The Journal of Prediction Markets 2017 Vol 11 No 2 pp 60-76

EVIDENCING THE FORECASTING PERFORMANCE OF PREDICATION MARKETS: AN EMPIRICAL COMPARATIVE STUDY

*Patrick Buckley**
Department of Management and Marketing, University of Limerick, Ireland
Email: Patrick.Buckley@gmail.com

Elaine Doyle
Department of Accounting and Finance, University of Limerick, Ireland
Email: Elaine.Doyle@ul.ie

ABSTRACT

Accurately forecasting uncertain outcomes to inform planning processes and aid decision making is a perennial organisational challenge, and the focus of a substantial body of research in management science, information systems and related disciplines. Academic research suggests that prediction markets may be of significant benefit to organisations in meeting this challenge. However most of the empirical studies assessing prediction market performance are laboratory based and suffer from limits to their generalizability. Recent literature has called for research which analyses the performance of prediction markets in ecologically valid settings in order to evidence their effectiveness to potential organisational users. This paper answers these calls by designing a prediction market to forecast an uncertain real world event. The study then compares the forecasting performance of the prediction market with a number of more traditional forecasting approaches regularly used by organisations. The study is contextually situated in a low information heterogeneity problem space, where relevant information is freely available. The results suggest that in this context prediction markets outperform the other forecasting methods studied.

Keywords: Prediction Markets, Field Experiments, Business Forecasting, Group Decision

INTRODUCTION

Accurately forecasting uncertain outcomes to inform planning processes and aid decision making is a perennial organisational challenge. It has been a pre-eminent theme in management science and information systems research, with a large and growing body of work focused on understanding how

* Corresponding author.

technology in general and information systems in particular can address this challenge. As a form of Group Decision Support System (GDSS), prediction markets seek to leverage "the wisdom of crowds" by utilising information technology to aggregate the opinions and knowledge of large numbers of people.

While the potential of prediction markets as forecasting tools is recognised in the literature, many questions remain unanswered. Extant academic studies assessing prediction market accuracy have been generally encouraging. However, while research results motivate further study, they have limits in terms of supporting the organisational deployment of prediction markets or guiding further research. A more nuanced understanding of the forecasting strengths of prediction markets is required in order to facilitate their transition from academic curiosity to practical organisational tool. This is evidenced by calls in the literature for ecologically valid research which demonstrates evidence of prediction market performance compared with comparable forecasting approaches in realistic contexts. This paper contributes to that agenda by designing an ecologically valid prediction market aimed at forecasting a real-world event in a low information heterogeneity environment, and comparing its forecasting performance to that of a number of more traditional forecasting approaches.

We start by investigating whether a group of relatively inexperienced individuals using a prediction market can forecast more accurately than individual experts in the area of interest. Delving more deeply, we then investigate the performance of the prediction market relative to a number of other forecasting approaches used by organisations to aid decision making; namely the Key Informant and Combined Judgement techniques.

The remainder of the paper is structured as follows. First, the extant literature on institutional forecasting is briefly reviewed and the study is motivated. The methodology section describes the context of the study, as well as the data collection and analytical processes used. The subsequent section presents results, while the paper concludes with a discussion of findings, implications and suggestions for further research.

INSTITUTIONAL FORECASTING APPROACHES

Organisations continually need to make decisions based in whole or part on the forecast outcome of large, uncertain and complex systems. For example, when a manufacturing organisation needs to schedule production, its decision will be based in part on an estimation of future demand for its products, a challenging forecasting problem. This challenge is exacerbated by the rapid metabolism of business in the modern world, caused by a range of factors such as increased competition, globalisation, the emergence of new technologies, accelerating innovation and new regulatory, environmental and ethical constraints (Haase & Franco, 2011).

There are two macro level paradigms used by organisations to make forecasts in complex problem spaces (Armstrong, 2001). The first approach focuses on using statistical methods to develop quantitative models that can be used to derive forecasts. As well as established methods such as simulation and linear programming, many organisations now use techniques such as machine learning and neural networks. These techniques leverage the enormous datasets that can be created, maintained and analysed by modern information technology.

However, quantitative approaches to modelling large, complex systems face a number of serious limitations. First, the number of interconnected variables that may be required to model a realistically complex system may be computationally prohibitive. Second, the model maker may be unaware of important variables to include in the model. Third, it may be impossible to define the nature of relationships between variables, particularly in contexts where those relationships are in constant flux. Many variables of interest may be inherently inscrutable. For example, it is reasonable to suggest that consumer sentiment will influence customers buying preferences, but it is also evident that sentiment is a label attached to a fluid, multi-faceted construct that defies straightforward computation. All of these factors limit the accuracy that can be achieved with statistical approaches to forecasting.

