International Journal of Multidisciplinary and Current Research ISSN: 2321-3124
Research Article Available at: http://ijmcr.com

Identifying and Measuring Expertise: Insights from Prediction Market Research

Max Werner and Hubertus Hofkirchner

HSU, Hamburg, Germany
Prediki Prediction Services, Vienna, Austria

Accepted 15 May 2016, Available online 22 May 2016, **Vol.4 (May/June 2016 issue)**

Abstract

A series of forecasting experiments with just under 1,000 participating students over a period of more than six months provides new insights into the complex issue of identifying experts and measuring their level of expertise.

Keywords: *Measuring Expertise, Prediction Markets, Forecasting, Experts*

1. Introduction

We encounter experts in one way or another almost daily: on the evening news, at work, and at university. There are experts on finance, politics, medicine, and law. Experts are found in America, Europe, Asia, Australia, and Africa. It seems no matter where we look, experts are everywhere; in every area on our planet, there is expertise. As is so often the case with a supply surplus, it is hard to maintain an overview. The issue of quality comes up. What is the actual expertise behind these self-declared experts? And is this expertise actually in the declared area of expertise or, in retrospect, more a case of self-declaration or, in extreme cases, even a deception?

Misjudging the quality of an expert can cause damage in many areas of life. As a proxy for this scientists have estimated that the damage from bad advice on the pension scheme in Germany costs €50 billion annually.[1] In health care 44,000 to 98,000 deaths annually in the United States of America and Australia are due to treatment failures, and this is only the tip of the iceberg.[2] Scientifically, the actual amount of damage worldwide caused annually by the lack of expertise is still not established; however, we can imagine that this amounts to an enormous sum. Of course, these examples cannot be exclusively attributed to lack of expertise by self-confessed experts, but they clearly show that the quality of experts is of enormous relevance on the one hand, and is difficult to detect on the other. In this context the question arises whether there are possibilities for prevention. But how can we distinguish between actual experts having valuable expertise from self-proclaimed experts of deception? Often these apparent experts are themselves unaware that they do not possess the promised level of expertise in their declared area of expertise.

Countless studies on this problem have already exposed a wide range of perspectives on the subject. With our paper we want to add another perspective: identifying experts through the verification process of a prediction market. For this purpose we have relied on a much-used thesis as an accepted hypothesis. Following our main thesis, according to which the degree of a thematically-linked expertise is defined by how well the expert under scrutiny predicts future events—which are within the scope of his thematic expertise—the difficulty of the investigation lies in the topical binding. In an early study in 1986 about the evaluation of expert forecasts, it was already concluded that experts are always well above average in forecasts, but only if they are interviewed in their field of expertise.[3]

Because of the problem at hand, experiments for the identification and categorization of experts require an explicit forecast question in a one-dimensional environment in order to minimize the influence of external factors. We were able to find such a test environment with a suitable forecast question in the academic context; students participating in a specific course forecast their average grades. Simply stated this means that we arranged for students to predict their grades in various university courses.

In this article we will present the results of an extensive series of experiments on the described forecast goal. The study included just under 1,000 participating students over a period of more than six months. The results provide new insights into the complex issue of identifying and ranking experts based on measurable empirical evidence.

2. Methodology

2.1 Research subject

To add a new perspective to current knowledge on identifying experts—through the lens of a prediction market—we took a closer look at the subject of our expert identification research. According to current research, experts define themselves through the fact that they have invested a disproportionate amount of time and work in developing their ability.[4] The identification of experts is therefore inevitably bound to a certain observation period and corresponding activities in the said area of expertise. In this context, the question arises of whether it would make more sense to have a way of measuring the degree of expertise.

Practically this would mean that anyone, regardless of time frame or specific requirements, could be tested on their level of expertise in a specific area of expertise. As an example, one can imagine a man in his mid-30s who learns his first moves on a chessboard during a chess course. This probably represents an expertise at the lower end of the scale where the top end is likely to be a chess AI. Nevertheless, it is understandable that the 30-year-old man can improve his position on the scale through consistent learning and practicing. In the course of this study, we want to check whether the use of a prediction market for identifying experts can accurately deliver this scale.

