Elsevier B.V., pp. 1631–1637 (doi: 10.1016/j.physa.2012.11.038).
Ren, F., and Dewan, S. 2015. "Industry-Level Analysis of Information Technology Return and Risk: What Explains the Variation?," *Journal of Management Information Systems* (32:2), pp. 71–103 (doi: 10.1080/07421222.2015.1063281).
Rogers, E. M. 1995. *Diffusion of innovations*, New York, USA (doi: citeulike-article-id:126680).
Roscoe, R. D., Grebitus, C., O'Brian, J., Johnson, A. C., and Kula, I. 2016. "Online information search and decision making: Effects of web search stance," *Computers in Human Behavior* (56), pp. 103–118 (doi: 10.1016/j.chb.2015.11.028).
Rubin, A., and Rubin, E. 2010. "Informed Investors and the Internet," *Journal of Business Finance & Accounting* (37:7–8), pp. 841–865 (doi: 10.1111/j.1468-5957.2010.02187.x).
Sabherwal, R., and Jeyaraj, A. 2015. "Information Technology Impacts on Firm Performance: An Extension of Kohli and Devaraj (2003)," *MIS Quarterly* (39:4), pp. 1–30.
Sabherwal, S., Sarkar, S. K., and Zhang, Y. 2011. "Do internet stock message boards influence trading? Evidence from heavily discussed stocks with no fundamental news," *Journal of Business Finance and Accounting* (38:9–10), pp. 1209–1237 (doi: 10.1111/j.1468-5957.2011.02258.x).
Schlag, K., Tremewan, J., and Weele, J. 2013. "A Penny for Your Thoughts: A Survey of Methods for Eliciting Beliefs," *Experimental Economics*, pp. 1–34.
Schneider, M. J., and Gupta, S. 2016. "Forecasting sales of new and existing products using consumer reviews: A random projections approach," *International Journal of Forecasting* (32:2), Elsevier B.V., pp. 243–256 (doi: 10.1016/j.ijforecast.2015.08.005).
Schotter, A., and Trevino, I. 2014. "Belief Elicitation in the Lab," *Annual Review of Economics* (6), pp. 103–128 (doi: 10.1146/annurev-economics-080213-040927).
Schryen, G. 2012. "Revisiting IS business value research: what we already know, what we still need to know, and how we can get there," *European Journal of Information Systems* (22:2), pp. 139–169 (doi: 10.1057/ejis.2012.45).
Schwager, S., and Rothermund, K. 2013. "Motivation and affective processing biases in risky decision making: A counter-regulation account," *Journal of Economic Psychology* (38), Elsevier B.V., pp. 111–126 (doi: 10.1016/j.joep.2012.08.005).
Science and Technology Committee. 2016. "The big data dilemma," *House of Commons Science and Technology Committee Fourth Special Report of Session*, pp. 1–56 (available at http://www.publications.parliament.uk/pa/cm201516/cmselect/cmsctech/468/468.pdf; retrieved February 12, 2016).
Self, Z. T., Spence, A. J., and Wilson, A. M. 2012. "Speed and incline during Thoroughbred horse racing : racehorse speed supports a metabolic power constraint to incline running but not to decline running," *Journal of Applied Physiology* (113), pp. 602–607 (doi: 10.1152/japplphysiol.00560.2011).
She, S., Lu, Q., and Ma, C. 2012. "A probability–time & space trade-off model in environmental risk perception," *Journal of Risk Research* (15:2), pp. 223–234 (doi:

10.1080/13669877.2011.634515).
Shen, M., Carswell, M., Santhanam, R., and Bailey, K. 2012. “Emergency management information systems: Could decision makers be supported in choosing display formats?,” *Decision Support Systems* (52:2), pp. 318–330 (doi: 10.1016/j.dss.2011.08.008).
Sheppard, S. R. J., and Cizek, P. 2009. “The ethics of Google Earth: crossing thresholds from spatial data to landscape visualisation.,” *Journal of environmental management* (90:6), pp. 2102–17 (doi: 10.1016/j.jenvman.2007.09.012).
Shontell, A. 2012. “Here’s How Long It Took 15 Hot Startups To Get 1,000,000 Users,” *Business Insider*, (available at http://www.businessinsider.com/one-million-users-startups-2012-1?op=1&IR=T; retrieved 13 February, 2014).
Simon, H. 1955. “A behavioral model of rational choice,” *The Quarterly Journal of Economics* (69:1), pp. 99–118 (doi: 10.1080/02724980343000242).
Smales, L. A. 2014. “News sentiment in the gold futures market,” *Journal of Banking and Finance* (49), pp. 275–286 (doi: 10.1016/j.jbankfin.2014.09.006).
Spann, M., and Skiera, B. 2009. “Sports forecasting: A comparison of the forecast accuracy of prediction markets, betting odds and tipsters,” *Journal of Forecasting* (28:1), pp. 55–72 (doi: 10.1002/for.1091).
Speier, C. 2006. “The influence of information presentation formats on complex task decision-making performance,” *International Journal of Human Computer Studies* (64:11), pp. 1115–1131 (doi: 10.1016/j.ijhcs.2006.06.007).
Spence, A. J., Thurman, A. S., Maher, M. J., and Wilson, A. M. 2012. “Speed , pacing strategy and aerodynamic drafting in Thoroughbred horse racing Subject collections Speed , pacing strategy and aerodynamic drafting in Thoroughbred horse racing,” *Biology Letters* (8:4), pp. 678–681.
Sprenger, T., Sandner, P., Tumasjan, A., and Welpe, I. 2014. “News or Noise? Using Twitter to Identify and Understand Company-specific News Flow,” *Journal of Business Finance & Accounting* (41:7–8), pp. 791–830 (doi: 10.1111/jbfa.12086).
Stieglitz, S., and Dang-Xuan, L. 2013. “Emotions and Information Diffusion in Social Media—Sentiment of Microblogs and Sharing Behavior,” *Journal of Management Information Systems* (29:4), pp. 217–248 (doi: 10.2753/MIS0742-1222290408).
Stillwell, D. J., and Tunney, R. J. 2012. “Individuals’ insight into intrapersonal externalities,” *Judgment and Decision Making* (7:4), pp. 390–401.
Stoneman, P. 1995. *The Handbook of Economics of Innocation and Technological Change Blackwell*, Cambridge MA: Blackwell.
Sul, H. K., Dennis, A. R., and Yuan, L. I. 2016. “Trading on Twitter: using social media sentiment to predict stock returns,” *Decision Sciences*, pp. 1–35 (doi: 10.1111/deci.12229).
Sundararajan, A., Provost, F., Oestreicher-Singer, G., and Aral, S. 2013. “Information in Digital, Economic and Social Networks,” *Information Systems Research* (24:4), pp. 883–905.
Sung, M.-C., and Johnson, J. E. V. 2007. “Comparing the Effectiveness of One- and Two-Step Conditional Logit Models for Predicting Outcomes in a Speculative Market,” *Journal of Prediction Markets* (44:1), pp. 43–59.
Sung, M.-C., Johnson, J. E. V., and Peirson, J. 2012. “Discovering a Profitable Trading Strategy in an Apparently Efficient Market : Exploiting the Actions of Less Informed Traders in Speculative Markets,” *Journal of Business Finance and Accounting* (39:October), pp. 1131–1159 (doi: 10.1111/j.1468-5957.2012.02300.x).
Sung, M.-C., McDonald, D. C. J., and Johnson, J. E. V. 2016. “Probabilistic forecasting with

discrete choice models: Evaluating predictions with pseudo-coefficients of determination," *European Journal of Operational Research* (248:3), Elsevier Ltd., pp. 1021–1030 (doi: 10.1016/j.ejor.2015.08.068).

Sung, M., and Johnson, J. E. V. 2010. "Revealing Weak-Form Inefficiency in a Market for State Contingent Claims: The Importance of Market Ecology, Modelling Procedures and Investment Strategies," *Economica* (77:305), pp. 128–147 (doi: 10.1111/j.1468-0335.2008.00716.x).

Surowiecki, J. 2005. *The wisdom of crowds*, New York: Anchor.

Tambe, P., Hitt, L. M., and Brynjolfsson, E. 2012. "The Extroverted Firm: How External Information Practices Affect Innovation and Productivity The Extroverted Firm : How External Information Practices Affect Innovation and Productivity," *Management Science* (58:5), pp. 8443–859.

Tetlock, P. C. 2007. "Giving Content to Investor Sentiment: The Role of Media in the Stock Market," *Journal of Finance* (62:3), pp. 1139–1168.

Tetlock, P. C., Saar-Tsechansky, M., and MacSkassy, S. 2008. "More than words: Quantifying language to measure firms' fundamentals," *Journal of Finance* (63:3), pp. 1437–1467 (doi: 10.1111/j.1540-6261.2008.01362.x).

Tomlinson, R. F. 1968. "A Geographic Information System for Regional Planning," in *Symposium on Land Evaluation, Commonwealth Scientific and Industrial Research Organization*G. Stewart (ed.), Melbourne: MacMillan (doi: 10.5026/jgeography.78.45).

Torngren, G., and Montgomery, H. 2004. "Worse than chance? Performance and confidence among professionals and laypeople in the stock market," *Journal of Behavioral Finance* (5:3), pp. 148–153 (doi: 10.1207/s15427579jpfm0503).

Trautmann, S. T., and Kuilen, G. 2014. "Belief Elicitation: A Horse Race among Truth Serums," *The Economic Journal* (doi: 10.1111/ecoj.12160).

Trope, Y., and Liberman, N. 2010. "Construal-Level Theory of Psychological Distance," *Pyschoological Review* (116:2), pp. 440–463 (doi: 10.1037/a0018963.Construal-Level).