The second forecasting archetype seeks to utilise expert informants or groups of informants who can make accurate forecasts. Individual experts use knowledge, heuristics and experience to make forecasts. However, the literature recognises that there are clear limitations on the rationality and information processing capabilities of the human brain (March, 1999; Simon, 1997). Ultimately, intellectual artefacts such as forecasts which are derived from the human mind are subject to the cognitive, psychological and emotional limitations of the human brain (Chugh & Bazerman, 2007). There is a clear consensus in the literature that there are fundamental limitations on the ability of individual humans to create accurate forecasts.

In order to ameliorate these limitations, many approaches seek to use the combined judgement of groups of experts. Groups of individuals should have access to more information than a solitary individual (Hitt, Black, & Porter, 2005). A group should have access to more resources than an individual, particularly cognitive resources such as attention, as well as having access to the pooled skills and knowledge of all of the participants (Ellis & Fisher, 1994). By reducing the effect of psychological and emotional biases, groups often have a particular edge in tasks which call for judgement (Ellis & Fisher, 1994). Groups can leverage the "assembley effect", whereby social interaction can prompt creativity and the generation of novel solutions to problems (Laughlin, Bonner, & Miner, 2002).

However, group forecasting is not a universal panacea. There are a range of negative second order effects that can adversely affect the performance of groups such as groupthink (Janis, 1972), information cascades (Anderson & Holt, 1997), group polarization (Isenberg, 1986) and escalating commitment

(Sprenger, Bolster, & Venkateswaran, 2007). Such effects are caused by the social nature of a group forecasting context. By their nature group forecasting contexts have both task and social dimensions, and in many cases social considerations can dominate task considerations. To minimise these socialisation effects, structured group forecasting approaches such as brainstorming, the Nominal Group Technique and the Delphi method have evolved (Hitt et al., 2005).

Prediction markets are a recently developed tool that can leverage information technology to enable large groups of individuals to collaborate in a structured way to create forecasts and reach decisions (Geifman, Raban, & Sheizaf, 2011). They are exciting increasing interest from both the academic and practitioner community. They are based on Hayek's conceptualisation of markets as near perfect transmitters of information (Hayek, 1945). They use a market mechanism to aggregate information held by a diverse population of participants and use that information in the form of market values to make predictions about specific future events (Tziralis & Tatsiopoulos, 2007). By way of example, consider an organisation that wishes to forecast whether a project will meet a specific deadline. A market maker begins by offering for sale a contract on the outcome of the deadline. The contract will pay a holder $1 if the deadline is reached or $0 otherwise. The initial price of the contract would be set to 50 cents and then offered for sale to individuals participating in the project. Under these circumstances, if an individual believes the deadline will be achieved, they will buy the contract in the expectation of a making a profit in the future. Equally, if a rational participant believes the deadline will not be achieved, then they will sell (or 'short') the contract. This dynamic changes the price of the contract, which ultimately moves to reflect the consensus of the group as a whole of the likelihood of the project achieving its deadline. This binary model can be extended to allow a range of outcomes. Equally, they can be used to allow participants to forecast values rather than select from a particular set of options. Other modifications to the basic concept allow specialised prediction markets, referred to as imagination markets, to rank product ideas (Horn & Ivens, 2015; LaComb, Barnett, & Pan, 2007).

Prediction markets have a number of advantages vis-à-vis comparable group decision making techniques such as polls or expert groups (Servan-Schreiber, Wolfers, Pennock, & Galebach, 2004; Garvey & Buckley, 2010). First, prediction markets incentivise truthful information revelation by providing an individualised financial incentive (Hall, 2010). This individualised incentive also serves to encourage participants to seek out new, relevant information (Berg & Rietz, 2003). The operation of the market mechanism serves to simultaneously share and aggregate participants' information. The market also provides a mechanism that allows participants to weight forecasts by varying the number of contracts they buy (Graefe & Weinhardt, 2008). When enabled by information technology, prediction markets can scale efficiently to very large groups with negligible overheads

(Hahn & Tetlock, 2006b), a significant advantage. They can operate for extended periods of time, allowing participants to re-assess their forecasts in light of newly revealed information (Spann & Skiera, 2003). Finally, they can enable anonymous participation, a feature that can be used to moderate the potential negative effects of power relationships and social interactions in a group context (Remidez & Joslin, 2007).

