To what extent can new web-based technology improve forecasts? Assessing the economic value of information derived from Virtual Globes and its rate of diffusion in a financial market

Citation for published version:
Green, L, Sung, M-C, Ma, T & Johnson, JEV 2019, 'To what extent can new web-based technology improve forecasts? Assessing the economic value of information derived from Virtual Globes and its rate of diffusion in a financial market', *European Journal of Operational Research*, vol. 278, no. 1, pp. 226-239. https://doi.org/10.1016/j.ejor.2019.04.014

Digital Object Identifier (DOI):
10.1016/j.ejor.2019.04.014

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
European Journal of Operational Research

To what extent can new web-based technology improve forecasts? Assessing the economic value of information derived from Virtual Globes and its rate of diffusion in a financial market.

Green, L.*, Sung, M.*, Ma, T.*[1], Johnson, J.E.V.*

*Centre for Risk Research, Southampton Business School, University of Southampton
UK SO17 1BJ
[1] Corresponding author

Email: {lg2g08, ms9, tiejun.ma, jej} @soton.ac.uk

ABSTRACT

As the rate of information availability increases, the ability to use web-based technology to improve forecasting becomes increasingly important. We examine Virtual Globe technology and show how the arrival of unprecedented types of web-based information enhances the ability to forecast and can lead to significant, measurable economic benefits. Specifically, we use market prices in a betting market over an eighteen-year period to examine how new elevation data from Virtual Globes (VG) enabled improved forecasting decisions and we explore how this information diffused through the betting market. The results demonstrate how short-lived, profitable opportunities arise from the arrival of novel information, and the speed at which markets adapt over time to account fully for new data.

Keywords: Forecasting, Adaptive markets, Virtual Globes, market efficiency, longitudinal study

To what extent can new web-based technology improve forecasts? Assessing the economic value of information derived from Virtual Globes and its rate of diffusion in a financial market.

Introduction

The information environment surrounding financial markets has rapidly developed in recent years due to advances in Information Technology (IT). In particular, the internet has made available rich and diverse sources of information that can be used to improve forecasts that can impact markets (Yu et al. 2018). Data sources such as Wikipedia, Twitter, Facebook, Google and YouTube offer ubiquitous and rapidly changing data. The impact of this ever-increasing volume of data has been the focus of many studies in multiple domains, including business analytics (Wang et al. 2016).

The rapid adoption of web-based technology and the shift from the proprietary information systems of the 1990s to more standardized, affordable, and simpler to acquire web technologies has offered new opportunities to analyze the relationship between new technology and forecasting accuracy (Chae et al. 2014). Information availability has enabled superior forecasting in a wide range of fields. For example, social networking data, twitter data and product reviews have been used to forecast box office statistics (Kim et al. 2015), elections (Huberty 2015) and sales of new and existing products (Schneider and Gupta 2016), respectively.

Recent developments in technology have led to a rapid growth in location analytics, spatial analysis and geographic information systems (Pick et al. 2017). For example, tools such as Google maps provide unprecedented access to spatial information. However, there is a need for more research exploring to what extent Google mapping technology has improved forecasting accuracy (Jun et al. 2017; Habjan et al. 2014). Consequently, we examine “virtual worlds” and the extent to which they can create and capture value, by providing information that can enhance forecasting capabilities (Drnevich and Croson 2013).

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, more empirical evidence is needed to demonstrate how VGs improve decision-making in general and forecasting in particular (Liu et al. 2011). To help fulfil this objective, we measure the speed and extent to which VG information diffuses through a financial market, highlighting the changing economic value of this data.

To achieve this, we avoid using shares prices as a measure of value, as new information has an unpredictable effect on share prices (Johnstone 2016). Rather, we explore a setting where the three-dimensional geospatial information provided by VGs offers the prospect of more effective forecasts that can be used to extract economic value from a financial market. In particular, we explore how VG information related to the configuration and topography of racetracks can be used to improve forecasts of racehorses' winning probabilities and we measure the economic value of these forecasts by assessing the profits that can be achieved by employing betting strategies based on these forecasts.

The Efficient Market Hypothesis (EMH: Fama 1970) states that new information should be immediately incorporated into market prices, such that no consistent positive returns are possible. There is a large body of research supporting the EMH (e.g., Xu and Zhang, 2013). The semi-strong form of the EMH, suggests that markets respond quickly and swiftly to new publicly available information, such that abnormal profits cannot be secured by trading on this information. Early research supported this view (Hillmer and Yu 1979). However, recent evidence exposes problems with the EMH, as profitable opportunities have been shown to arise at times in certain markets (Fischer and Krauss 2018; Ng et al. 2017). The adaptive market hypothesis (AMH), proposed by Lo (2004) offers a theoretical framework to explain how markets can evolve over time, adapting to new environmental and technological conditions. This theory suggests that profitable opportunities can arise in periods when the value of new information is not fully appreciated by market participants. We use the AMH as a basis for a hypothesis related to the changing economic value of forecasts arising from using VG data. We show that substantial profits could be made by employing VG data when it was first released. However, as its value for improving forecasts became more widely understood, the profitable opportunities decreased and were eventually eliminated.

Croxson and Reade (2013) show that football markets react swiftly and fully to new information, such that abnormal profits cannot be earnt around the event of a goal in a football match. However, transaction costs and other costs of acquiring new information often prevent information being instantly discounted in market prices (e.g., Chordia et al., 2005). For example, Mills and Salaga (2018) showed that information associated with umpire related decisions could be used to make a return of nearly 10% on bets related to the total runs scored in major league baseball. Similarly, Hwang and Kim (2015) identified a misunderstanding of extreme events in volleyball, which led to a four-year period when participants failed to update fully their posterior probabilities in the light of this information. Therefore, there is some contrasting evidence of the speed of market convergence, but the weight of evidence points to markets undergoing a period of diffusion during which new information becomes discounted in market prices.

It is clear from the discussion in section 2.2, that there is a substantial literature which demonstrates that betting markets are highly efficient, prices (odds) fully accounting for a wide range of publicly available information, including that derived from sophisticated mathematical prediction models incorporating a variety of fundamental information (e.g., Spann, 2003; Smith et al., 2006 ,2009; Tziralis and Tatsiopoulos, 2007; Franck et al., 2010; Croxson and Reade, 2012; Štrumbelj and Vračar,, 2012; Baker and McHale, 2013; Baboota and Khaur, 2018). In the light of this literature, and the fact that the configuration and topography of racetracks have been readily observable by bettors and available to them on large-scale topographic maps for many years, one might expect that bettors would fully discount this information in betting odds. However, despite this, we believe that the more accessible nature of VG-based topographic information may have led bettors to appreciate its value in helping them predict winning probabilities. Consequently, the first contribution of this paper is to explore the degree to which VG related information can improve winning probability forecasts beyond those that can be discerned from, what are widely regarded, as efficient market prices.

Second, we show, via a longitudinal study, the speed at which the VG information diffuses through the market as participants learn to use the data to produce better market forecasts. The vast majority of research exploring betting market efficiency examines to what extent particular types of information are accounted for in market prices *at a specific point in time*. Clearly, this may fail to capture the degree to which market participants adapt their behavior to use new information over time. Our study seeks to overcome this limitation.

Third, we estimate the economic value of using geospatial information by using out-of-sample forecasts as the basis of betting strategies. We show that the profits achievable from employing VG information vary through time in the manner predicted by the AMH. In particular, when the VG information first appeared, it was possible to make substantial profits based on forecasts that employed this information. However, these profits disappeared as market participants learned to employ the data, market odds fully discounting the information.

The remainder of the paper is organized as follows: In Section 2 we motivate and develop our hypotheses. In section 3, we describe the data and methodology used to test the hypotheses. We present the results in Section 4. In Section 5, we discuss the results and the implications of these findings in terms of information diffusion and adaptive markets and draw some conclusions.

2 Virtual Globes and market efficiency: Hypotheses

In this section, we develop the hypotheses to be tested. To motivate these hypotheses, in Section 2.1 we outline the nature and history of VGs. This serves to indicate the nature of the data that they provide and when it became available to different groups of the population. In Section 2.2, we outline the broad themes that emerge in the literature that explores the degree to which betting odds discount publicly available information. We draw on this literature to develop our first hypothesis. In Section 2.3, we discuss the nature of the AMH and use this to develop our second hypothesis.