Main hypothesis: *The degree of topical expertise can be identified by how well the expert under scrutiny predicts future events, which are forecast within the thematic spectrum of the investigated expertise.*

Based on our assumption that we once again would like to represent as exploratory, the two central questions we want to answer in the course of our study are:

1. Does the performance of a prediction market participant on a particular forecast question determine subsequent performance on similar issues?
2. What are the factors contributing to the performance of a prediction market participant for the scenario described in question 1?

We are aware that this is an exploratory study with the aim of initiating an academic discourse. There are limits to our study—which we have knowingly taken into account—due to the experimental nature of the subject being investigated. Therefore, we have limited ourselves to experts on knowledge, according to Hayek's understanding, and have systematically excluded other forms of expertise.[5] The significant advantage of such a limitation resides in the simple integration and comparability of the methodology of the prediction market.

2.2 Research structure

According to the two questions formulated above, our research was divided into two consecutive experiments. Both experiments had the same framework conditions, which are described in detail below. The division of our research is of particular relevance to guarantee comparability and minimize the influence of external interference.

2.2.1 Structure

In the course of the first experiment, 12 prediction markets for 12 university courses were opened at a university campus in Germany. The students enrolled in the courses—just under 600 students—had an opportunity to predict the average course grades for each course. Students were allowed to forecast for courses on which they themselves were enrolled as well as on all other courses of the experiment. The aim of the first experiment was to find out whether correlations between the individual prediction performances of students could be established. An evaluation about the percentile rank of individual students was methodically planned.

The second experiment went a step further. With 12 prediction markets in only 4 university courses and almost 400 participating students, there was a clear discrimination. Based on the results of the first experiment, specific factors that were considered to have an impact were studied separately from each other. This allowed for quantifying the influence of each factor and explaining the actual forecast performance of students in detail. Similar to the first experiment, within the selected groups individual forecast performance was evaluated based on the percentile ranks. With completion of the second experiment, the intent was to have examined identifying the experts (first question) as well as the categorization of the identified expertise (second question).

2.2.2 Framework conditions

Prediction target: The average course grade of each tested course. The average course grade was composed of the average exam results of all enrolled students.

Participants: Students at a university campus in Germany. All students were financially secure and had no distracting ancillary employment. Interference factors outside the academic operation were therefore considered negligible.

Incentive mechanism: 30 Euros were given to the best performers in any market. The three best students per market received certificates for excellent performance. All participating students who were able to demonstrate active participation were given a certificate of participation.

Prediction market: The platform from Prediki Prognosedienste GmbH was used for conducting the experiments. Contracts with final payment mode were selected for the experiment.[6]

Prediction question: Multiple-choice questions were used for the prediction market. For each question there were

10 response options, which corresponded to the usual university grade range (see Figure 1). The probability of each possible response changed as a result of the choice made by the students, based on a price algorithm.[7]

Table 1 Possible response options

1.0 to 1.3	2.0 to 2.3	3.0 to 3.3	4.0 to 5.0
1.3 to 1.7	2.3 to 2.7	3.3 to 3.7	
1.7 to 2.0	2.7 to 3.0	3.7 to 4.0	

Price algorithm: A so-called "automated market maker" was used for calculating the price or probability. Due to the relatively small number of market participants—in comparison to the stock exchange—and the corresponding liquidity restriction, an automatic execution of transactions using an algorithmic "market maker" was useful for facilitating fluid trading.[8]

2. Evaluation of Results

2.1 First prediction market experiment

Does good forecasting performance in a prediction market also indicate comparably good forecasting performance in another prediction market with similar questions? This question has been investigated in the first experiment. This chapter is divided into two sections. The first section examines the relationship between market participation and forecast performance. In the second section the relationship between consecutive forecast performances is analyzed.

3.1.1 Active market participation

As a prerequisite for the study of identifying experts, the overall forecast accuracy of the first experiment was evaluated. As shown in Table 1, the average standard error oscillates between 0.08 and 1.41, with 32 to 69 active performers and 175 to 422 transactions per market. The calculation of the average standard error was based on the median forecast of the cumulative probabilities of all response options per market. The median was chosen because it is particularly resistant to large variations in transactions. That is not negligible, especially due to the nature of the participating students.[9] All in all, the framework conditions can be considered as stable according to current scientific understanding [10].

For the evaluation of the first prediction market experiment in terms of the given question, all participants were sorted according to their respective forecast performance. This means that the correlation between the number of active participations in prediction markets and the position in the ranking must now be determined.

Since the number of participants in the various markets was variable and comparability of all markets is a prerequisite for the investigation, the regular ranking was expanded by percentile ranking.

