ACCURACY AND BIAS IN EUROPEAN PREDICTION MARKETS

Sveinung Arnesen[1],

Department of Comparative Politics, University of Bergen, Norway

Oliver Strijbis

Department of Comparative Politics, University of Hamburg, Germany, and Wissenschaftzentrum Berlin – WZB, Germany

Received: 12th March 2014/ Accepted: 15th September 2015

Abstract. *Prediction markets have a good overall track record in accurately predicting election outcomes. In this paper, however, we observe a bias in the aggregate predictions of election and referendum outcomes in three European countries. Analyzing 62 vote share contracts from Norway, Germany, and Switzerland, we find that small-sized vote share contracts systematically are predicted higher than what the actual outcome is, while large-sized vote share contracts in contrast tend to be underpredicted. We discuss three possible explanations for the bias. These are, firstly, the horse-race logic of political campaigns which tends to portray winning margins to be narrower than they actually are; secondly the winner's curse in combination with cash constraints for the traders in the market; and thirdly, the properties of the logarithmic market scoring rule (LMSR) which serves as an automated market maker. The data supports the last explanation, namely that the LMSR market maker causes the predictions to be more distant from the extremes than the electoral outcomes turn out to be.*

Keywords: *Prediction markets, Election forecasting, Automated market maker, Prediction bias.*

1. INTRODUCTION

The modern history of prediction markets as tools for forecasting elections starts with the 1988 U.S. Presidential election. Back then, researchers at the University of Iowa set up an experimental market and invited students at the campus to bet on the outcome of the election. There were two questions the students needed to consider: Who would win the election? And, how big a share of the votes would the

[1] Sveinung Arnesen, email: Sveinung.Arnesen@isp.uib.no

candidates receive? In the first Iowa Political Stock Market – later known as the Iowa Electronic Markets (IEM) – the students would call in their bids and offers by phone. They were allowed to invest between $5 and $500 of their own money, and the market ran for six months in the run-up to the election in November.

These two questions were set up with two different types of contracts – winner-take-all contracts and vote share contracts. The winner-take-all contracts are binary contracts that liquidate at $1 if the selected candidate wins the election and $0 if another candidate wins. A trader who possesses ten contracts on the winning candidate will thus have earned $10. For the vote share contracts, the payoff is $1 multiplied by the relative vote share taken by the candidate or party in the election. For example, if the Republican candidate wins 48% of the referendums, the trader receives 48 cents for each share she owns.

The prices of the vote share contracts translate invariantly into predictions about the candidates' share of the votes on election day. If a vote share contract is traded at the price of $0.48, then the prediction of the market is that this candidate will receive 48% of the votes.[2] The last trade represents the current prediction, and as the market is open 24 hours, the predictions are updated continuously. Similarly, the prices of the winner-take-all contracts translate directly into probabilities about who will win the election.[3] If a winner-take-all contract is traded at the price of $0.48, then the market prediction is that this candidate has 48% probability of winning the election.

The vote share predictions are comparable to polls, and their accuracy can be compared directly against the election result. The probability predictions of the winner-take-all contracts are not directly comparable to polls, and their accuracy cannot as easily be determined by comparing one or a few election results. They are however popular among the traders and provide additional interesting information about the candidates' winning chances.[4]

---

2 The IEM operate on the basis of the two-party vote, that is, third party candidates are ignored.

3 Importantly, the question to be determined is who will win the popular majority, and not who will become the next U.S.President. Normally the winner of the popular vote will become President, but – as Al Gore experienced in 2000 – this is not always the case.

4 For example, by analyzing the vote share predictions and winner-take-all predictions in conjunction, one may study the uncertainty of a candidate's winning chances given his predicted vote share level. One can imagine that in an election with a polarized electorate where voters rarely switch parties or candidate preferences, the vote share predictions are less volatile than in an election where the voters are less polarized. All else being equal, a candidate who has a vote share lead in the polarized election will consequently have a higher chance of winning in the election than a candidate with a similar lead in an election with a less polarized electorate.