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. Such data have been freely available over the internet since the release of GLOBE in 1999. A series of DEM product releases have followed. The most significant resulted from the Shuttle Radar Topography Mission (STRM), a collaborative effort by multinational agencies including NASA. On the 11th January 2003 the resulting SRTM dataset 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 (e.g., academics who developed Geographical Information Systems (GIS)). GIS scientists used specialist software (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 products 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 and free, but had less analytical functionality and were widely used by the public (Goodchild et al. 2012).

SRTM 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.[1] The advent of VGs and particularly KeyHole, has made elevation data freely and easily accessible to the public (Sheppard and Cizek 2009), thereby improving the public's potential geospatial decision-making capabilities.

Table 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

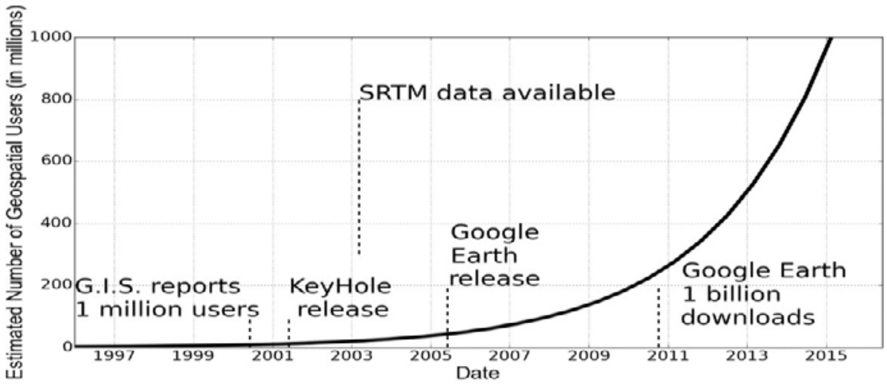
[1] See Appendix A for further information, historical context and media coverage related to VGs.

	regional planning.”
1970	Burgeoning of GIS software.
November 1997	Microsoft released Encarta Virtual Globe 98 (offline CD version); one of the first virtual globes allowing users to navigate and experience the world in 3-D.
December 1999	ArcGIS 8.0 released: Professional software providing a command line interface for (experienced) users to create, combine and analyse statistical mapping information.
2000	An estimated 1 Million Geospatial users.
June 2001	Keyhole Inc. released Earth Viewer with elevation data.
February 2003	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	An estimated 100 million users of Google Earth.
2011	Google Earth reaches 1 billion downloads.
2016	An estimated 1 billion users of Google Earth.

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. For example, Keyhole, released in 2001, had 250,000 consumers in 2003 (Keyhole 2003) and Google Earth, released in 2005, had 100 million users a year later (Sheppard and Cizek 2009); dwarfing (by 100 times) the launch year user numbers of the most popular social media technologies, Facebook and Twitter (Shontell 2012). The context and chronology of important events related to VGs are summarised in Table 1.

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

Figure 1: Estimated growth in Geospatial users



2.2 The use of publicly available information in horserace betting markets

The principal theoretical framework employed when investigating the degree to which horserace-betting market prices (odds) account for information is the EMH (Fama, 1970). Adapted to horserace betting markets, this theory postulates that the bettors' combined best estimate of a horse's winning chance (represented by the odds) should correspond to its true winning probability. In particular, since bettors have strong economic motives to appropriately use all available information when estimating each horse's probability of winning, the EMH would predict that odds fully account for all publicly available information related to past and current performance-related information, and should fully adjust for the arrival of new information. As Sauer (2005, p.416) observes in reviewing betting market literature, 'it is the benchmark result that prices [in betting markets] are approximately efficient that yields insights to which economics should be proud to lay claim. For example, odds at the racetrack imbed subjective estimates of the probability of winning that are (a) quite close to their empirical counterparts, (b) similarly close to those obtained using sophisticated statistical methods, and (c) very difficult to exploit on a systematic basis.' Sauer points out that there is considerable evidence in the literature related to sport betting markets that odds provide well-calibrated forecasts of each contestant's winning probability (e.g., Johnson and Bruce, 2001; Smith, 2003; Deschamps and Gergaud, 2008; Lessmann et al., 2010). A wide range of studies have demonstrated that predictions based on final odds are better than forecasts based on many other methods, including lay persons' aggregate fast and frugal predictions (Serwe and Frings, 2006; Spann and Skiera, 2009), experts' predictions (Forrest and Simmons, 2000) and statistical models using fundamental variables (Croxson and Reade, 2012; Štrumbelj and Vračar,, 2012; Baker and McHale, 2013; Baboota and Khaur, 2018)). For example, Baboota and Khaur (2018) found that odds were better calibrated than a state-of-the art artificial intelligence algorithm they had developed to predict football results. Similarly, Štrumbelj and Vračar, (2012) developed several sophisticated models to predict basketball matches, including a Markov model (using the arrival of new information in play-by-play data), and Baker and McHale (2013) developed a model to predict exact scores at NFL, but both sets of authors found that these models' predictions could not out-perform odds. Equally, Croxson and Reade (2012 found that odds accurately account for the arrival of new information (in-play) in soccer matches.

Many of these betting market efficiency studies test whether simple, single factor pieces of information (e.g., newspaper and expert tipsters' predictions: Vaughan Williams, 1999, 2000; Smith, 2003) are accounted for in odds (for review, see Sung and Johnson, 2008). Since, manipulating these types of information for making probability estimates is relatively easy, it is perhaps not surprising that odds account for the information. However, some more recent studies suggest that complexity can inhibit the degree to which information is discounted in odds. This complexity may relate to computational complexity, such as the complex manner in which information needs to be manipulated and combined, or to the complexity of the modelling required to effectively use the information. For example, whilst it has been shown in numerous studies that individual expert tipsters information is fully accounted for in the odds in horse racing (e.g., Sung et al., 2005), it has been shown that *combining* different expert tipsters' predictions, can produce small profits in football betting markets (Brown and Reade 2019). Similarly, Rosenbloom (2003) demonstrated that a variable combining (in a regression model) ten previous speed ratings of a horse, was not fully discounted in odds. Equally, some models that *combine* a range of variables related to horse or jockey performance, capturing non-linear relationships, can produce probability estimates that are not fully discounted in odds (e.g., Lessmann, 2009, 2010, 2012). Equally, some variables based on publicly available information derived from complex modelling procedures, are also not fully accounted for in odds (e.g., horses' recovery times/duration between wins: Ma et al., 2016; starting position: Johnson et al., 2010a; performance against common opponents: Knottenbelt, 2012). Importantly, with an increase in complexity, bettors may not be able to scrutinise fully all important relationships between variables and performance outcomes and feedback may become more ambiguous (Brehmer and Allard, 1991). Consequently, this may impair bettors' ability to effectively develop decision models that are logically correct, thus making it difficult to understand and predict performance outcomes.

We suspect that VG enabled elevation data can be useful for developing forecasts of the winning probabilities of horses, by accounting for horses' preferences for different racetrack topography. However, in order to estimate these winning probabilities, it is necessary to undertake substantial, complex data gathering and manipulation (as outlined in Section 3.2). Consequently, despite the fact that odds have been shown to be efficient predictors of winning probabilities we suspect that VG enabled elevation data related to racetrack topography will not be fully discounted in odds. To explore this view, we test the following hypothesis:

Hypothesis 1 (H1): VG information can be used to develop forecasts of horse performance that outperform market forecasts (odds).

2.3 Adaptive markets

Some economists (e.g., Lim and Brooks 2011) have questioned the assumption of the EMH, that market prices reflect all new information instantly. Equally, psychologists have suggested that the EMH's human rationality assumptions do not accord with individual behavior (e.g., Goodwin, et al., 2010; Frino et al., 2008). In order to accommodate these criticisms, a revised version of the EMH has been proposed, namely the Adaptive Market Hypothesis (AMH) (Lo 2004, 2005, 2012). This theory proposes that markets evolve as participants learn to adapt to new information and conditions.

The AMH assumes that market efficiency is related to the environment and to the market participants' adaptability to emerging technology (Lo 2005). Importantly, advocates of the AMH argue that when environmental conditions change, market participants require a period of learning new heuristics that better suit the new environmental conditions (Lo 2004). Furthermore, that competition drives the rate at which individuals adapt to the new informational environment, such that "prices reflect as much information as dictated by the combination of environmental conditions and the number and nature" of market participants (Lo 2005, p.31). This implies that market prices do not always perfectly account for all available information, and they can vary in terms of how much of this information is discounted (Lo 2012).