Table 2 Prediction results and average standard error for all 12 prediction markets of the first experiment

Course	Final prediction	Average course grade	Average Standard Error
Advanced Macroeconomics	2.93	2.87	0.08
Market and State	3.17	3.10	0.10
Game Theory	2.95	2.70	0.26
Advanced Macroeconomics	3.24	3.39	0.39
Macroeconomics	3.33	3.62	0.51
Mathematics II/III	3.62	3.86	0.60
Advanced Empirical Economic Research	3.21	3.60	0.63
Introduction to Social Psychology	2.82	2.05	0.65
Macroeconomics	3.43	4.14	1.05
Market and State	3.14	4.16	1.18
Responsible Project Management	2.24	1.54	1.18
Economics in the Public Sector	3.09	4.40	1.41
		Average	0.67

The resulting table of participating students and their percentile rank achieved per market could then, in retrospect, be transformed into a distribution scheme as shown in Table 2. At this point it is important to mention that a possible positive correlation only means a positive relationship in a distribution scheme in accordance with the above formulated questions. Without a distribution scheme, a negative correlation would have been expected. This is in particular due to the fact that a lower percentile rank represents good forecast performance.

Finally, the correlation coefficients between the distribution and number of active markets, number of transactions, and the average number of transaction per market were calculated. The results are shown in Table 2. While no definite statement can be derived from the investigation of relationships between the number of active markets (i.e. markets on which at least one transaction was made) and the forecast performance (primarily due to the comparatively similar positive correlation in all areas of distribution), there is a clear outstanding distribution correlation for the number of transactions and the average transactions per market. In both contexts a significantly positive correlation for the first 25 percentiles is apparent, while in the other percentiles there remains a negligible positive/negative correlation. Therefore, it can be assumed that there is a significant positive correlation (0.64) between the number of transactions that a participant has performed and his forecast performance. Similarly, the somewhat less pronounced positive correlation (0.32) between the average number of transactions per active market and forecast performance of a participant was formulated. Significance of the results can be assumed given the high number of value pairs (808 in all) and the causal logic of results.

Table 3 Correlation coefficient of the first experiment

	1st quartile	2nd quartile	3rd quartile	4th quartile
Active markets	0.58	0.44	0.60	0.53
Transactions	0.64	0.13	0.23	0.22
Average transactions	0.32	-0.05	-0.05	-0.08

For the evaluation of the first prediction market experiment in terms of the given question, all participants were sorted according to their respective forecast performance. This means that the correlation between the number of active participations in prediction markets and the position in the ranking must now be determined.

Since the number of participants in the various markets was variable and comparability of all markets is a prerequisite for the investigation, the regular ranking was expanded by percentile ranking. The resulting table of participating students and their percentile rank achieved per market could then, in retrospect, be transformed into a distribution scheme as shown in Table 2. At this point it is important to mention that a possible positive correlation only means a positive relationship in a distribution scheme in accordance with the above formulated questions. Without a distribution scheme, a negative correlation would have been expected. This is in particular due to the fact that a lower percentile rank represents good forecast performance.

Finally, the correlation coefficients between the distribution and number of active markets, number of transactions, and the average number of transaction per market were calculated. The results are shown in Table 2. While no definite statement can be derived from the investigation of relationships between the number of active markets (i.e. markets on which at least one transaction was made) and the forecast performance (primarily due to the comparatively similar positive correlation in all areas of distribution), there is a clear outstanding distribution correlation for the number of transactions and the average transactions per market. In both contexts a significantly positive correlation for the first 25 percentiles is apparent, while in the other percentiles there remains a negligible positive/negative correlation. Therefore, it can be assumed that there is a significant positive correlation (0.64) between the number of transactions that a participant has performed and his forecast performance. Similarly, the somewhat less pronounced positive correlation (0.32) between the average number of transactions per active market and forecast performance of a participant was formulated. Significance of the results can be assumed given the high number of value pairs (808 in all) and the causal logic of results.

The results of the correlation coefficients of the first experiment seem promising, taking into account the fact that the mere passive participation in a prediction market experiment does not yet suggest an above average forecast performance. Only in the case of active participation, represented by a corresponding transaction volume, can a better than average forecast performance be assumed.