The organizers of the Iowa Electronic Markets would initially set up a call market with Iowa University students, where the students could phone the organizers to make trades that were processed in bulk. The telephones were quickly replaced by internet-connected computers, which now use the continuous double auction (CDA) process for matching trades: traders bid or ask prices, and whenever there is a match with another participant, a trade goes through. This set-up is similar to that of stock markets and betting exchanges, where traders both offer and accept prices.

In 1988, the predictions were surprisingly accurate. On election eve, the absolute percentage error of the vote share predictions was 0.2 percentage points, while the polls were off by more than 2.5 percentage points. Since then, the Iowa Electronic Markets have continued to outperform the polls. Displaying results from 237 contracts in 49 markets from 13 countries, Berg et al. (2008a) find that the average market error for all the predictions was 1.49 percentage points when measuring election eve closing prices. The corresponding poll error was 1.91 percentage points. Berg et al. (2008b) present an analysis of the long-run forecasting ability of markets relative to polls for the U.S. Presidential elections from 1988 through 2004. For the final 100 days of the election campaigns, the market predictions were closer to the election results on 74% of the days.

Erikson and Wlezien (2008) and Erikson and Wlezien (2012) have in two separate articles demonstrated how polls may be adjusted to anticipatable rating developments during the Presidential campaigns. Such enhanced polls were more accurate than market predictions when they were tested against election outcomes that have already occurred. They demonstrated how polls may perform just as well as the IEM, if the polls are used to make regression model forecasts which take expected poll developments into account.[5] Despite this qualification, market predictions tend to outperform the polls as they are presented to the public.

Furthermore, the organizers reported no obvious biases on an aggregate scale (Berg et al., 2008b). With the data under scrutiny in this paper, however, we do indeed find the market predictions to be biased on an aggregate scale. How and why is this so? The task of this paper is first to present the market predictions of 62 prediction market contracts from elections and referendums in Norway, Germany, and Switzerland. These are real money markets run with software that uses

[5] Arnesen and Bergfjord (2014) do however find that these poll adjustments are less accurate than market predictions when they are tested against the out-of-sample elections of the 2008 and 2012 US Presidential elections.

Hanson's logarithmic market scoring rule as an automated market maker. We find that there is a tendency for the markets to overpredict small-sized vote share contracts and underpredict large-sized vote share contracts. That is, the actual outcomes tend to be more distant from the extreme than the predictions anticipate.

Second, we discuss three potential explanations for the biases we observe. We discuss the hypothesis, introduced by Jacobsen et al. (2000), that a combination of the winner's curse and cash constraints for the traders in the market causes the cheaper contracts to be overpredicted while the more expensive contracts in the market tend to be underpredicted. The data does not seem to support that particular combination of causes. However, the argument that cash constraints have an impact somewhat overlaps the next explanation: the automated market maker implemented in the software has a logarithmic cost function which makes it increasingly expensive to push the predictions towards the set minimum and maximum values, especially in a context of cash constraint as is the case in our data. This may cause the bias in the observed predictions. The third explanation is political. The nature of political campaigns is to make the races look closer than they actually are. This horse-race media logic might easily spill over into the expectations of the participants in prediction markets, since their information might be based to an important degree on media reports. We discuss the alternative explanations and provide some analysis where we have relevant data. But first, we present the prediction markets.

2. THE ELECTIONS, THE MARKETS, AND THE PREDICTIONS

Presented below are predictions of 62 vote share contracts collected from elections and referendums in Germany, Norway, and Switzerland. The data has been gathered over a five-year period and includes prediction market contracts from the 2009 Norwegian parliamentary election, the 2011 Swiss federal election, the 2013 German federal election, and Swiss referendums and initiatives in the 2012-2014 period.