Motivation for Study

Academic research to date suggests that prediction markets "*provide accurate forecasting and effective aggregation*" (Hall, 2010, p. 45). However, some authors caution against drawing definitive conclusions, summarising the existing empirical evidence as cautiously optimistic (K.-Y. Chen, Fine, & Huberman, 2003; Einbinder, 2006; Gruca, Berg, & Cipriano, 2008; Ledyard, 2006; Wolfers & Zitzewitz, 2006). As practitioner interest in prediction markets increases the importance of investigating their reliability and underlying mechanisms grows (Boulu-Reshef, Comeig, Donze, & Weiss, 2016). Graefe and Armstrong (2011) note that studies are limited and often of a small scale. Most of the extant studies are laboratory based and suffer from limits to their generalizability (Buckley & O'Brien, 2015). Such studies have limited numbers of participants in the market, are of limited duration and offer stylised contracts for trade. However, many of advantages ascribed to prediction markets are explained by reference to them allowing large numbers of participants interact over a period of time. By limiting these interactions, experimental work risks underweighting the potential benefits of prediction markets.

In an applied context, the limitations of laboratory based studies are often a serious impediment to practitioner acceptance (Deck, Lin, & Porter, 2013). Laboratory based studies do not offer reassurance to managers and decision makers who are considering the deployment of prediction markets but are concerned with the generalisability of observed results to real world settings. Slamka et al (2013) call for further research which analyses the performance of prediction markets in real-world settings, while Jian and Sami (2012) echo this concern with a call for field experiments with larger groups.

This research aims to answer calls by authors such as Graefe and Armstrong (2011, p. 195) who observe "*Future research should evaluate the relative performance of the methods for more complex problems in more realistic environments*". Our study answers calls in the literature for research that moves beyond experimental work and has ecological validity.

The prediction market analysed in this study operates using industrial parameters. It has a large number of participants and runs over a period of time measured in weeks. We answer calls to move beyond evaluations of absolute performance to investigate relative performance compared with more established institutional forecasting approaches such as Combined Judgement and Key Informant Forecasting. The study focusses on a particular context, namely a low information heterogeneity environment. By focussing on a specific context, this study contributes a body of empirical evidence supporting the suitability of prediction markets for particular contexts which can serve both to inform practitioners and guide further research.

Research Questions

In order to compare the relative performance of forecasting mechanisms, the first step is to identify a suitable forecasting challenge, which can then be used to evaluate the relative performance of various forecasting mechanisms. In this study, the objective was to forecast the tax policies that might be introduced or altered as part of a national budget. This is a low information heterogeneity forecasting space, in that most of the relevant information is widely available, and there is little unique information available to participants (Van Bruggen, Spann, Lilien, & Skiera, 2010).

Our first research objective was to investigate whether a large group of relatively inexperienced individuals using an ecologically valid prediction market can forecast more accurately than individual experts within the relevant field. In order to address this objective we used two distinct forecasting populations with differing levels of expert knowledge and experience. The prediction market participants were undergraduate business students with a limited knowledge of tax. In contrast, the forecasters recruited to act as expert informants were practitioners working in the tax industry for periods ranging from 2 to 30 years.

The second objective was to answer calls in the literature for comparative studies that evaluate prediction markets relative to other organisational forecasting approaches. Qualitative approaches to forecasting can be broadly divided into two categories, namely using a Key Informant or Combined Judgement Forecasts. The Key Informant approach involves an individual decision maker applying themselves to a forecasting problem. From an organisational perspective, the individual may be selected because of attributes such as experience or expertise. However the defining characteristic of such approaches is that an individual makes a forecast alone, with all the advantages and disadvantages resulting from individual decision making.

It is reasonable to suppose that there would be variance across individuals in terms of their forecasting ability. Individuals who made the best and worst forecasts are easy to identify post hoc, however, if an organisation uses the Key Informant forecasting approach it must identify the individual who will be the key informant beforehand. Experience is commonly used in organisations as a proxy for knowledge and ability. For the purposes of this study, we selected our Best Key Informant (BKI) and Worst Key Informant (WKI) as being the individuals who had the longest and shortest experience in the industry respectively. This selection rule is consistent with organisational norms whereby higher level managerial functions such as forecasting are assigned on the basis of experience. It is important to note here that the BKI and WKI selected before the study were not the actual best or worst individual forecaster as evidenced after the budget was published. Combined Judgement approaches, on the other hand, use groups of individuals to make decisions.