In addition to the calculation of correlation coefficients, to understand any occurring relationship, it is relevant to look at the proportional distribution according to the number of markets in which a performer actively participates. The development of the percentage distribution in the percentile ranks shown in Figure 2 clearly shows that a shift in distribution takes place with an increasing number of actively traded markets.

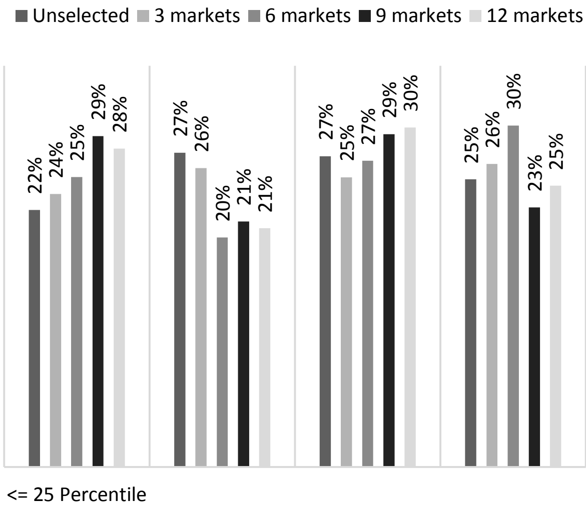

Fig. 1. Distribution shift

With an increasing number of actively traded markets, a percentage increase in the <= 25 percentile distribution is to be expected. On average, each market can be expected to rise by 0.53%. Between the 25th and 50th percentile a proportional decline of 0.72% per actively traded market can be assumed. The trend between the 50th and 75th percentile is almost constant. Here, a proportional decrease of 0.07% is expected. Finally, in all the percentile ranks >75, an increase of 0.23% can be expected.

From the results of the first experiment it can be assumed that there is no statistically significant relationship between the sheer number of markets in which a performer participates and good forecast performance. This can be explained in a distribution scheme by an increasing formation of extremes. With an increasing number of actively traded markets, the proportionate number of very good participants as well as that of very poor participants grows at the expense of the statistical midfield. On the other hand, there is a statistically significant correlation between the number of transactions in a prediction market experiment and the forecast performance, as well as between the average number of transactions in a prediction market experiment and the forecast performance.

3.1.2 Recurring forecast performance

But what about the relationship between good forecasting performance in a prediction market and good forecasting performance in following markets? To examine this question more closely we calculated the correlation of all combinations of percentile rankings of the 12 forecast markets between each other and arrived at the conclusions illustrated in Table 3.

Table 4 Average correlation coefficient of all the prediction market pairs

Participants	Average correlation coefficient	Value pairs
without discrimination	0.25	6666
>5 transactions	0.41	3300
>10 transactions	0.36	1711
>15 transactions	0.17	280

The correlation of the percentile rankings without participant selectivity is relatively weak, but clearly positive. Thus, a weak but significant positive correlation between the percentile rankings of consecutive prediction market experiments, as based on data, is detectable. If a participant thus reaches a good percentile ranking through good forecasting performance, it is likely that this participant will also provide good performances in other prediction markets.

This relationship becomes even clearer when isolating participants with more than five transactions. Due to the nature of the research concept, many students took part in the experiment without claiming long-term participation. The described problem may be counteracted by selecting the analyzed percentile rankings of participants with over five transactions. Thus, the resulting correlation of the active participants is also significantly more pronounced and it can be assumed that there is a medium-strong connection. The result of the discrimination is equivalent to halving the relevant value pairs and thus reduces its significance. We consider correlation coefficients with more than 800 value pairs to be a significant result. Only two findings can be derived for stronger discrimination of participants with 10 or 15 transactions. On the one hand, the correlation does not increase through further discrimination. On the other hand, further discrimination of participants with more than 12 transactions means that the number of value pairs goes down to less than 800 and must therefore be considered as not significant.

In summary, for recurring forecast performance it can be considered that a statistically significant medium-strong (0.41) positive relationship exists between forecast performance and the following forecast performances. The participants of a prediction market thus have a higher probability of repeatedly predicting well if they could already achieve a good forecast performance in the previous prediction markets.

2.2 Selected prediction market experiment

First insights into correlations between prediction market participation and successful forecast performance was already proven with the first experiment. In the second experiment we went one step further and examined what factors affect good forecasting performance and to what extent. To do this, we looked once more at the research concept. What are the possible parameters? Due to the structure of the experiment, interference factors outside regular academic operation could be ignored. All surveyed students were financially secure campus students and without secondary activities. The question therefore was: What can affect the successful forecast of average course grades?