Table 1: Overview of Prediction Market Contracts

COUNTRY (YEAR)	CONTRACT	PREDICTION	RESULT	TYPE
NO (2009)	Labour party	35.0%	35.4%	Election
NO (2009)	Labour/Socialists/ Center Party	49.0%	47.8%	Election
NO (2009)	Liberal party/	4.9%	3.9%	Election
NO (2009)	Liberals/Christian Dem. / Conservatives/Progressives	47.9%	49.5%	Election
CH (2011)	Swiss People's Party	28.6%	26.6%	Election
CH (2011)	Social Democratic Party	23.2%	18.7%	Election
CH (2011)	Green Party	10.4%	8.4%	Election
CH (2011)	Conservative Democratic Party	6.4%	5.4%	Election
CH (2011)	The Liberals	13.7%	15.1%	Election
CH (2011)	Christian Democratic Party	14.0%	12.3%	Election
CH (2011)	Green Liberal Party	5.9%	5.4%	Election
CH (2011)	Other	6.9%	8.1%	Election
CH (2012)	Secondary residences	46.7%	50.6%	Referendum
CH (2012)	Home savings I	48.5%	44.2%	Referendum
CH (2012)	Five weeks of mandatory holiday	39.1%	33.5%	Referendum
CH (2012)	Gambling games	68.4%	87.1%	Referendum
CH (2012)	Book price regulation	45.6%	43.9%	Referendum
CH (2012)	International treaties	43.6%	24.8%	Referendum
CH (2012)	Health insurance	43.9%	24.0%	Referendum
CH (2012)	Home savings II	46.2%	31.1%	Referendum
CH (2012)	Shopping hours I	47.8%	29.3%	Referendum
CH (2012)	Choice of schools	43.0%	18.2%	Referendum
CH (2012)	Cultivated land	54.1%	54.5%	Referendum
CH (2012)	Promotion of youth music	76.9%	72.7%	Referendum
CH (2012)	Mortage regulaton of elderly	49.3%	47.4%	Referendum
CH (2012)	Protection of nonsmokers	46.0%	34.0%	Referendum
CH (2012)	Epizootic diseases act	78.9%	68.3%	Referendum
CH (2012)	Basic level school I	45.0%	28.7%	Referendum
CH (2012)	Basic level school II	54.3%	45.2%	Referendum
CH (2012)	Rentin market	46.8%	52.3%	Referendum
CH (2013)	Family politics	60.9%	54.3%	Referendum
CH (2013)	Manager bonuses I	62.2%	67.9%	Referendum
CH (2013)	Land use planinng	59.0%	62.9%	Referendum
CH (2013)	Personnel management of teachers	47.6%	47.3%	Referendum
CH (2013)	Tax law	59.8%	59.3%	Referendum
CH (2013)	Protection of nonsmokers	54.8%	44.1%	Referendum
CH (2013)	Law of secondary education	68.9%	69.3%	Referendum

Table 1 (follows): Overview of Prediction Market Contracts

COUNTRY (YEAR)	CONTRACT	PREDICTION	RESULT	TYPE
CH (2013)	Public occupation pensions	55.7%	65.2%	Referendum
CH (2013)	Election of the federal government	35.4%	23.7%	Referendum
CH (2013)	Asylum law	66.1%	78.4%	Referendum
CH (2013)	Fight on hooliganism	66.0%	85.5%	Referendum
CH (2013)	Tax on assets	47.2%	38.4%	Referendum
CH (2013)	Compulsory military service	37.0%	26.8%	Referendum
CH (2013)	Epidemics act	63.0%	60.0%	Referendum
CH (2013)	Shopping hours II	56.0%	55.8%	Referendum
CH (2013)	Voting rights of foreigners	37.0%	25.0%	Referendum
DE (2013)	CDU/CSU	40.1%	41.5%	Election
DE (2013)	Social Democratic Party	28.9%	25.7%	Election
DE (2013)	Alliance ’90/The Greens	13.3%	8.4%	Election
DE (2013)	Free Democratic Party	8.7%	4.8%	Election
DE (2013)	The Left	8.9%	8.6%	Election
DE (2013)	Pirate Party	4.5%	2.2%	Election
DE (2013)	Others	8.3%	8.8%	Election
CH (2013)	Manager bonuses II	35.6%	34.7%	Referendum
CH (2013)	Taxation of families	47.8%	41.5%	Referendum
CH (2013)	Road tax disc	50.7%	39.5%	Referendum
CH (2013)	Cantonal fusion (Berne)	37.7%	29.2%	Referendum
CH (2013)	Cantonal fusion (Jura)	72.4%	76.6%	Referendum
CH (2014)	Immigration quotas	46.9%	50.3%	Referendum
CH (2014)	Railway network	59.9%	62.0%	Referendum
CH (2014)	Abortion	34.0%	30.2%	Referendum
CH (2014)	Building law	50.2%	54.6%	Referendum