Results of some recent financial market studies support the view that markets go through cyclical change as they evolve and market participants learn new mechanisms of trading. Prices therefore vary, at different times failing to reflect, partially reflecting or fully reflecting different pieces of publicly available information (e.g., Charles, et al., 2012; Doyle and Chen 2013; Urquhart and Hudson, 2013; Urquhart and McGroarty, 2014; Urquhart et al., 2015; Fischer and Krauss, 2018). For example, Urquhart (2017) showed, using rolling subsamples that prices of three precious metals go through periods of predictability and unpredictability in line with the AMH, and Al-khazali and Mirzaei (2017) showed that the AMH better describes the way in which Islamic stock markets adapt to changing environmental conditions.

The majority of studies demonstrating support for the AMH employ longitudinal analysis (e.g.; Kim et al., 2011; Ghazani, 2014; Urquhart and McGroarty, 2014; Urquhart, 2017; Fischer and

Krauss, 2018) with rolling subsamples. This approach enables tests of market efficiency through time, with the same experimental setup in each period.

By contrast, the vast majority of studies that have examined the degree that odds account for publicly available information use a single dataset covering a limited period. Lim and Brooks (2011) argue that the dynamic nature of efficiency cannot be captured in these arbitrarily chosen sub-samples, and they highlight the need for research that examines the speed at which market prices adjust to new information. Clearly, drawing conclusions about the degree to which market prices account for particular information using a sample drawn from a specific period, fails to capture the speed at which market prices adjust to new information (Khuntia and Pattanayak, 2018). Hwang and Kim (2018) highlight this problem in the betting market literature, arguing that most studies only employ small subsets of data and, therefore, fail to explore the degree to which odds account for a particular piece of information over a sufficiently long period (see Sung et al., 2005 for review).

Online betting markets have facilitated the development of in-play betting, whereby bets are placed on the result of an event whilst it is taking place. There is evidence from these markets that bettors react quickly and appropriately to information related to events which can be anticipated and for which historical data to develop appropriate models of behavior is available. For example, there is evidence that in-play odds correspond with predictions made by mathematical models based on information that arrives during the course of an event (e.g., a goal being scored in soccer or a point being won in tennis: McHale and Morton, 2011; Croxson and Reade 2014). These studies confirm that bettors have the ability to learn to handle even complex information, provided its arrival can be anticipated. Equally, Johnson et al. (2010a) demonstrated that bettors can learn, via feedback, to cope with even complex dynamic information.

In summary, most betting market studies examine the degree to which odds account for particular types of information at a specific moment. Consequently, they do not examine variations in the degree to which odds account for information through time. In-play studies provide some evidence that bettors can react to dynamic information. However, there is a dearth of studies examining how bettors react to *new* information. This is surprising, since the AMH predicts that market inefficiencies can arise when new information becomes available, but learning, competition and evolutionary forces, eventually force prices to levels that account for the new information.

Prior to the arrival of VG information, bettors could assess the topographical features of racetracks based only on visual-based assessments or from topographical maps. Geospatial technology offers more easily accessible information that is more readily capable of statistical manipulation. In addition, its accessibility is likely to have increased bettors awareness of its potential value for predicting winning probabilities. The AMH predicts that it may take some time for bettors to make more informed judgments using novel information, such as VG elevation information. We explore this prediction by testing the following hypothesis:

Hypothesis 2 (H2): Betting market participants take time to learn how to fully account for VG elevation information in odds.

3 Methodology

3.1 The advantages of exploring speed of diffusion of VG information in betting markets

We selected the horserace betting market as a suitable environment to study the speed of diffusion of VG information, as it offers several advantages. In particular, bettors make forecasting decisions regarding horses' winning chances. These are more likely to be accurate if they account for the extent that gradients and cambers at different racetracks afford advantages to certain horses. Bettors have always been able to observe the configuration and topography of racetracks and more detailed data has been available to them via large-scale topographic maps. However, we suspect that bettors only fully discounted this information in their betting decisions when it became readily available via VGs.

Horserace betting markets also afford the advantage that odds provide a means of assessing the degree to which bettors employ VG data. Odds represent the betting public's combined forecast of the horses' winning probabilities and it is possible to derive winning probability estimates based on market odds alone and from models which combine both odds and variables derived from VG elevation data. Comparing these two sets of probabilities with many race results, enables us to assess their relative calibration. Equally, we can compare the performance of betting strategies based on these two sets of probabilities; enabling, the economic value (in terms of betting profit) of VG data to be determined. Furthermore, each race is a separate decision task and there are several thousand races each year. Consequently, the outcomes of races over a period can provide an historical record of the changing degree to which individuals incorporate VG information into their betting decisions.

Finally, betting markets have the advantage that they capture the 'real world' decision-making of individuals who have strong (financial) incentives to make accurate judgements and where the bettors' process of learning how to use the VG-based information is facilitated by immediate feedback (i.e. the winner of each race is announced immediately after the event). These are important advantages over research on technology diffusions conducted in laboratory settings, where participants are offered either no incentives or artificial incentives of limited value and are often not provided with the immediate feedback that can occur in real-world settings.

3.2 Data

In order to test the hypotheses, we secured data relating to 18 years of horserace data (1997 to 2014, inclusive), including 75,750 races (incorporating 76,406 different horses) run at all 34 UK racetracks. This period encompasses the dates of release of VGs. In particular, Geocontext (www.geocontext.com), a web service that became available from 2010, provides access to SRTM elevation data from the Google elevation API. This data enabled us to develop topographical 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, we also measured the camber 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.[2] This procedure produced a dataset of topographies related to 300 different 'courses.'

Spence et al. (2012) found that horseraces 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 topography in the final section of the race was crucial. Consequently, topography 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 certain horses may gain from topographical features, we consulted the equine literature related to 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 (e.g. weight, leg length, body weight distribution, running action etc.) might give them an advantage on flat, upward or downward sloping tracks or on undulating tracks.

In brief, we created four variables that capture different topographical 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. This is important because speed is influenced by adaptation to curved motion (Hobbs et al. 2011). We determined the number of MPs in the last quarter of the race where the positive or negative gradient was less than 10 degrees (based on Hobbs et al.'s (2011) definition), and defined these as 'flat'). We then determined the proportion of MPs with flat cambers in relation to the total number of MPs in the last quarter:

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

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

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 $camber_f$. We assessed the effect of cambers by using the proportion of flat MPs divided by the total number of MPs in the last quarter of each track ($\frac{n_f}{4}$)) in order to nullify the 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*). We measure *DOWNSLOPE* because downward sloping tracks reduce a horse's maximum speed differentially, depending on the horse's physique (Self et al. 2012). Therefore, we determined the cumulative decline in the last quarter of the race for track k, as follows:

$$DOWNSLOPE_k = \frac{\sum_{f=1}^{\left(\frac{n_f}{4}\right)} (Df - Df_{+1})}{\frac{n_f}{4}} \quad (2)$$

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*). We include this because the degree of undulation differentially affects 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:

$$UNDULATION_k = \sqrt{\frac{1}{4}\sum_{f=1}^{4}(Df - \mu)^2} \quad (3)$$

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) The average width of the track in the last furlong (*WIDTH*). This variable is included because track width affects horses differently, depending on their running styles (Spence et al. 2012). Even though the rails can be moved during a meeting, 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, we calculate the average width for track k, as follows:

$$Width_k = \frac{\sum_{f=1}^{4} W_f}{4} \quad (4)$$

There is 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 pricing in horseracing betting markets (Bruce et al. 2009). In order to ensure that we had a reasonably comprehensive picture of how VG elevation data can be employed by bettors to improve on their winning probability estimates, we tested a number of other features related to racetrack topography that horseracing experts suggested might differentially influence the winning chances of horses. These included, for example, the curve of the track and the change in height from start to finish. We also examined a variety of interactions between topographical features and weather, ground conditions and distance of the race. However, none of these features significantly affected winning probabilities. Clearly, the nature and complexity of the horserace betting markets make it difficult to be categorical that we have captured every feature of topography that might influence winning probabilities. However, the final set of four topography-related variables defined above were those most strongly recommended by racing experts and they provided reliable and consistent variables that could be clearly shown to affect a horse's winning probability (further details of these variables are provided in online Appendix B).

3.3 Statistical procedures

3.3.1 Preference variables

In order to test whether the four variables associated with racetrack topography (*CAMBERS, DOWNSLOPE, UNDULATION* and *WIDTH*) confer advantages/disadvantages to horse *i* in race *j*, we examined horse *i*'s past performances, measured by its normalized finish position (Brecher 1980) in race *k*, (where *k*=1,…, *j*-1) as follows:

$$NFP_{ik} = 0.5 - \frac{Ordinal\ Finishing\ Position_{ik} - 1}{Number\ of\ runners_{k} - 1} \tag{5}$$

where the first and last placed horses are assigned values of 0.5 and -0.5, respectively.