3.2.1 Definition of the selection criteria

To answer this question we assumed Hayek's understanding of knowledge, which defines three types of knowledge as parameters having potential influence.

Static knowledge: Forecasts on the basis of static knowledge are not linked to active information gathering. In our understanding static knowledge is defined by means of a general, already existing knowledge on a specific subject area. Independent from new information gathering, static knowledge is quickly available and applicable to many questions within a specific area. The disadvantage of static knowledge is in the detail and the lack of dynamism. Trend-setting information can be missed by mere static knowledge and can therefore not be exploited.

In relation to our prediction experiment, static knowledge can be considered as a general understanding of the subject area, the lecturers, and the academic environment of the course in which the average course grade is to be predicted. Studied prediction markets where static knowledge is the dominant form of knowledge are identified by the following characteristics:

- The students have no previous experience with prediction markets.
- The students are already enrolled at the university.
- The degree of expression of static knowledge depends on the duration, for which students are already enrolled at the university.
- The students do not even attend the classes for which they provide forecasts.

Dynamic knowledge: In contrast to static knowledge, dynamic knowledge is described—in our understanding—as active information gathering. Given this active information acquisition, dynamic knowledge is highly specialized and especially difficult to convey. Due to the high degree of specialization, there is usually no basis for comparison, which can lead to strong over/undervaluation of information. The most predominant advantage of dynamic knowledge is the fast reaction time and ability to adapt to changes in information.

In relation to our prediction market experiment, dynamic knowledge is defined by the active participation of students in a course for which they forecast an average course grade. Studied prediction markets where dynamic knowledge is the dominant form of knowledge are identified by the following characteristics:

- The students have no previous experience with prediction markets.
- The students attend the classes for which they provide forecasts.
- The degree of dominance of dynamic knowledge is dependent on the level of static knowledge that students naturally acquire with each passing semester.

Mechanical knowledge: After static and dynamic knowledge—which are relatively strongly associated with the prediction target of a forecast—mechanical knowledge is described as knowledge, understanding, and information on the prediction market system. Almost independent of the actual forecast target, mechanical knowledge is universally applicable; however, it requires a comprehensive understanding of the prediction instrument. Similar to the regular stock market on which the largest part of all transactions is technical (i.e. not conditioned by content in terms of information, but by the counter value of the shares), mechanical knowledge is used primarily for arbitrage and price-smoothing.[11]

In relation to our forecast market experiment, mechanical knowledge is described by prior successful experience in dealing with prediction markets. Studied prediction markets where mechanical knowledge is the dominant form of knowledge are identified by the following characteristics:

- The students do not even attend the classes for which they provide forecasts.
- The students were the 20 most successful participants in the first prediction market experiment.
- The degree of dominance of mechanical knowledge is dependent on the level of static knowledge that students naturally acquire with each passing semester. To prevent a possible conflict of dominance, no member of the so-called performance group enrolled for more than four semesters at the university was included.

Similar to the first experiment, the students were asked to provide forecasts for the average course grade of four university courses. The difference was in the detail and in the arrangement of the forecast markets. Thus, all prediction markets were investigated separately. This separation was used to ensure that only students with certain properties performed on the respective markets. Through this arrangement the students could assign the selected markets to a dominant form of knowledge. This allowed for evaluation according to the dominant forms of knowledge.

3.2.2 Knowledge form-dependent forecast performance

Due to strong discrimination, participation in our second experiment was significantly lower. Although nearly 400 students took part in the experiment, no significant share of transactions/participants could be achieved, particularly with students at the seventh semester level. The results shown in Table 4 can therefore only be understood as a trend statement, specifically for students at the seventh semester level.[12]

Table 5 Forecast results and average standard error for the second experiment according to types of knowledge

Knowledge	Trans-actions	MSF Start	Avg. MSF	MSF End	Delta
Mechanical	284	1.26	1.08	0.94	-0.33
Dynamic	494	1.33	1.21	0.98	-0.35
Static (1).	331	0.85	0.90	0.99	0.14
Static (4).	193	0.89	0.87	0.90	0.01
Static (7).	63	1.00	1.19	1.20	0.20

Static knowledge shows the lowest average standard error of 0.99 (0.89 without seventh semester students). Intra-static knowledge shows that with an increase in the duration of the stay, the forecast accuracy on average can be improved. After static knowledge, mechanical knowledge has the lowest average standard error of 1.08. The tail light of the experiment is dynamic knowledge with an average standard error of 1.21.