The advantages ascribed to group decision making processes have been outlined previously. There are a wide variety of approaches that can be used to combine judgements. Combined judgement forecasts can use relatively informal methods whereby a group collectively arrives at a forecast by conversing, to more structured methods such as the Delphi Method or the Nominal Group technique. We examine two methods of combining judgements in this study. Simple Combined Judgement Forecasting (SCJF), involves combining the judgements of individuals using an simple average. We also examine Weighted Combined Judgement Forecasting (WCJF), where the forecasts of individual experts are combined using experience based weighted averages.

Plott's (2000) analysis of markets suggests that they perform three main information processing roles. First, they gather information distributed across a system. Second, they aggregate that information together into a collective estimate. Third, they disseminate that information to participants. Following the approach used by Van Bruggen et al (2010), we use this taxonomy to guide hypotheses development.

In a low information heterogeneity environment, the majority of information is available to all participants. Many authors suggest that prediction markets encourage individuals to seek out information or prompt truthful revelation of information which participants may otherwise wish to conceal. However, in a context where all the relevant information is widely available, and no participant has access to privileged information, nor a plausible means of obtaining privileged information, there is no reason to suppose that a prediction market has an inherent superiority over other forecasting methods.

On the other hand, prediction markets can aggregate information effectively using a market mechanism. Moreover, the use of this market mechanism allows individual participants to weight their contribution to the collective estimate by signalling their confidence via the volume of contracts they buy or sell. The trading of contracts on the market also enables information dissemination within the group. Individual participants can observe movements in the collective estimate, infer the reasons for those movements and modify their own judgement accordingly.

The above analysis suggests that the ability of a prediction market to allow participants to weight their forecasts and receive and integrate feedback from the group should position them as superior forecasting tools when compared with Combined Judgement Forecasting Methods. Of the two combined judgement methods examined, it is reasonable to suppose that the WCJF method would outperform the SCJF since it places a premium on the experience of individual forecasters.

When comparing the performance of individuals to that of the prediction market and the CJF methods, we would expect that group based forecasting methods would outperform individual mechanisms. In the context of an individual making a decision, there are no other individuals to gather information from or to dissemination information to We would expect Combined Judgement Forecasting Methods to outperform individual forecasting because groups have access to more information. From the above analysis, a hypotheses chain emerges which suggests that:

PM > WCJF > SCJF > BKI > WKI

Our research objectives therefore lead us to five distinct hypotheses as follows:

H1: The forecasting outcome of a prediction market will be more accurate than the forecasting outcome of the median expert informant, as identified post hoc.
H2: The performance outcomes of the Prediction Market would be significantly more accurate than the performance outcomes of the WCJF forecasting approach.
H3: The performance outcomes of the WCJF forecasting approach would be significantly more accurate than the performance outcome of the SCJF forecasting approach.
H4: The performance outcome of the SCJF forecasting approach would be significantly more accurate than the performance outcome of the BKI forecasting approach
H5: The performance outcome of the BKI forecasting approach would be significantly more accurate than the performance outcome of the WKI forecasting approach.

METHODOLOGY

Our study compares the performance of a number of forecasting approaches. On an annual basis the Minister for Finance introduces a National budget to manage the economy and fund government spending. The fundamental forecasting activity that occurs was that of providing prior forecasts as to what policy initiatives would be introduced as part of the national budget.

The policy initiatives introduced in the budget impact upon the operation of business in a wide variety of ways. For example, changes to the personal tax regime may require organisations to modify the processes they use to reimburse employees. At the other end of the scale, government tax policies can have a significant strategic impact on organisations by altering corporation tax rates or modifying the availability of, for example, tax credits for research and development. The policies implemented by government have a direct, measurable effect on the operation of companies, and so forecasting these is of clear practical interest to many business organisations.

The national budget is often the subject of intense media speculation. There is a large quantity of information publicly available. This is usually distributed either by news organisations that have a vested interest in reaching many people, or stakeholder organisations who seek to influence the budget by promulgating position papers. In either case, the systemic effect is to ensure that any information that is publicly available about the budget is rapidly circulated and widely available. There are individuals who will have privileged information about the content of the budget, such as the Minister for Finance, his cabinet colleagues and senior members of the civil service, but this is a relatively small group of individuals. Thus, the forecasting problem is a low information heterogeneity space, in the sense that most individuals have access to the same information. In the context of this research project none of the market participants or key informants had access to privileged information about the budget in advance of its publication.

Our forecasting problem space thus provides a domain that matches the calls in the literature. It presents a realistic forecasting problem that is sited in a particular context, namely a low information heterogeneity environment.