The NFP_{ik} provides a means of assessing how well the horse performed compared to the other horses in race *k*. For example, if a horse had performed better in races where the cambers were generally flat, then this might indicate that the horse has an advantage in those conditions. In order to find such patterns, we used Benter's (1994) preference variable technique. Specifically, for each horse, a linear regression was 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.

This technique is performed for each of the four topographical variables and for each race in each horse's history. This approach provides four 'preference variables' (*prefCAMBERS, prefDOWNSLOPE, prefUNDULATION and prefWIDTH*), which take values in a particular race depending on the regression line for each geospatial variable.

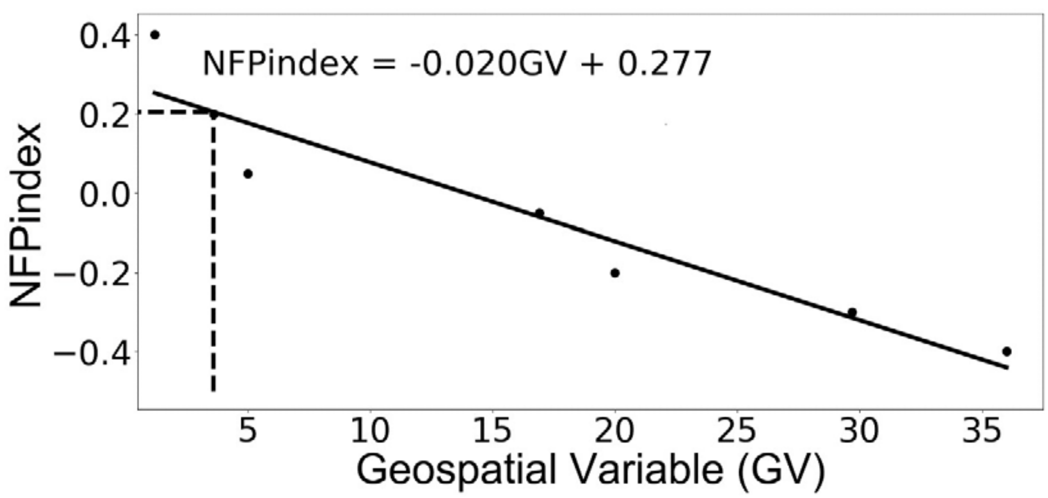
3.3.2 Two stage conditional logit model

We then developed a model with horse *i*'s predicted NFP in race *j* as the dependent variable and the four preference variables for horse *i* in race *j*, together with the horse's average NFP across all races up to today's race (i.e. across races 1,….., *j*-1)), $averageNFP_i$, as independent variables:

$$predictedNFP_{ij} = \beta_1 prefCAMBER_{ij} + \beta_2 prefDOWNSLOPE_{ij} + \beta_3\, prefUNDULATION_{ij} + \beta_4 prefWIDTH_{ij} + \beta_5 averageNFP_i \tag{6}$$

The model is estimated for each 3-year period from 1997 to 2014 using OLS regression.

Figure 2: Preference Variable Technique



The $predictedNFP_{ij}$, derived from (6) are used as independent variables in a conditional logit model (CL, McFadden, 1974) to estimate winning probabilities for each horse. 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 i in race j:

$$P_{ij}{}^{e} = P_1, P_{2^e j}, \ldots P_{n^e jj} \tag{7}$$

where n_j is the number of horses in race j. To achieve this, we define a 'winningness index' W_{ij}, for horse i in race j, as follows:

$$W_{ij} = \sum^{m}{}_{k=1} \beta(k) x_{ij}(k) + {}_{ij} \tag{8}$$

where $\beta(k)$ (for $k = 1,2\ldots m$) 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 ${}_{ij}$ 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:

$$p_{ij} = \frac{\exp(\sum_{k=1}^{m} \beta_k x_{ij}(k))}{\sum_{i=1}^{n_j} \exp(\sum_{k=1}^{m} \beta_k x_{ij}(k))} \tag{9}$$

We use the winning probabilities estimated by (6) ($predictedNFP_{ij}$), across races in the 3-year period (e.g., 1997-1999), as the sole inputs for a CL model, in order to estimate (9). This estimation is conducted by employing all races run in the same 3-year period (1997-1999). This model, referred to as the $Geospatial_{simple}$ model predicts the winning probability of each horse i in race j, as follows:

$$p_{ij}^{s} = \frac{\exp(\sum_{k=1}^{m} \beta_k predictedNFP_{ij}(k))}{\sum_{i=1}^{n_j} \exp(\sum_{k=1}^{m} \beta_k predictedNFP_{ij}(k))} \tag{10}$$

In order to test if the betting market efficiently accounts for the VG elevation data, we explore if a model to predict winning probabilities based on these VG-elevation variables together with odds, outperforms a model involving odds-implied probabilities alone. This mirrors the methodology employed in the vast majority of papers exploring the degree to bettors account for various types of information (e.g., Sung and Johnson, 2008; Lessmann et al., 2009, 2010, 2012; Štrumbelj and Vračar, 2012; Baker and McHale, 2013; Ma et al., 2016; Baboota and Kaur, 2018). As a form of robustness check, we also explore whether VG elevation data is already accounted for in variables related to horse and jockey performance which have been shown are not fully accounted for in odds (these results are presented and discussed in Appendix C)

For the analysis reported in the main body of the paper, the predicted probabilities from the $Geospatial_{simple}$ model, are used as independent variables, alongside probabilities implied by the market odds for horse i in race j, $p_{ij}{}^{m}$, in a second CL model (second stage), referred to as the $Geospatial_{Full}$ model, in order to estimate winning probabilities $p_{ij}{}^{full}$, as follows:

$$p_{ij}^{full} = \frac{\exp(\alpha \ln\left(p_{ij}^{s}\right) + \gamma \ln(p_{ij}^{m}))}{\sum_{i=1}^{n_j} \exp\left(\alpha \ ln\ (p_{ij}^{s}\right) + \gamma \ln(p_{ij}^{m}))} \tag{11}$$

The odds implied probability for horse *i*, $p_{ij}{}^{m}$, is given by 1/(X_i-1), where odds for horse i , X_i , mean that a winning £1 bet produces a profit of £X-1 (suggesting that market participants as a whole believe the horse has a 1/(X_i-1) chance of winning). The natural log of $p_{ij}{}^{s}$ and $p_{ij}{}^{m}$ is used in the CL model, as this transformation provides a better fit to winning probabilities (Benter, 1994).

The parameters α and γ in (11) are estimated using maximum likelihood procedures, again using the three-year sample of races that was used to develop model (10). This ensured that as large a sample as possible was used to estimate the coefficients of the two variables, based on market odds-implied probabilities $p_{ij}{}^{m}$ and the predicted probabilities from the $Geospatial_{simple}$ model (p_{ij}^{s}). This in-sample approach may have the effect of slightly overestimating the effect of the p_{ij}^{s} in the $Geospatial_{Full}$ model. Consequently, to ensure that the VG elevation data was definitely not fully accounted for in odds, we conduct *out-of-sample* betting simulations (see Section 3.3.3, below) when testing H2. These simulations measure directly the economic benefit that would have accrued to bettors had they appropriately used VG elevation data in the years after the VG data became available.

The two-stage procedure adopted here has been shown to provide more accurate winning probabilities than those derived by simply incorporating market 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.