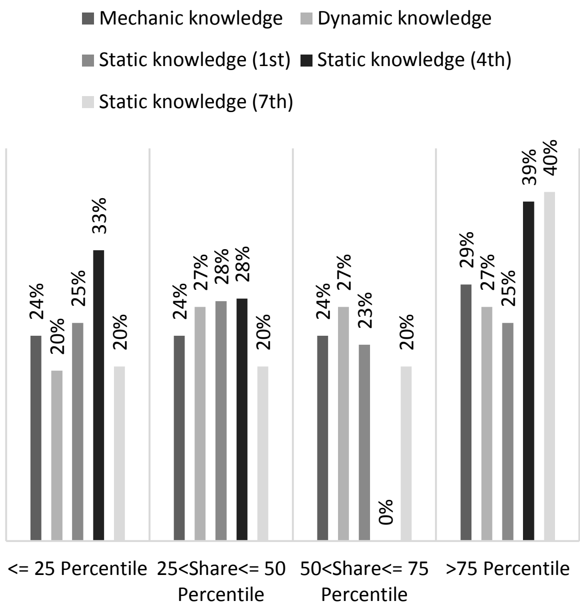

Fig.2 Distribution of the percentile rankings for the various forms of knowledge

The activity of information gathering shows up when viewing the deltas between initial and final medium standard error. While with mechanical knowledge a different type of information is processed, dynamic (improvement of 0.35) and mechanical (improvement of 0.33) knowledge results in a similar activity. The missing dynamics of static knowledge is even clearer in the results. Over the period of the experiment the average standard error on markets with static knowledge as the dominant form of knowledge, deteriorates on average by 0.11 (or 0.07 without seventh semester students). The results show a clear causal relationship with popular prediction market research.[13]

Strongly pronounced static knowledge provides the largest contribution in the formation of correct forecasts. In second place we have mechanical knowledge, while dynamic knowledge takes up the last position. But how does this finding affect identifying the experts? How likely is it that a participant with a dominant form of mechanical knowledge achieves a percentile rank below 25?

Similar to the first experiment, these questions can be answered by looking at the percentile distribution of the different forms of knowledge. The results shown in Figure 3 do not seem surprising. The probability of being within the first 25 percentiles is highest with a strong static knowledge, followed by mechanical and finally dynamic knowledge. The ranking is different in the next 25 percentiles. Although static knowledge is still always the most represented with 28%, dynamic knowledge follows immediately with 27%. Mechanical knowledge is significantly behind with 24%. Between the 50th and 75th percentile, the distribution corresponds to causal logic. Static knowledge is comparatively the least represented. The largest percentage of representation is dynamic knowledge, followed by mechanical knowledge with a 3% difference. The probability of being within the last 25 percentiles is again highest with a strong static knowledge with an average of 35%. Mechanical knowledge comes 6% behind and finally dynamic knowledge with 27%.

Knowledge of the knowledge form-dependent distribution and shares for forecast accuracy is followed by an investigation of the relationship of previous forecast performance and future forecast performances with the respective forms of knowledge. As shown in Table 5, none of the calculated correlation coefficients can be described as significant. Even the number of value pairs for dynamic knowledge falls by more than 50% under the required minimum number of value pairs.

Table 6 Average correlation coefficient of all dominant forms of knowledge (>5 transactions)

Knowledge	Average correlation coefficient	Value Pairs
Static (1).	-0.05	72
Static (4).	0.40	36
Static (7).	-0.40	20
Dynamic	0.36	360
Mechanical	0.23	85

As a trend and when disregarding the results of the first and seventh semester students, static knowledge again dominates the rankings of the forms of knowledge; although this is not the remarkable feature found in the data. Static knowledge is also below the average correlation coefficient of the first prediction market experiment with a positive correlation maximum of 0.4. This fact could indicate that the number of used forms of knowledge increases the likelihood that a participant reaches a similar forecast result as in a previous forecast. More conclusions in particular on ranking the forms of knowledge cannot be derived from this calculation due to the problematic evidence.