There were two primary processes by which data was collected for this project. The first mechanism was a prediction market, which was called the National Budget Forecasting Project (NBFP). This project consisted of a prediction market which asked student participants to forecast the outcome of a range of potential policy changes. An example of such a question would be "*What will the Irish corporation tax rate be after the next budget?*" A range of answers were offered:

- < 11%
- >= 11% and < 12.5%
- = 12.5%
- > 12.5%% and < 13%
- >= 13.0%

Students were given €5,000 in virtual cash when the market opened. They used this to invest in the outcome they considered most likely for each question (the contract). The NBFP commenced on 19th October and remained open for five weeks until 18th November. Three questions were originally posted. For each question, the contracts available were initially set as being equally likely. In the example above, each contract was originally set to a probability of 20%. As participants traded, the contract prices changed accordingly. Additional questions where posted throughout the operation of the prediction market. By the closing date, 12 forecasting problems were posed to participants. 57 undergraduate students in 3rd year of a 4 year undergraduate business degree participated in the NBFP. They received marks for their overall forecasting performance as individuals, which were calculated on the 6th of December after the National Budget was actually announced. Over the course of the project, 3,474 trades were executed.

A second data collection procedure was used to collect data from the Key Informants. This consisted of a survey including the 12 forecasting questions the students were asked on the prediction market, with participants asked to forecast the result by placing a tick next to the outcome they believed most likely to occur. It also asked participants to note how long they had been working in the tax industry. In order to collect this data, an electronic survey was distributed to tax practitioners accessed through the professional network of one of the researchers. Ten individuals were contacted directly, and they were asked to pass the survey to other individuals in their professional network. The survey was distributed on 17th November. In order to avoid the possibility that newly revealed information would give an informant an advantage, only responses received before t19th November were considered.

The dataset used to investigate our hypotheses consisted of the aggregated forecasts of the student participants as derived from the prediction market and the survey responses from the Key Informants. The forecasts were collected on 18th November, while the event of interest, the National Budget, occurred on 5th December.

Performance Measures

There are a number of metrics that can be used to assess the prediction accuracy of forecasting methods, including the Absolute Error, the Quadratic Score and the Logarithmic Score (Servan-Schreiber et al., 2004). In this study, we use the Quadratic Score, which is one of the most commonly used metrics in evaluating forecasting accuracy (Y. Chen, Chu, Mullen, & Pennock, 2005). The Quadratic Score is calculated as per Figure 1.

$$Quadratic\_Score = 100 - 400\ X\ (Prob\_Lose^2)$$

Figure 1: Quadratic Score

Prob_Lose is the probability assigned to the outcome(s) that do not in fact happen. The Quadratic Score can be positive or negative, and a prediction with a higher quadratic score is more accurate.

Deriving Predictions

In order to investigate our research hypotheses, we needed to calculate a quadratic score for each question for each forecasting method, which in turn required the calculation of a probability distribution for each question for each method. To create these probability distributions, we followed the approach used by Chen et al. (2005). The calculation of a probability distribution for a prediction market is relatively straightforward. On the 18th of November, we took the price of each individual contract, and divided it by 100 to obtain the market's prediction of the outcome represented by that security occurring. This provided the probability distribution implicit in the final price of the contracts being traded.

In order to derive a probability distribution for an individual respondent, we assigned a probability of 1 to the outcome the respondent indicated that they would believe happen, and 0 to all the other outcomes for a given question.

In order to calculate the Simple Combined Judgement Forecasts (SCJF), we calculated the aggregated mean probability across all the respondents for each outcome for each question (see Figure 2).

$$Probability_{qo} = \frac{\sum_{i=0}^{n} Probability_{qoi}}{n}$$

Figure 2: Simple Combined Judgement Forecast Score

$Probability_{qo}$ is the aggregated probability assigned to outcome *o* for question *q* with *n* being the number of respondents. The Weighted Combined Judgement Forecasts are computed as being the experience weighted aggregated mean probability across all the respondents for each outcome for each question.

$$Probability_{qo} = \frac{\sum_{i=0}^{n} Probability_{qoi} \,.\, Experience_i}{\sum_{i=0}^{n} Experience_i}$$

Figure 3: Weighted Combined Judgement Forecast Score

$Probability_{qo}$ is the aggregated probability assigned to outcome *o* for question *q*. $Experience_i$ is the experience of respondent *i* and *n* is the number of respondents.

RESULTS

Since the data collected for this study were not normally distributed, this study conducted non-parametric significance tests. We begin by presenting the median quadratic score for each method in Table 1.