3.3.3 Testing hypotheses.

To test H1, that VG information can be used to develop forecasts of horse performance that outperform market forecasts (odds), we first examine whether the preference variables do contain valuable predictive information. To achieve this, we test whether model 10 (($Geospatial_{simple}$; incorporating geospatial preference variables) can provide winning probability estimates which improve on those based on a random selection of the winning horse. This is achieved by testing whether the variable coefficients β in model 10, estimated for each three-year period, 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, is also calculated to show how much variation in the win probabilities are explained by model 10 (Bolton and Chapman 1986):

$$R^2 = 1 - \frac{LL(model)}{LL(random)} \tag{12}$$

where $LL(model)$ is the log-likelihood of the $Geospatial_{simple}$ is given by:

$$LL\ (model) = \sum_{j=1}^{N} \sum_{i=1}^{n_j} y_{ij} \ln_{p_{ij}} \tag{13}$$

and $LL(random)$ is the log-likelihood of the random choice model, where each horse is assigned an equal winning probability, determined as follows:

$$LL(random) = \sum_{j=1}^{N} \ln(\frac{1}{n_j}) \tag{14}$$

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

The results show that the geospatial variables can help to predict winning probabilities (see below). Consequently, to complete our testing of H1, we explore *to what extent* the information contained in these variables is not discounted by the betting public in market odds. This is achieved by examining to what extent the estimates of winning probabilities based on the CL model incorporating both market odds-implied probabilities and the preference variable probabilities (model 11: $Geospatial_{Full}$), are more accurate than winning probabilities derived from a CL model simply including a Ln transformation of market odds implied probabilities (we refer to this latter model as the *Odds* Model). This follows the methodology employed in most semi-strong form efficiency studies conducted in betting markets (e.g., Bolton and Chapman, 1986; Spann and Skiera, 2009; Lessmann et al., 2009, 2010, 2012; Johnson et al., 2009, 2010; Štrumbelj and Vračar, 2012; Baker and mchale, 2013; Croxson and Reade, 2014; Baboota and Kaur, 2018). The *Odds* Model is used as a baseline because, as discussed in Section 2.2, odds have been shown to account for a wide variety of information related to a horse's chance of winning (Sung et al., 2005). As a result, odds produce very well calibrated winning probability forecasts (e.g., Johnson and Bruce, 2001; Lessmann et al., 2010).

The Log Likelihood Ratio (LLR) test is employed to compare the maximum likelihood values (LL) of the $Geospatial_{Full}$ and the $Odds$ (null model) models, using the following statistic:

$$LLR = -2\frac{LL(\ Geospatial_{Full})}{LL(Odds)} \tag{15}$$

LLR is χ^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 is used to test the null hypothesis that these models predict winning probabilities with the same degree of accuracy.

This process is repeated via a series of sliding windows of consecutive 3-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 the VG data over time. In particular, this approach provides us with a means of determining, for each 3-year period, whether $predictedNFP_{ij}$ provides information that can help predict the probability of horse i winning race j (i.e. whether horse i does have a 'preference' for certain racetrack topography). It also enables us to assess the extent to which $Geospatial_{Full}$ provides information to help predict winning probabilities, beyond that accounted for by the betting public in the odds.

To test H2, that betting market participants take time to learn how to fully account for VG elevation information in odds, the winning probabilities estimated using $Geospatial_{Full}$ are compared to probabilities derived from a model containing only the odds information (*Odds*) for each three year period commencing in 1997. 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 wider public could harness this data to make probability estimates. If this elevation information diffused immediately in the horserace betting market, then from 2003 onwards the LLR test will be insignificant. However, we anticipate that the public will take time to fully appreciate the value of the data. If this is the case, then the LLR test statistic comparing the maximum likelihood values (LL) of the $Geospatial_{Full}$ and the $Odds$ (null model) models would be 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 would become insignificant.

To demonstrate the economic value of the decision-making capabilities which VG enabled, we undertake out-of-sample, betting simulations. In particular, models (10) and (11) are first estimated using the three-year sample of races run from 1997-1999. The coefficients of model (11) are then fixed and these are then used to predict winning probabilities for races run over the next year (2000). The sliding window then moves forward one year and we estimate models (10) and (11) using data from 1998-2000. The new model (11) is then used to estimate winning probabilities in 2001, etc. We then explored if a betting strategy based on the out-of-sample winning probabilities derived from model (11) (i.e. a CL incorporating geospatial information and market odds ($Geospatial_{Full}$)), can provide profitable returns. To develop an appropriate betting strategy, the probability estimates from $Geospatial$, p_{ij}^{full}, are used to calculate the expected return for a £1 bet on each horse in each of the out-of-sample races, as follows:

$$E(R_{ij}) = p_{ij}^{full} * G_{ij} - 1 \tag{16}$$

where 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 shown 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_j horses in race j 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. If horse h wins race j, wealth increases by the following proportion after race j:

$$1 - \sum_{i=1}^{n_j} b_{ij} + b_{hj} . r_{hj}. \tag{17}$$

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

$$\max b_{hj} \sum_{h=1}^{m_j} p_{hj} . \ln(1 - \sum_{i=1}^{n_j} b_{ij} + b_{hj} . r_{hj}) \tag{18}$$

where m_j represents the total number of races in the dataset. Research has suggested that restricting the size of individual bets provides optimal returns in the long run (Maclean et al. 2010), and ensures that expected returns are not biased by one 'lucky win or loss'. Consequently, a maximum limit per bet of 1% of current wealth was 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: *Odds* model) this will show that VG data provide significant measurable value.

4 Results

4.1 Market diffusion

Results of estimating a CL model including *predictedNFP* as the only independent variable (i.e. $Geospatial_{simple}$) using data from races for each of the 3-year periods 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. In addition, the coefficient (β) for *predictedNFP* is positive for each of the 3-year periods (range: 6.0051 to 6.4308; mean: 6.2264), indicating that greater values of predicted NFP are associated with higher probabilities of winning. The mean R^2 of the annual models is 0.0230. This apparently low figure is not surprising for scholars working in the field of horserace market predictions, as complex models involving a multitude of variables have been shown to have R^2 of only around 0.15; yet these models have been shown to achieve significant profits (Lessmann et al. 2012). Consequently, the fact that elevation data alone produces an R^2 of 0.0230 suggests that elevation data provides potentially valuable information that can help forecast winning probabilities.

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).

We also estimated CL models incorporating the single independent variable, *averageNFP*, using data from races for each of the 3-year periods. None of these models were significant at the 1% level, further confirming that horses' *preferences* for features of racetrack topography, rather than simply their previous success rate, are the key predictive ingredients of the variable *predictedNFP*.

Table 2 Results from estimating $Geospatial_{simple}$ for three year periods, 1997-2014.

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****

** denotes significance at 1% level, 2-tailed test.

4.2 Rate of market diffusion

Having confirmed that VGs provide the opportunity to make more accurate winning probability estimates, we then explored how quickly the market discounts this information. To achieve this the predictive accuracy of a model ($Geospatial_F$) incorporating market odds-implied probabilities and probabilities derived from the $Geospatial_{simple}$ model is compared with that of a CL model simply incorporating market odds-implied probabilities ($Odds$). Each of these models is estimated using data from races for each of the 3-year periods from 1999-2014 (inclusive). The predictive accuracy of these two models in each of these 3-year periods is then compared using a LLR. The results are reported in Table 3.

Table 3: Model summary for $\boldsymbol{Geospatial_{Full}}$ and $\boldsymbol{Odds}$ 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.

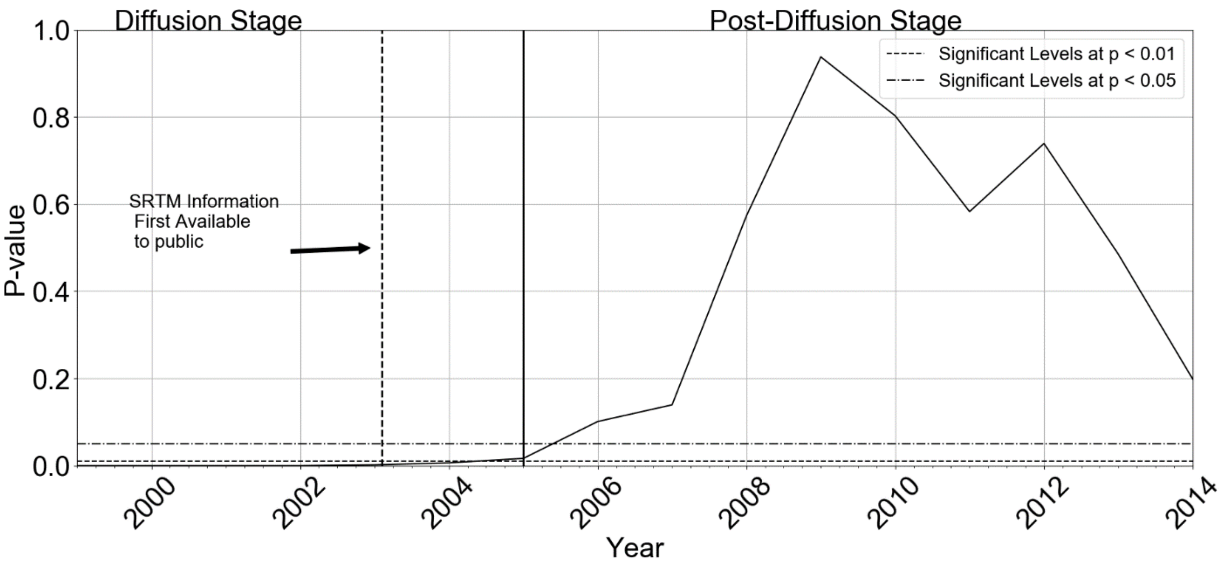
In each of the 3-year periods between 1999-2005 (inclusive), the LLR test statistic is significant at the 5% level, suggesting that the $Geospatial_{Full}$ model better predicts winning probabilities than the $Odds$ model; providing support for H1, that VG information can be used to develop forecasts of horse performance that outperform market forecasts (odds).