4. Analysis

Much data and many numbers were statistically processed and presented in the previous chapter. But what about the two above-formulated questions and resulting conclusion on the subject matter of identifying experts? The core findings after evaluating statistical data regarding this relationship will be discussed in the course of this chapter.

4.1 Identifying experts

An expert is primarily defined as a person who works and trains more than average on his abilities. This is one of the most important scientific findings about experts in the last decade.[4] This realization is reflected in our results in which the number of transactions with a significant positive correlation (0.63) with good forecasting performance within the first 25 percentiles can be seen as a quantitative measure for the learning effect. The mere experimental participation in many markets with only one transaction could therefore not have any significant outstanding, positive correlation. However, with an increasing number of markets, a distribution shift within the meaning of the following percentage changes is to be expected. The first and last 25 percentiles grow by 0.53 or 0.23%, while the middle 50 percentiles decrease by 0.72 and 0.07%. In the long term, the mere participation, without learning effect, leads to the formation of extremes in the sense of percentage distributions under the first and last 25 percentiles. This is a finding which can be transferred to current knowledge on experts.[14] The formulated thesis is also confirmed in terms of the correlation of the percentile rankings of all 12 forecast markets between them. The highest significant correlation among the selected percentile rankings of all 12 markets for participants with more than five transactions (the average number of transactions in the first prediction market experiment was three transactions per participant and market) is 0.41. Without such a selection, we can observe a halving of the correlation coefficient to 0.25.

Thus, as formulated above in terms of concept, to be able to identify experts with a prediction market the forecast performance must be seen in connection with the number of transactions. Good forecasting performance without a larger number of transactions poses a significantly higher risk of incorrectly choosing an expert. Actual experts have a higher probability of providing good forecasting performance as well as above-average transaction numbers. In terms of specifically identifying an expert, this means that prediction market participants with an above-average transaction count and good forecasting performance (within the first 25 percentiles) have a 39% higher probability of providing good forecasting performance in a subsequent forecast market experiment. In the event of an equally distributed probability the probability of being within the first 25 percentiles, is 25%. the probability of being within the first 25 percentiles with an increased transaction number and previous good forecast performance is significantly higher with 64%.

4.2 Categorizing expertise

While we have already shown how the probability of identifying experts can still be significantly increased, the question initially formulated that relates to the area of expertise of the expert arises. In connection with our prediction market experiment, we identified three forms of knowledge as specialized areas of expertise: dynamic, static, and mechanical knowledge. Different results to the first prediction market experiment can be identified by subdividing the participants of a prediction market experiment according to those specialties. These variations will be described in detail in the next section and classified with regard to the category of expertise.

The first and most obvious difference was already apparent when considering the general market-linked forecast accuracy. It was on the markets in which the dominant knowledge form was static knowledge that the lowest average standard error (0.89-0.99) is measured. In the middle field of forms of knowledge we find mechanical knowledge with an average standard error of 1.08, followed by dynamic knowledge with 1.21. The specificity of this knowledge lies less in the various forms of knowledge as in the fact that all average standard errors have been situated above the average standard error of the first experiment. Therefore, it can be assumed that only the combination of the different forms of knowledge leads to a higher level of forecast accuracy. This thesis is also confirmed by the correlation coefficients of individual knowledge form-specific percentile ratings with each other. Here too a difference between the relationships is found, but no higher relationship than in the first prediction market experiment. Participants with more than one strong knowledge form can thus not only achieve better forecasts, as a matter of principle, but also with a higher probability. An isolated use of static knowledge reduces the forecast accuracy by 32.3% and the probability of a good forecast by 2.5%. An isolated use of mechanical knowledge reduces the forecast accuracy by 37.9% and the general probability of a good forecast by 43.9%. In isolated use dynamic knowledge leads to a reduced (by 44.2%) forecast accuracy and a reduced probability of occurrence for good forecasts of 12.2%.

In conclusion, it can be derived from the collected data that with categorization of examined forecast market participants regarding the dominant knowledge form, the accuracy of future forecast results would improve significantly. This means that a participant with an above-average transaction count and good forecasting performance should be examined with respect to his active forms of knowledge. Depending on the result, the 39% increased probability for good future forecasts can be adjusted by the displayed changes to the probability of occurrence. For example, a participant with dominant mechanical knowledge, an above-average number of transactions, and previous good forecast performance has a probability of 25% + 39% - 43.9% = 20.1% to succeed.