Table 1: Median Quadratic Score for each method

	Prediction Market	Combined Judgement		Key Informant		
		Weighted	Simple	Best	Worst	Median
Median Quadratic Score	-15.28	-21.74	-32.65	-300	-300	-300

Hypothesis H1 posits that the forecasting outcome of a prediction market will be more accurate than the forecasting outcome of the post hoc median expert. A Wilcoxon signed rank test supported this hypothesis. The test revealed a statistically significant difference in performance score between the Prediction Market Performance and the Performance of the Median Key

Informant, as selected post hoc, z = -2.275, p< .05 with a large effect size (r = .66). The observed change in performance score between the two methodologies decreased from -15.28 to -300.

The data collected also supports the hypothesis chain expressed in Figure 1. The average quadratic score of each method decreases from the prediction market to the combined judgement methods to the key informant methods. This analysis is supported by a Friedman Test that indicates there was a statistically significant difference in Performance Scores across the five methods (PM, CJFW, CJFS, KIB, KIW χ^2 (4, n=12) = 14.815, p <= 0.001). Inspection of the median values showed an decrease in quadratic scores in line with initial hypothesis (PM = -15.28, CJFW = -21.74, CJS = -32.65, KIB = -300, KIW = -300). In order to investigate the specific hypothesis generated by this analysis, we conducted post hoc analysis with Wilcoxon Sign-rank tests. In accordance with generally accepted procedures, a Bonferroni correction was applied, resulting in a significance level set at p < 0.01 for further hypothesis testing. This testing produces the following results:

Hypothesis 2: A Wilcoxon signed rank test revealed no statistically significant difference in Performance Score between the Prediction Market and CJW. However, the observed change in performance score, decreasing from -15.28 to -21.74 is consistent with the change expected.

Hypothesis 3: A Wilcoxon signed rank test revealed no statistically significant difference in Performance Scores between CJW and CJS. The observed change in the performance score, decreasing from -21.74 to -32.65 is consistent with our hypothesis.

Hypothesis 4: The observed change in Performance Scores between CJS and KIB again matched expectations, with the median score decreasing from -32.56 to -300.

Hypothesis 5: The performance scores between KIB and KIB remained static at -300. This relationship did not reach statistical significance.

CONCLUSION

This study compares the relative forecasting performance of a number of commonly used organisational forecasting methods to that of a prediction market. A number of conclusions emerge from our findings.

This research was contextually situated in a low information heterogeneity environment, where all the forecasters had access to the same information. Our first hypothesis was that a prediction market would have superior forecasting performance to the median individual forecaster as identified post hoc. Our data supports this hypothesis, suggesting that prediction markets have a significant performance advantage in this problem space. Additionally, in this context, our analysis suggests that the advantages of group forecasting should allow both CJF methods and the prediction market to dominate KI approaches. This conclusion is supported by our results. The hypotheses chain identified in this research is supported by the data collected. In general, these results suggest that when forecasting in a low information heterogeneity context, group forecasting methods will outperform KI approaches. These results also bolster suggestions in the literature that the ascribed advantages of prediction markets mean that they are superior to simple combined judgement forecasting approaches in a low information heterogeneity context.

The hypotheses chain examined in this research was created by theoretically identifying the relative performance of forecasting methods. For example the literature suggests that weighting experts' forecasts will produce a better forecast than simply averaging forecasts. The hypotheses chain is supported by the data. The intuitively obvious ways of improving forecasting methods do offer incremental improvements.

However, it is important to note that relative performance is not the sole variable of concern when organisations engage in forecasting. Collecting and aggregating forecasts from a group of participants is likely to consume considerably more time and resources than simply asking one individual to forecast. These costs are magnified when one considers using a prediction market, as in most organisations, such an endeavour is likely to require additional overheads such as recruiting and training participants and software acquisition. Our research demonstrates that different approaches to forecasting do differ in terms of performance, a result of clear relevance in informing the selection of forecasting methods by practitioners. However, in an organisational setting, this information is only one input into a decision making process that must also account for the actual benefits that will accrue from improved forecasting and the costs associated with obtaining those improved forecasts.

This research project begins the process of developing a more nuanced understanding of the strengths and weaknesses of prediction markets as tools for supporting organisational forecasting. Considerable work remains to be undertaken if this goal is to be meaningfully achieved. This study focussed on investigating prediction market performance in a specific context, namely that of low information heterogeneity. Our results show that the information environment has an impact on the relative performance of prediction markets vis-à-vis other forecasting methods. More generally, greater insight into how prediction markets performance varies across different contexts will offer re-assurance and guidance to practitioners who are considering using them to supplement or replace existing organisational forecasting approaches. Empirical research investigating how prediction market performance varies in different contexts such as high information heterogeneity, as well as how performance varies according the number and attributes of participants, will provide the evidence required to develop this deeper understanding.