However, for all 3-year periods, commencing with the period 2004-2006, the LLR test statistic is insignificant at the 5% level; confirming that prior to the period 2003-2005, the market odds do not fully reflect horses' topographical preferences that could be determined from data available in VGs. However, VG information was fully discounted in odds thereafter.

In Figure 3, the corresponding p-value of the χ^2 tests are plotted over time. A vertical line at 2005[3] separates the 'Diffusion Stage', a period when the $Geospatial_{Full}$ model produces significantly more accurate winning probabilities than $Odds$ model, and the 'Post- Diffusion stage', when market odds fully incorporate the valuable information regarding racetrack topology offered by VGs.

These results lead us to accept H2, that betting market participants take time to learn how to fully account for VG elevation information in odds. Although the SRTM was available to specialists from 1999, it was not until 2003 that the information first became available to the public. The results, therefore, demonstrate a delayed diffusion process, where the full value of information is only realised over a number of years and after the information is made freely available to the public. This finding highlights the need for research that explores efficiency via longitudinal studies.

Figure 3: Diffusion of technology enhanced information over time



4.3 Economic value of technology enhanced decisions

Betting simulations, based on probability estimates derived from employing VG information to account for horses' topographical preferences, show the economic importance of the VG data.

---

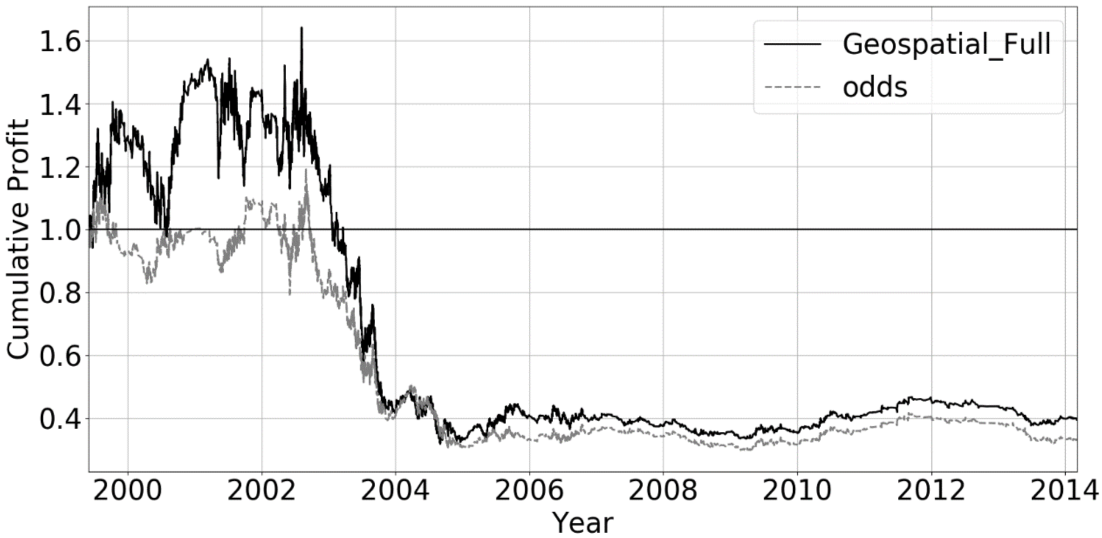
[3] Although the analysis uses 3 year windows, 2005 is selected since the $patial_{Full}$ model produces significantly more accurate winning probabilities than $Odds$ model, in all periods up to an including the 2003-2005 period.

Specifically, the profit achieved by betting strategies employing winning probabilities estimated using a CL model incorporating probabilities derived from the geospatial preferences of horses together with their odds implied probabilities (the $Geospatial_{Full}$ model) are compared with the profits achievable using winning probability forecasts derived from a CL model simply incorporating odds implied probabilities (the $Odds$ model). A comparison of these two sets of profits provides a measure of the economic value of the decisions of bettors using VG data. To ensure that the profits we identified were truly achievable in practice, we estimate these on out-of-sample data. In particular, we estimate the $Geospatial_{Full}$ and the $Odds$ models using three years of race results (e.g., using 1997-1999). The coefficients of these models are then held fixed for forecasting the winning probabilities for races run in the following year (2000, in this example). We then develop separate Kelly betting strategies based on the predicted winning probability estimates for that year (2000, in this example) from both the $Geospatial_{Full}$ and the $Odds$ models. Subsequently, the next three years of data (e.g. 1998-2000) are used to estimate $Geospatial_{Full}$ model and the $Odds$ models. The coefficients of these models are then held fixed for forecasting the winning probabilities for races run in the following year (2001, in this example). This process continues for each out-of-sample dataset from 2000-2014. The results are displayed in Figure 4.

Between 2000 and 2003, the Kelly betting strategy based on Geospatial information ($Geospatial_{Full}$ model) produces 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 decreased. After 2004, the VG information was completely diffused into the market as the cumulative wealth based on a Kelly betting strategy employing the $Geospatial_{Full}$ model dropped below the initial wealth level.

The predicted probabilities from the $geospatial_{Full}$ model provide consistently higher returns compared to those from $Odds$; as depicted by the difference between the two lines in Figure 4. In 1999, 2000, 2001, 2002, 2003, 2004 and 2005 the VG information provides 33%, 16%, 3%, 58%, 145%, 125% higher returns than those generated simply with $Odds$, before plateauing at about 12% higher from 2006-2014. Therefore, using the VG information, betting strategies achieve consistently higher returns than using market odds information alone. However, once the VG information became freely available to the public in 2003, the market began to incorporate the information and the profitable opportunities were eroded, such that abnormal profits could not be earned using this information from 2004 onwards.

Figure 4: Cumulative wealth from Kelly betting strategies based on winning probabilities estimated using information available in (a) VGs and in odds and (b) only in odds.



5. Discussion, implications and conclusion

The results demonstrate the extent and speed that VG-based elevation information diffused through the betting market. Although geospatial data was available to some users in 1999 and to the wider public in 2003, it still took the market until 2004 to fully discount the information.

It has been shown that individuals learn to use new technologies in ways that increase their long-term welfare (Stillwell and Tunney 2012) and our results suggest that profitable opportunities do arise when novel data becomes available. However, we find that these opportunities are eroded as market participants learn to adapt their behavior. In particular, we show that those who were able to exploit the new information offered by VGs could generate abnormal profits for a period, but that these profits disappeared once the market fully accounted for the new information. More generally, our results suggest that technology-enhanced information takes time to diffuse through a market. Consequently, the results are in line with the AMH and these results show how the AMH can be used in sports betting to understand how markets evolve in relation to new information.

An important consideration is that horserace-betting markets are uncertain and dynamic environments (Johnson and Bruce 2001). The uncertainty relates to the uniqueness of each event in terms of the participants, location, conditions and a range of other factors that affect the predictability of a participant's performance. The market is also dynamic, in the sense that information related to these different factors, is constantly updating. In fact, the changing odds during the build up to each race reflect the dynamic updating of new information and odds have been shown in numerous studies to provide well calibrated forecasts of event outcomes (e.g. Johnson and Bruce 2001; Forrest et al., 2005; Baker and McHale, 2013; Štrumbelj and Vračar, (2012), Baboota and Khaur (2018). However, because of its dynamic nature and the fact that there are so many factors that can influence the result of a race, it is very difficult for market participants to account for every piece of information. Some studies, as indicated in Section 2.2, confirm that betting market participants have difficulty in fully discounting certain types of complex information in betting odds. In particular, bettors do not always fully discount data that involves a combination of inter-related complex variables and those that are derived from complex modelling procedures (e.g., Johnson et al., 2010a; Lessmann et al., 2012; Ma et al., 2016). However, these studies often examine market efficiency in a specific period and therefore do not have the ability to measure the speed at which the market participants learn to use information effectively. In addition, because none of these studies examine *novel* information (i.e. not available in a particular form before a specific moment in time), they cannot measure the rate at which *new* information diffuses through the market. Our study, seeks to overcome these limitations by conducting a longitudinal study of the impact on betting market odds of information that was only available in a readily usable format after the arrival of VG elevation data.