Venkatesh, V., Thong, J., and Xu, X. 2012. "Consumer Acceptance and User of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology," *MIS Quarterly* (36:1), pp. 157–178.

Vessey, I. 1991. "Cognitive Fit: A Theory-Based Analysis of the Graphs versus Tables Literature," *Decision Sciences* (22:2), pp. 219–240.

Vessey, I., and Galletta, D. 1991. "Cognitive Fit: An Empirical Study of Information Acquisition," *Information Systems Research* (2:1), pp. 63–84.

Viscusi, W. K., and Evans, W. N. 2006. "Behavioral probabilities," *Journal of Risk and Uncertainty* (32:1), pp. 5–15 (doi: 10.1007/s10797-006-6663-6).

Wallenius, J., Dyer, J. S., Fishburn, P. C., Steuer, R. E., Zionts, S., and Deb, K. 2008. "Multiple Criteria Decision Making, Multiattribute Utility Theory: Recent Accomplishments and What Lies Ahead," *Management Science* (54:7), pp. 1336–1349 (doi: 10.1287/mnsc.1070.0838).

Wang, H., Xu, Z., Fujita, H., and Liu, S. 2016. "Towards felicitous decision making: An overview on challenges and trends of Big Data," *Information Sciences* (367–368), Elsevier Inc., pp. 747–765 (doi: 10.1016/j.ins.2016.07.007).

Werner, C., Bedford, T., Cooke, R. M., Hanea, A. M., and Morales-Nápoles, O. 2017. "Expert judgement for dependence in probabilistic modelling: A systematic literature review and future research directions," *European Journal of Operational Research* (258:3), Elsevier B.V., pp. 801–819 (doi: 10.1016/j.ejor.2016.10.018).

Van Wesep, E. D. 2016. "The Quality of Expertise," *Management Science* (62:10), pp. 2937–2951 (doi: 10.2139/ssrn.2257995).

White, L. 2016. "Behavioural operational research: Towards a framework for understanding behaviour in OR interventions," *European Journal of Operational Research* (249:3), pp. 827–841 (doi: http://dx.doi.org/10.1016/j.ejor.2015.07.032).

Xu, S., and Zhang, X. 2013. "Impact of Wikipedia on Market Information Environment: Evidence on Management Disclosure and Investor Reaction," *MIS Quarterly* (37:4), pp. 1043-A10.

Yang, S. Y., Li, T., and Van Heck, E. 2015. "Information transparency in prediction markets," *Decision Support Systems* (78), Elsevier B.V., pp. 67–79 (doi: 10.1016/j.dss.2015.05.009).

Yang, S. Y., Liu, A., Chen, J., and Hawkes, A. G. 2017. "Applications of a Multivariate Hawkes Process to Joint Modeling of Sentiment and Market Return Events," *Quantitative Finance*, Routledge, pp. 1–16 (doi: 10.2139/ssrn.2954079).

Yu, L., Zhao, Y., Tang, L., and Yang, Z. 2018. "Online big data-driven oil consumption forecasting with Google trends," *International Journal of Forecasting* (Forthcomin), Elsevier B.V. (doi: 10.1016/j.ijforecast.2017.11.005).

Yuan, Y. 2015. "Market-wide attention, trading, and stock returns," *Journal of Financial Economics* (116:3), pp. 548–564 (doi: 10.1016/j.jfineco.2015.03.006).

Zhang, J. L., Härdle, W. K., Chen, C. Y., and Bommes, E. 2016. "Distillation of News Flow into Analysis of Stock Reactions," *Journal of Business & Economic Statistics* (34:4), pp. 547–563 (doi: 10.1080/07350015.2015.1110525).

Zhang, W., Li, X., Shen, D., and Teglio, A. 2016. "Daily happiness and stock returns: Some international evidence," *Physica A: Statistical Mechanics and its Applications* (460), pp. 201–209 (doi: 10.1016/j.physa.2016.05.026).

Zhang, X., and Zhang, L. 2015. "How Does the Internet Affect the Financial Market? An Equilibrium Model of Internet-Facilitated Feedback Trading," *MIS Quarterly* (39:1), pp. 17–38.

Zhang, X., Zhang, Y., Wang, S., Yao, Y., Fang, B., and Yu, P. S. 2018. "Improving stock market prediction via heterogeneous information fusion," *Knowledge-Based Systems* (0), Elsevier B.V., pp. 1–12 (doi: 10.1016/j.knosys.2017.12.025).

Zhu, B., and Chen, H. 2005. "Using 3D interfaces to facilitate the spatial knowledge retrieval: a geo-referenced knowledge repository system," *Decision Support Systems* (40:2), pp. 167–182 (doi: 10.1016/j.dss.2004.01.007).

Zipf, G. K. 1949. *Human Behavior and the Principle of Least Effort. An Introduction to Human Ecology*, Cambridge, MA: Addison-Wesley.