Conclusions

The subject field of identifying experts and measuring expertise needs investigating, but is also complex. Countless studies have already delivered a wide range of perspectives on this topic. In the course of our experiments using the prediction market method we can add a further perspective to the basic understanding of identifying experts and ranking their expertise based on their relative performance.

With empirical data from a prediction market we could show a significant positive correlations between past forecast results, transaction numbers, and future forecast results. In addition, we quantified these relationships, and categorized influencing parameters. We investigated expertise relating to forecasts as a combination of static, dynamic, and mechanical knowledge. The highest forecast accuracy, with an average standard error of 0.67, is observed with their combination. Forecasts on the basis of static knowledge only will lose 48% of their accuracy with an average standard error of 0.99. The forecasts are significantly worse if they only rely on mechanical knowledge with an average standard error of 1.08 (-61%) or just dynamic knowledge with an average standard error of 1.21 (-81%).

When considering the probability of occurrence of a forecast performance in the top percentile, the numbers look similar. Without any discrimination it is equally distributed at 25%. If an above-average number of transactions and a previous forecast performance are selected in the first quartile, the likelihood of recurrence of forecast performance in the first quartile increases significantly by 39%. Forecast performance on the basis of static knowledge will only rise by 36.5%. The probability based on mechanical knowledge is worse. Here, the likelihood decreases by 4.9%. In contrast to the forecast accuracy, the probability of occurrence of a forecast performance in the first quartile based on dynamic knowledge increases by 26.8%.

Despite promising results, we should at this point once again be aware of the limitations of our study. The most authoritative problem, already mentioned several times in the previous pages, is the lack of significance in the results from 3 of the 24 prediction markets. The required number of active participants could no longer be guaranteed due to the already described discrimination.

Given these non-significant results, the derived results can only be understood as trend statements. In addition, the study has an exploratory nature and is limited to a relatively simple application area.

Based on our results, we hope to have prompted the academic discourse with regard to further experiments. From our point of view a more extensive forecasting market series in the same field of application for subsequent review and optimization of significance is necessary. Subsequently, other applications with more complex experimental scenarios should be tested and evaluated. If our exploratory results are replicable, experts can be identified and the degree of their expertise can be measured objectively, based their performance on prediction markets in their respective fields of knowledge.

References

[1] A. Oehler, Die Verbraucherwirklichkeit: Mehr als 50 Milliarden Euro Schäden jährlich bei Altersvorsorge und Verbraucherfinanzen. … und Lösungsmöglichkeiten, 2012.

[2] T. W. Johnson, Epidemiology of Medical Errors. 2003.

[3] W. A. Wagenaar and G. B. Keren, "Does The Expert Know? The Reliability of Predictions and Confidence Ratings of Experts," in Intelligent Decision Support in Process Environments, vol. 21, no. 6, Berlin, Heidelberg: Springer Berlin Heidelberg, 1986, pp. 87–103.

[4] K. A. Ericsson, N. F. Charness, and P. Feltovich, PJ & Hoffman RR (Eds.),(2006). Cambridge handbook of expertise and expert performance.

[5] F. A. Hayek, "The use of knowledge in society," The American economic review, 1945.

[6] C. Slamka, W. Jank, and B. Skiera, "Second-Generation Prediction Markets for Information Aggregation: A Comparison of Payoff Mechanisms," Journal of Forecasting, vol. 31, no. 6, pp. 469–489, Sep. 2012.

[7] J. Wolfers and E. Zitzewitz, "Prediction Markets in Theory and Practice," Mar. 2006.

[8] A. Othman, "Automated Market Making: Theory and Practice," pp. 1–212, May 2012.

[9] L. Von Mises and B. Mayes, Human action. 1990.

[10] J. Wolfers and E. Zitzewitz, "Prediction Markets," May 2004.

[11] W. Sharp, G. Alexander, and J. Bailey, Investments. Prentice-Hall International, 1999.

[12] J. D. Christiansen, "Prediction markets: Practical experiments in small markets and behaviours observed," The Journal of Prediction Markets, pp. 1–25, May 2007.

[13] J. E. Berg, F. D. Nelson, and T. A. Rietz, "Prediction market accuracy in the long run," International Journal of Forecasting, vol. 24, no. 2, pp. 285–300, Apr. 2008.

[14] M. T. H. Chi, R. Glaser, and M. J. Farr, The Nature of Expertise. Psychology Press, 2014.