For all contracts, the participants have invested real money. The amount has varied somewhat, but typically each participant has received € 30 by the organizer to invest in each market. For any given election, its market consists of all the vote share contracts included in the election. For example, the prediction market on the 2009 Norwegian parliamentary election consisted of four contracts. Regarding the referendums, each market consists of the contracts for the referendums that are held in the same month. Typically, between 50 and 100 traders participated in each market, and the markets were open during the last few months of the run-up to the election/referendum.

All the contracts have been run with the same prediction market software, which was originally developed for a PhD project on election forecasting methods (Arnesen, 2012).[6] The software employs the logarithmic market scoring rule (LMSR), invented by Robin Hanson (Hanson, 2003; Hanson, 2007), as an automated market maker.

As mentioned, the Iowa Electronic Markets use the continuous double action (CDA) procedure, which appeals to operators of real-money markets as it poses no financial risk to the organizer (Spann and Skiera, 2003). Its major disadvantage is the risk of illiquid markets. The thin market problem occurs when few traders care to participate since there are so few other traders, which in the long run leads to the breakdown of the market (Pennock and Sami, 2007). Markets with few traders – a typical situation in election prediction markets – are particularly vulnerable because this leads to low liquidity and few trades in the market (Hanson, 2003). Hence, most prediction markets nowadays prefer to facilitate trading by introducing an automated market maker into the software.

An automated market maker is an algorithm that automatically offers a new price for the contract after a trade has been made. The organizer then ensures there is a price offer for any trader to take at any time. The obvious benefit is that the liquidity in the market is infinite. The LMSR has become the most popular automated market maker among prediction market organizers (see for example Berg and Proebsting, 2009).[7] The reason for this is its simple yet proper design. Sliding on a logarithmic scale normally between 0 and 100, Hanson´s scoring rule works like a two-sided market maker, thus allowing traders to both sell and buy contracts at will. The traders must depart from the current price, and how far they can push the price in either direction depends on what level the price is at, how much money the trader has, and how much the organizer has decided that one unit of money should affect the price at any given level.

Since their outset, prediction markets have grown beyond forecasting election outcomes and are currently used to predict a range of other topics. Hanson’s automated market maker is widely used in prediction market software due to its easy implementation and the desirable properties mentioned above. For the most part, this automated market maker is used for probability markets, where an event

---

6 An elaborate study of the 2013 German Federal election is provided in Strijbis and Schnapp (2015).

7 An alternative is the dynamic pari-mutuel market maker developed by David Pennock (2004). Dynamic pari-mutuel markets combine the characteristics of the aforementioned continuous double auction and pari-mutuel markets. Also, variations of Hanson’s market scoring rule have recently appeared in the literature (Othman et al., 2013).

either occurs or does not occur. When it comes to elections, however, forecasting vote shares are of pivotal concern. In multiparty systems, for instance, what the public wants to know is how large a share of the vote each party will collect. Poll ratings, which are the usual information the public uses, reflect a current snapshot of the support for each party that competes for seats in the legislature. The same applies for upcoming referendums and initiatives, where pollsters measure the share of citizens who are for or against the upcoming questions to be considered.