As Papagiannidis et al. (2015) observe, isolating diffusion is non-trivial, since Web information diffuses at varying speeds depending on the region and industry. Equally, isolating the effects of one type of information, such as VG information, is difficult because the environment within which diffusion takes place is itself subject to significant change through time. Indeed, the information diffusion we observe in our VG study could to some extent be attributed to the rise in online betting services such as Betfair and Bet365 (founded in 2000), since these changed the nature of the horserace betting market (Johnson et al. 2010b). The arrival of these new online betting organisations has been linked to a significant increase in market efficiency. However, we show that even after the arrival of these new betting organisations, diffusion of VG information took several years. This suggests that even in apparently highly efficient markets (e.g., Smith et al., 2009) where the rewards to effectively using new information are high, market convergence can take time.

A further difficulty arises in isolating the effects of diffusion of VG elevation information, since other factors may affect the performance of the forecasting model used to estimate winning probabilities. In particular, it might be argued that changes in the accuracy of winning probability forecasts over time might be due to changes in other covariates related to horseracing and not due to the diffusion of geospatial information in the market. For instance, horses and their riders might improve over time (e.g., through training) to overcome challenges that persist with certain terrains. Consequently, it might be argued that the dynamic nature of the performance of horses and riders could explain the apparent reduction in the value which geospatial information provides for predicting winning probabilities.

To confirm that the VG-based elevation data offered information that was not accounted for in data from other sources and was not fully accounted for in the market we conducted a form of robustness analysis. In particular, we explored the degree to which VG based data could be used to produce winning probability estimates which outperformed predictions from a model incorporating both odds and fundamental factors related to horse and jockey performance, shown in previous studies to be predictive (see Sung et al. (2005)). The results of this analysis are presented in Appendix C. These results show that even after controlling for the most important fundamental features, the geospatial information is still the key ingredient for superior forecasts and profitable returns.

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 probabilities, the simple CL model employing only geospatial variables is most predictive in 2005-2008, the years when the market became efficient. Consequently, minor variations in the R^2 of the geospatial variables reported in table 2 show that the models do not deteriorate over time and the declining diffusion rate was not linked to a reduction in predictive value of the geospatial variables.

A further potentially confounding factor is that the topography of racetracks might change following reconstruction, maintenance efforts or following changes of layout. To control for this possibility, elevation data was collected in 2010 and again in 2018, to see how much the data changed. These two samples show no significant changes in the topography/layout of the tracks, indicating that the forecasting model is built on topography information that does not change considerably over time.

Attributing the declining trend in profits achievable from a Kelly betting strategy based on the topography preference variables also suffers from some potential limitations. In particular, an important assumption inherent in the interpretation of the findings is that bettors are indeed utilizing geospatial information to make betting decisions. The analysis focusses on the rate at which the aggregate-level market incorporates information. This is important because betting markets, as in most financial markets, have been shown to be made up of sophisticated (informed) bettors and less informed bettors. It requires only a few well-informed bettors to use sophisticated analysis techniques to extract the value form the VG information and to bet based on this, for market odds to move to levels that accurately reflect each horse's winning chance. We speculate that this may cause the market convergence we observe.

Web-based information diffusion is important in the context of modern markets, since increasing amounts of information are available online. Equally, information diffusion is important in a wide range of fields beyond financial markets, including disaster response, product diffusion and epidemiology (Nagarajan et al. 2012). Information transfer is key, not just within markets but between markets (demonstrated by the fact that similar commodity futures and equity markets are interconnected: Bekiros et al. 2017). We believe that the longitudinal method employed here, employing rolling sub-sample windows over a long period, can be used to good effect to understand evolution from inefficiency to efficiency in different markets and to show how markets adapt to the arrival of new information.

In summary, our results suggest that the full value of novel, technology-enabled elevation information was diffused through the market over a period. However, the economic gains from employing the new data decreased almost immediately when the information became available. This suggests that it is the incentives which markets offer (betting profits in the market we observe) that serve to help the market adapt to new technology-enhanced information.

This eighteen-year longitudinal study demonstrates that VG information can be used to improve market-based forecasts by generating variables that can improve winning probability forecasts. It also shows how the information diffuses through the market as the market progresses towards efficiency. Furthermore, we show that despite the need to employ sophisticated procedures to extract the full value from the VG topographical information, the data was fairly quickly discounted by the aggregate market odds when the information became available to the public. This finding therefore fully supports the view that markets rapidly adapt to the arrival of new, even complex, information.

References

Al-khazali, O., and Mirzaei, A. 2017. "Stock Market Anomalies, Market Efficiency and the Adaptive Market Hypothesis: Evidence from Islamic Stock Indices," *Journal of International Financial Markets, Institutions & Money* (51), 190–208.

Baboota, R., and Kaur H. 2018. "Predictive analysis and modelling football results using a machine learning approach for English Premier League," *International Journal of Forecasting* (in press).

Baker, R.D, and McHale, I. G. 2013. "Forecasting Exact Scores in National Football League Games,"*International Journal of Forecasting* (29), 122–130

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), 945–961.

Benter, W. 1994. "Computer Based Horse Race Handicapping and Wagering Systems: A Report," in Efficiency of Racetrack Betting Markets, London: Academic Press, 183–198.

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), 1040–1060.

Brecher, S. L. 1980. "Beating the Races with a Computer", Long Beach: Software Supply.

Brehmer, B. and Allard, R. 1991. “Dynamic Decision Making: The Effects of Task Complexity and Feedback Delay”. In J. Ramussen, B. Brehmer, and J. Leplat (eds.), *Distributed Decision Making: Cognitive Models Of cooperative Work.* Chichester: Wiley.

Brown, A., and Reade, J. J. 2019. “The Wisdom of Amateur Crowds: Evidence from an Online Community of Sports Tipsters,” *European Journal of Operational Research* (272:3), 1073–1081. (https://doi.org/10.1016/j.ejor.2018.07.015).

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.

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.

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.

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. (https://doi.org/10.1016/j.jfineco.2004.06.004).

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.

Croxson, K. and Reade, J. J. (2014). “Information and efficiency: Goal arrival in soccer betting”. *Economic Journal*, 124, 62–91.

Deschamps, B. and Gergaud, O. (2008). “Efficiency in Horserace Betting Markets. The Role of the Professional Tipster”. In D. B. Hausch, & W. T. Ziemba (Eds.), Handbook of sports and lottery markets. North Holland: Elsevier

Doyle, J. R., and Chen, C. H. 2013. “Patterns in Stock Market Movements Tested as Random Number Generators,” *European Journal of Operational Research* (227:1), 122–132.

Drnevich, P. L., and Croson, D. C. 2013. “Information Technology and Business-Level Strategy: Toward an Integrated Theoretical Perspective,” *MIS Quarterly* (37:2), 483–509.

Fama, E. F. 1970. “Efficient Capital Markets : A Review of Theory and Empirical Work,” *The Journal of Finance* (25:2), 383–417.

Fischer, T., and Krauss, C. 2018. “Deep Learning with Long Short-Term Memory Networks for Financial Market Predictions,” *European Journal of Operational Research* (270:2), 654–669.

Flaxman, M., and Vargas-Moreno, J. C. 2012. “Introduction: Science, Technology, and Engineering (Tools and Methods),” in Restoring Lands - Coordinating Science, Politics and Action, H. Karl, L. Scarlett, J. C. Vargas-Moreno, and M. Flaxman (eds.), London: Springer, 21–27.

Forrest D. K., and Simmons R. 2000. “Forecasting Sport: The Behaviour and Performance of Football Tipsters”. *International Journal of Forecasting*, 16, 317–331.

Forrest, D., Goddard, J. and Simmons, R. 2005. “Odds-setters as Forecasters: The Case of English Football. International Journal of Forecasting, 21, 551-564.

Franck, E., Verbeek, E. and Nuesch, S. 2010. Prediction Accuracy of Different Market Structures – Bookmakers versus a Betting Exchange. *International Journal of Forecasting*, 26(3), 448-459.

Frino A, Grant J, and Johnstone D. 2008. “The House Money Effect and Local Traders on the Sydney Futures Exchange”. *Pacific-Basin Finance Journal*, 16(1):8-25.

Ghazani, M. M., and Araghi, M. K. 2014. “Research in International Business and Finance Evaluation of the Adaptive Market Hypothesis as an Evolutionary Perspective on Market Efficiency : Evidence from the Tehran Stock Exchange,” *Research in International Business and Finance* (32), 50–59.

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), 11088–94.

Goodwin, P., Önkal, D., and Thomson, M. (2010). “Do forecasts expressed as prediction intervals improve production-planning decisions?” *European Journal of Operational Research*, 205(1):195-201.