This study investigated the forecasting performance of prediction markets compared with a number of commonly used forecasting approaches. However, the approaches investigated in this study are far from exhaustive.

Other approaches such as the Delphi Method or the Nominal Group Technique can be used as forecasting tools. Some of these approaches explicitly allow for communication between forecasters, an attribute which may mean their performance matches or even exceeds the performance of prediction markets in low information heterogeneity environments. Further studies which investigate, in an ecologically valid manner, the relative performance of prediction markets compared with such approaches will also offer valuable guidance to academics and practitioners alike.

BIBLIOGRAPHY

Anderson, L. R., & Holt, C. A. (1997). Information Cascades in the Laboratory. *American Economic Review*, *87*(5), 847–62.

Armstrong, J. (2001). *Principles of forecasting a handbook for researchers and practitioners* (1st ed.). Boston MA: Kluwer Academic.

Berg, J. E., & Rietz, T. A. (2003). Prediction Markets as Decision Support Systems. *Information Systems Frontiers*, *5*(1), 79–93.

Boulu-Reshef, B., Comeig, I., Donze, R., & Weiss, G. D. (2016). Risk aversion in prediction markets: A framed-field experiment. *Journal of Business Research*. https://doi.org/10.1016/j.jbusres.2016.04.082

Buckley, P., & O'Brien, F. (2015). The effect of malicious manipulations on prediction market accuracy. *Information Systems Frontiers*, 1–13. https://doi.org/10.1007/s10796-015-9617-7

Chen, K.-Y., Fine, L. R., & Huberman, B. A. (2003). Predicting the Future. *Information Systems Frontiers*, *5*(1), 47–61.

Chen, Y., Chu, C.-H., Mullen, T., & Pennock, D. M. (2005). Information Markets vs. Opinion Pools: An Empirical Comparison. In *Proceedings of the 6th ACM Conference on Electronic Commerce* (pp. 58–67). New York, NY, USA: ACM. https://doi.org/10.1145/1064009.1064016

Chugh, D., & Bazerman, M. H. (2007). Bounded awareness: what you fail to see can hurt you. *Mind & Society*, *6*(1), 1–18. https://doi.org/10.1007/s11299-006-0020-4

Deck, C., Lin, S., & Porter, D. (2013). Affecting policy by manipulating prediction markets: Experimental evidence. *Journal of Economic Behavior & Organization*, *85*, 48–62. https://doi.org/10.1016/j.jebo.2012.10.017

Einbinder, M. (2006). Information Markets: Using Market Predictions to Make Administrative Decisions. *Virginia Law Review*, *92*, 149.

Ellis, D. G., & Fisher, B. A. (1994). *Small Group Decision Making: Communication and the Group Process* (4th ed.). New York: McGraw-Hill Book Company.

Garvey, J., & Buckley, P. (2010). Teaching the Concept of Risk: Blended Learning Using a Custom-Built Prediction Market. *Journal of Teaching in International Business*, *21*(4), 346–357. https://doi.org/10.1080/08975930.2010.526032

Geifman, D., Raban, D. R., & Sheizaf, R. (2011). P-MART: Towards a classification of online prediction markets. *First Monday; Volume 16, Number 7 - 4 July 2011*. Retrieved from http://firstmonday.org/ojs/index.php/fm/article/view/3203/3019

Graefe, A., & Armstrong, J. S. (2011). Comparing face-to-face meetings, nominal groups, Delphi and prediction markets on an estimation task. *International Journal of Forecasting*, *27*(1), 183–195. https://doi.org/10.1016/j.ijforecast.2010.05.004

Graefe, A., & Weinhardt, C. (2008). Long-Term Forecasting with Prediction Markets A Field Experiment on Applicability and Expert Confidence. *The Journal of Prediction Markets*, *2*(2), 71–91.

Gruca, T. S., Berg, J. E., & Cipriano, M. (2008). Incentive and Accuracy Issues in Movie Prediction Markets. *The Journal of Prediction Markets*, *2*(1), 29–43.

Haase, H., & Franco, M. (2011). Information sources for environmental scanning: do industry and firm size matter? *Management Decision*, *49*(10), 1642–1657. https://doi.org/10.1108/00251741111183807

Hahn, R. W., & Tetlock, P. C. (2006a). A New Tool for Promoting Economic Development. In R. W. Hahn & P. C. Tetlock (Eds.), *Information Markets: A New Way of Making Decisions* (pp. 170–194). Washington D.C: AEI-Brookings Joint Center for Regulatory Studies.