Using the LMSR on vote share contracts is not widespread, but has been suggested in various informal fora.[8] To our knowledge this paper has the largest set of real world vote share predictions made with prediction markets that use the LMSR. For most markets, the price range had a minimum of zero and maximum of 100%. However, for one contract on the 2009 Norwegian market (the Liberals), the price range was 0-10%, and for all the contracts in the 2003 German election market the range was 0-60%.

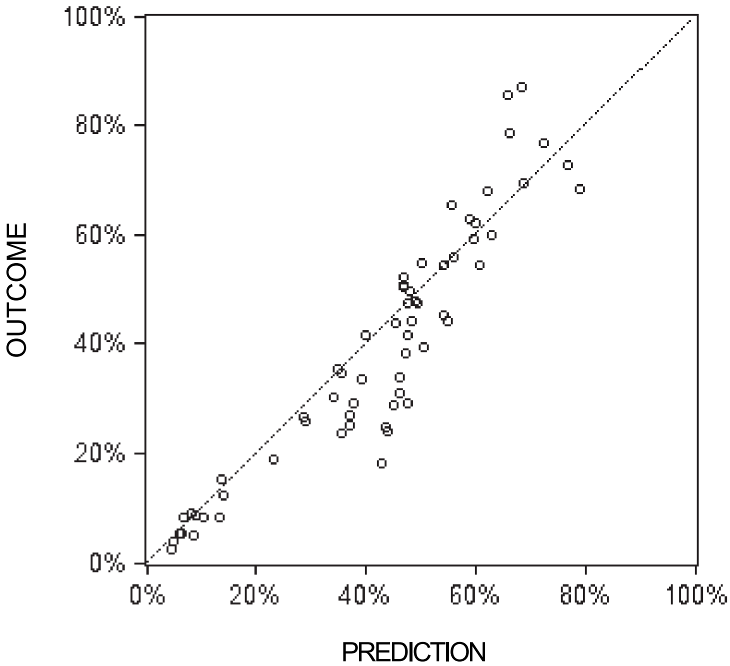

Figure 1: Predicted Versus Actual Outcomes

---

8 See, for example, the discussion in blog post by David Pennock (2006).

Figure 1 displays the calibration of predicted share of yes-votes and the actual outcomes. It shows that more contracts are predicted to have a higher share of yes-votes than what they actually received, so there is a general tendency to overpredict the share of yes-votes. The errors – whose sum should be zero if they are unbiased – amount to 172 percentage points, with a mean of 3.4 (Table 2). This tendency is particularly strong for the contracts that were predicted to end up between 30 and 50%. However, when the errors are measured relative to the sizes of the predicted vote shares, it is clear that the contracts with smaller predicted vote share sizes are about equally far off the mark (Figure 2).