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), 481–502.

Hillmer, S., and Yo, P. 1979. “The Market Speed of Adjustment,” *Journal of Financial Economics* (7), 321–345.

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), 686–94.

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), 992–1007.

Hwang, J. H., and Kim, M. 2015. “Misunderstanding of the Binomial Distribution, Market Inefficiency, and Learning Behavior : Evidence from an Exotic Sports Betting Market,” *European Journal of Operational Research* (243:1), 333–344.

Johnson, J. E. V, Brien, R. O., and Sung, M. 2010a. “Assessing Bettors ’ Ability to Process Dynamic Information : Policy Implications,” *Southern Economic Journal* (76:4), 906–931.

Johnson, J., Bruce, A., Yu, J., Johnson, J., Bruce, A., and Yu, J. 2010b. “The Ordinal Efficiency of Betting Markets : An Exploded Logit Approach The Ordinal Efficiency of Betting Markets : An Exploded Logit Approach,” *Applied Economics* (42:29), 3703–3709.

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), 265– 290.

Johnstone, D. 2016. “The Effect of Information on Uncertainty and the Cost of Capital,” *Contemporary Accounting Research* (33:2), 752–774.

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), 1116–1127

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* (https://doi.org/10.1016/j.techfore.2017.11.009).

Kelly, J. L. 1956. "A New Interpretation of Information Rate," Information Theory, IRE Transactions on 2.3, 185–189.

Keyhole. 2003. "Keyhole: Industry Solutions," Web.archive.org, (http://web.archive.org/web/20030801173507/http://www.keyhole.com/industry_solutions/index. html; retrieved 19 July, 2013).

Khuntia, S., and Pattanayak, J. K. 2018. "Adaptive Market Hypothesis and Evolving Predictability of Bitcoin," *Economics Letters* (167), 26–28.

Kim, J. H., Shamsuddin, A., and Lim, K.-P. 2011. "Stock Return Predictability and the Adaptive Markets Hypothesis: Evidence from Century-Long U.S. Data," *Journal of Empirical Finance* (18:5), 868–879.

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), 364–390.

Knottenbelt, W., Spanias, D. and Madurska, A. 2012. "A Common-Opponent Stochastic Model for Predicting the Outcome of Professional Tennis Matches". *Computers and Mathematics with Applications*, 64, 3820-3827.

Lessmann, S., Sung, M. and Johnson, J. 2009. "Identifying Winners of Competitive Events: A SVM-based Classification Model for Horserace Prediction". *European Journal of Operational Research*, 196 (2), 569–577.

Lessmann, S., Sung, M.-C., and Johnson, J. E. V. 2010. "Alternative Methods of Predicting Competitive Events: An Application in Horserace Betting Markets," *International Journal of Forecasting* (26:3), 518–536.

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), 163–174.

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), 69–108.

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), 688– 700.

Lo, A. 2004. "The Adaptive Markets Hypothesis: Market Efficiency from an Evolutionary Perspective," *Journal of Portfolio Management* (30:5), 15–29.

Lo, A. 2005. "Reconciling Efficient Markets with Behavioral Finance: The Adaptive Markets Hypothesis," *Journal of Investment Consulting* (7:2), 21–44.

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

Ma, T., Tang, L., McGroarty, F., Sung, M.-C., and Johnson, J. E. V. 2016. “Time Is Money: Costing the Impact of Duration Misperception in Market Prices,” *European Journal of Operational Research*, 255(2), 397-410.

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), 681–687.

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

McHale, I. and Morton, A. 2011. “A Bradley-Terry Type Model for Forecasting Tennis Match Results”. *International Journal of Forecasting*, 27, 619-630.

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.

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), 601–629.

Mills, B. M., and Salaga, S. 2018. “A Natural Experiment for Efficient Markets: Information Quality and Influential Agents,” *Journal of Financial Markets* (40), Elsevier B.V., pp. 23–39. (https://doi.org/10.1016/j.finmar.2018.07.002).

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), 1039–46.(

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), 810–819.

Newton, P. K. and Aslam, K. 2009. “Monte Carlo Tennis: A Stochastic Markov Chain Model”. *Journal of Quantitative Analysis in Sports*, 4 (3), 1–42.

Ng, P., Wong, W., and Xiao, Z. 2017. “Stochastic Dominance via Quantile Regression with Applications to Investigate Arbitrage Opportunity and Market Efficiency,” *European Journal of Operational Research* (261:2), 666–678.

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), 308– 321.

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), 1–8.

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), 241–262.

Rosenbloom, E. 2003. "A Better Probability Model for the Racetrack using Beyer Speed Numbers". *Omega: The International Journal of Management Science*, 31, 339-348.

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), 243–256.

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), 602–607.

Serwe, S. and Frings, C. 2006. "Who will win Wimbledon? The Recognition Heuristic in Predicting Sports Events". *Journal of Behavioural Decision Making,* 19(4), 321-332.

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), 318–330.

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), 2102–17.

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-20121?op=1&IR=T; retrieved 13 February, 2014).

Simon, H. 1955. "A Behavioral Model of Rational Choice," *The Quarterly Journal of Economics* (69:1), 99–118.

Smith, M. A. 2003. "The Impact of Tipster Information on Bookmakers' Prices in UK Horserace Markets". In L. Vaughan Williams (Ed.), The Economics of Gambling. London: Routledge.

Smith, M. A., Paton, D. and Vaughan Williams, L. 2006. "Market Efficiency in Person-to-Person Betting". *Economica*, 73, 673–689.

Smith, M., Paton, D., and Vaughan Williams, L. 2009. "Do Bookmakers Possess Superior Skills to Bettors in Predicting Outcomes?" *Journal of Economic Behavior and Organization* (71:2), 539–549.

Spann, S. 2003. "Internet-based virtual stock markets for business forecasting". *Management Science*, 49, 1310-1326.

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, 55-72.

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), 678–681.

Stillwell, D. J., and Tunney, R. J. 2012. "Individuals' Insight into Intrapersonal Externalities," *Judgment and Decision Making* (7:4), 390–401.

Štrumbelj, E., and Vračar, P. 2012. "Simulating a Basketball Match with a Homogeneous Markov Model and Forecasting the Outcome," *International Journal of Forecasting* (28), 532–542.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), 43–59.

Sung, M.-C., and Johnson, J. E. V. 2008. “Semi-Strong Form Efficiency in the Horse Race Betting Market,” in D. B. Hausch and W.T. Ziemba (eds.) *Handbook of Sports and Lottery Markets*, 275-306, North-Holland, Amsterdam.

Sung, M.-C., 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), 128–147.

Sung, M.-C., Johnson, J.E.V., and Bruce, A. C. 2005. “Searching for Semi-Strong Form Inefficiency in the UK Racetrack Betting Market,” in *Information Efficiency in Financial and Betting Markets*, 179–192, Cambridge University Press: Cambridge.

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. (https://doi.org/10.5026/jgeography.78.45).

Tziralis, T. and Tatsiopoulos, I. 2007. “Prediction Markets: An Extended Literature Review. *Journal of Prediction Markets*, 1, 75-91.

Urquhart, A and Hudson, R (2013). Efficient or Adaptive Markets? Evidence from Major Stock Markets using Very Long Run Historic Data. *International Review of Financial Analysis*, 28, 130-142.

Urquhart, A., and Mcgroarty, F. 2014. “International Review of Financial Analysis Calendar Effects , Market Conditions and the Adaptive Market Hypothesis : Evidence from Long-Run US Data,” *International Review of Financial Analysis* (35), 154–166.

Urquhart, A., Gebka, B., and Hudson, R. 2015. “How Exactly Do Markets Adapt ? Evidence from the Moving Average Rule in Three Developed Markets,” *Journal of International Financial Markets, Institutions & Money* (38), 127–147.

Urquhart, A. 2017. “How Predictable Are Precious Metal Returns?” *The European Journal of Finance* (23:14), 1390–1413.

Vaughan Williams, L. 1999. “Information Efficiency in Betting Markets: A Survey,” *Bulletin of Economic Research* (51:1), 1–39.

Vaughan Williams, L. 2000. “Can Forecasters Forecast Successfully? Evidence from UK Betting Markets,” *Journal of Forecasting* (19), 505–513.

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), 747– 765.

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

Yu, L., Zhao, Y., Tang, L., and Yang, Z. 2018. “Online Big Data-Driven Oil Consumption Forecasting with Google Trends,” *International Journal of Forecasting*, (https://doi.org/10.1016/j.ijforecast.2017.11.005).