Hahn, R. W., & Tetlock, P. C. (2006b). Introduction to Information Markets. In *Information Markets: A New Way of Making Decisions* (pp. 1–12). Washington D.C: AEI-Brookings Joint Center for Regulatory Studies.

Hall, C. (2010). Prediction Markets: Issues and Applications. *The Journal of Prediction Markets*, *4*(1), 27–58.

Hayek, F. A. (1945). The Use of Knowledge In Society. *American Economic Review*, *35*(4), 519–530.

Hitt, M., Black, S., & Porter, L. W. (2005). *Management* (2nd Revised US ed). Prentice Hall.

Horn, C. F., & Ivens, B. S. (2015). Corporate Prediction Markets for Innovation Management. In A. Brem & É. Viardot (Eds.), *Adoption of*

Innovation (pp. 11–23). Springer International Publishing. Retrieved from http://link.springer.com/chapter/10.1007/978-3-319-14523-5_2

Isenberg, D. J. (1986). Group polarization: A critical review and meta-analysis. *Journal of Personality and Social Psychology*, *50*(6), 1141–1151. https://doi.org/10.1037/0022-3514.50.6.1141

Janis, I. L. (1972). *Victims of Groupthink: A psychological study of foreign-policy decisions and fiascoes*. Houghton Mifflin Company.

Jian, L., & Sami, R. (2012). Aggregation and Manipulation in Prediction Markets: Effects of Trading Mechanism and Information Distribution. *Management Science*, *58*(1), 123–140. https://doi.org/10.1287/mnsc.1110.1404

LaComb, C. A., Barnett, J. A., & Pan, Q. (2007). The imagination market. *Information Systems Frontiers*, *9*(2–3), 245–256. https://doi.org/10.1007/s10796-007-9024-9

Laughlin, P. R., Bonner, B. L., & Miner, A. G. (2002). Groups perform better than the best individuals on Letters-to-Numbers problems. *Organizational Behavior and Human Decision Processes*, *88*(2), 605–620. https://doi.org/10.1016/S0749-5978(02)00003-1

Ledyard, J. (2006). Designing Information Markets for Policy Analysis. In *Information Markets: A New Way of Making Decisions* (pp. 37–66). Washington D.C: AEI-Brookings Joint Center for Regulatory Studies.

March, J. G. (1999). *The Pursuit of Organizational Intelligence: Decisions and Learning in Organizations*. Wiley-Blackwell.

Plott, C. R. (2000). Markets as Information Gathering Tools. *Southern Economic Journal*, *67*(1), 2–15. https://doi.org/10.2307/1061610

Remidez, H., & Joslin, C. (2007). Using Prediction Markets to Support IT Project Management - Research in Progress. In *International Research Workshop on IT Project Management 2007*. Montreal, Quebec, Canada: Association for Information Systems.

Servan-Schreiber, E., Wolfers, J., Pennock, D. M., & Galebach, B. (2004). Prediction Markets: Does Money Matter? *Electronic Markets*, *14*(3), 243–251. https://doi.org/10.1080/1019678042000245254

Simon, H. A. (1997). *Administrative Behavior, 4th Edition* (4th ed.). Free Press.

Slamka, C., Skiera, B., & Spann, M. (2013). Prediction Market Performance and Market Liquidity: A Comparison of Automated Market Makers. *IEEE Transactions on Engineering Management*, *60*(1), 169–185. https://doi.org/10.1109/TEM.2012.2191618

Spann, M., & Skiera, B. (2003). Internet-Based Virtual Stock Markets for Business Forecasting. *Management Science*, *49*(10), 1310–1326.

Sprenger, T., Bolster, P., & Venkateswaran, A. (2007). Conditional Prediction Markets as Corporate Decision Support Systems An Experimental Comparison with Group Deliberations. *The Journal of Prediction Markets*, *1*(3), 189–208.

Tziralis, G., & Tatsiopoulos, I. (2007). Prediction Markets: An Extended Literature Review. *The Journal of Prediction Markets*, *1*(1), 75–91.

Van Bruggen, G. H., Spann, M., Lilien, G. L., & Skiera, B. (2010). Prediction Markets as institutional forecasting support systems. *Decision Support Systems*, *49*(4), 404–416. https://doi.org/10.1016/j.dss.2010.05.002

Wolfers, J., & Zitzewitz, E. (2006). Five Open Questions About Prediction Markets. In *Information Markets: A New Way of Making Decisions* (pp. 13–36). Washington D.C: AEI-Brookings Joint Center for Regulatory Studies.