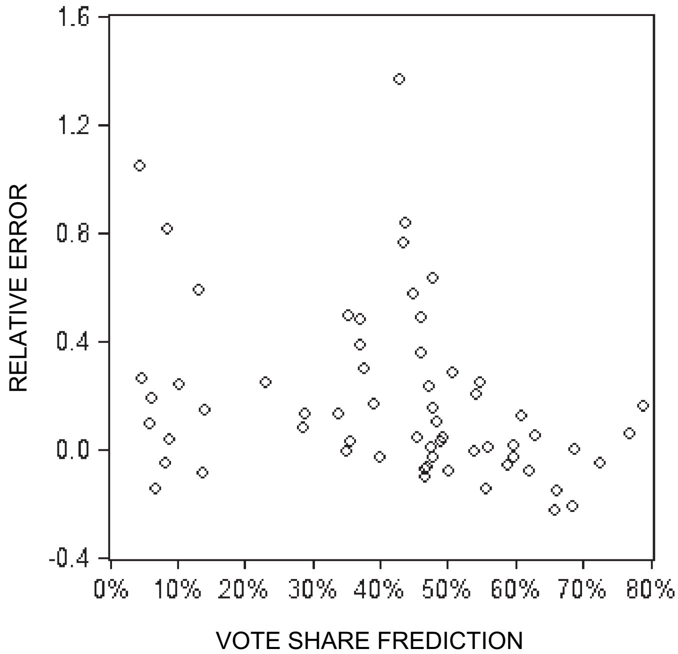

Figure 2: Predicted Vote Shares and Their Errors Relative to Size

Both Figures 1 and 2 indicate that for the higher predicted vote share sizes, the overprediction tends to diminish or disappear altogether. Indeed, Table 2 shows that when the predicted vote share sizes are split between those under and over 50%, the mean errors of the predictions above 50% are actually negative.[9] The histogram displays this visually (Figure 3).

[9] In this table, we have only included the contracts where the automated market maker goes from 0 to 100%, and has a midpoint of 50%. For the German contracts and one Norwegian contract, there were other min-max ranges.

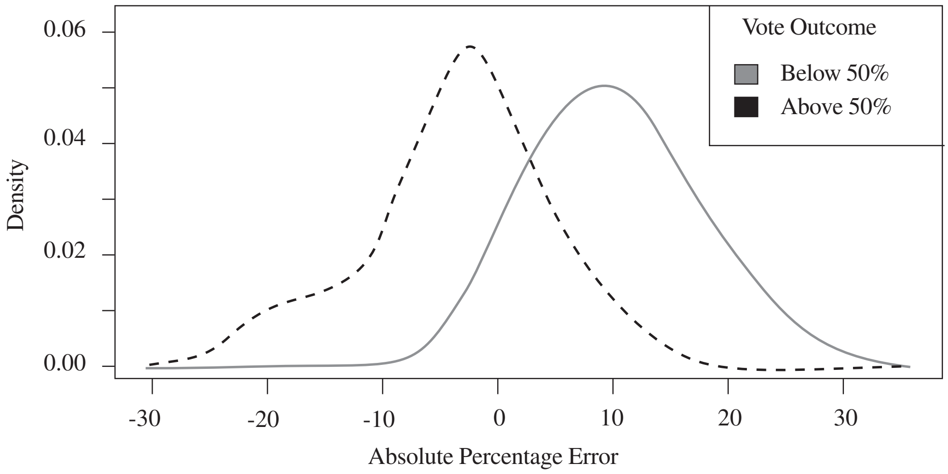

Figure 3: Density Plot of Prediction Bias for Vote Autcomes Under and Over 50%

Table 2: Prediction Errors in Percentage Points

	ERROR	ERROR>50%	ERROR<50%
Mean	3.30	-1.14	5.66
Median	1.95	-0.38	2.90
Maximum	24.80	11.20	24.80
Minimum	-19.48	-19.48	-5.50
Std. Dev.	8.68	8.96	7.67
Skewness	-0.06	-0.51	0.82
Kurtosis	3.70	2.63	2.72
Sum	178.14	-22.78	192.52
Sum Sq. Dev.	3993.79	1526.27	1942.35
Observations	54	20	34

3. SUGGESTED EXPLANATIONS FOR OBSERVED BIAS

What might explain the observations that small-sized vote share contracts are overpredicted while large-sized contracts are underpredicted? We shall discuss three possible explanations; one political, one psychological, and one technical.[10]

[10] According to the literature, we might expect a few alternative errors in prediction markets, which are however not relevant for us since they do not apply to vote share markets. In relation to probability markets it has for instance been discussed whether traders systematically overestimate low probabilities and underestimate high probabilities, also called "favorite-longshot bias", as prospect theory suggests (Kahneman and Tversky, 1973).

The political explanation is the horse-race logic of elections and referendums. Horse-race media coverage informs voters whether a political actor is doing well or not in the polls or according to experts. According to this media logic, like horse-races elections and referendumss are only interesting if they produce a winner and are competitive. Hence, especially in winner-take-all elections or referendums, there is a general tendency in the media to overestimate the narrowness of the race (Ansolabehere and Iyengar, 1994; Broh, 1980). This media logic might easily spill over into the expectations of the participants in prediction markets, since their information might be based to an important degree on media reports. This horse-race logic can generate a bias towards the midpoint if this midpoint is simultaneously the winning threshold. The horse-race logic does not apply to all votes, but only in winner-take- all contests. Winner-take-all votes consist of referendums and elections in two-party systems. If the horse-race logic hypothesis drives the bias in market predictions, then all predictions for referendums should move towards 50, while there should be no bias for party vote shares in the elections of Germany, Norway and Switzerland.

The second hypothesis is that a judgment failure, known as *the winner's curse in combination with cash constrained traders*, causes systematic bias in the predictions (Jacobsen et al., 2000; Potters and Wit, 1996).

The winner's curse occurs as bidders participating in common value auctions tend to put too much weight on their own private information and judgment. Put shortly, in an auction, the winner of a bid has the highest revealed estimation of the auctioned item and thus has bought an overpriced item when measured against the likely average estimation of that item. In itself, this phenomenon will in a prediction market lead to vote share contracts' being overpriced. When combined with cash constraints on the traders – as is the case in most prediction markets – Jacobsen et al. (2000) present an analytical model arguing that small-sized vote share contracts will be overpriced and large-sized ones will be underpriced. This is because cash constrained traders, all else being equal, will prefer a cheaper contract over a more expensive one.

Their analytic model of a two party system assumes that traders are motivated by differences in information and shows that the market's equilibrium price gives a correct prediction only if the true vote shares of the two parties are equal. The authors do not have an analytical result for a model with more than two parties, but simulations indicate that the model then consistently predicts a negative correlation between the true vote shares and the price biases. This expectation was supported by empirical evidence from their four prediction markets of Dutch political elections.

Jacobsen et al.'s model expects the *relative price* of the contracts to determine whether they will be under- or overpredicted. That is, in the same market, the cheap contracts will be overpriced, and the expensive contracts will be underpriced. If the Jacobsen hypothesis were valid, we should witness this pattern in our data. An important note in this regard is that the prediction market software allows for short selling, that is, a trader can sell a contract he does not possess. To make sure that the trader can buy back his shares, an amount is frozen in his account. This amount equals the number of short sold shares multiplied by the maximum possible price of each share. When the maximum value is 100, this means that short selling a contract, for example at the price of 30, costs the same as buying a contract at the price of 70: the seller earns 30 monetary units for selling the contract, but must freeze 100 monetary units until he has bought the contract back. Hence, when the min-max range of the possible contract value is 0-100, it is the distance from the midpoint – which is 50 in most cases – that determines how cheap it is to buy or short sell for the trader. When the price is above the midpoint, it is cheaper to short sell than to buy, and vice versa when it is below 50. A contract is cheaper the farther away its price is from the midpoint.

The third and last explanation for overprediction of small-sized contracts and underprediction of large-size contracts we shall consider is the influence of the automated market maker. The LMSR uses the cost function

$$C(q) = b\log\left(\sum\nolimits_i \exp\left(\frac{q}{b}\right)\right), \tag{1}$$

where the state of the market is kept by the quantity vector **q**, *i*th element of which determines the payout owed to traders if the *i*th element occurs, and *b*>0 is the liquidity parameter that the market organizer sets (Othman et al., 2013).

Due to its logarithmic function, the LMSR algorithm works in such a way that it becomes increasingly expensive to push the price further down from the midpoint and towards the minimum level, and equally increasingly expensive to push the price further up from the midpoint and towards the maximum level. In a context of cash constraint, this could lead to overprediction of small-sized vote share contracts and underprediction of large-sized contracts. Figure 4 illustrates the price function below in a market where the minimum value is 0 and the maximum value is 100.

These three hypotheses are not mutually exclusive. All explanations may, in principle, simultaneously introduce biases. With the observational data at hand, we cannot know for sure whether the predictions conceal some effects of one explanation for example because of the dominance of another. There are hence limits to how strongly we can make inferences with regards to which hypothesis seems to be the

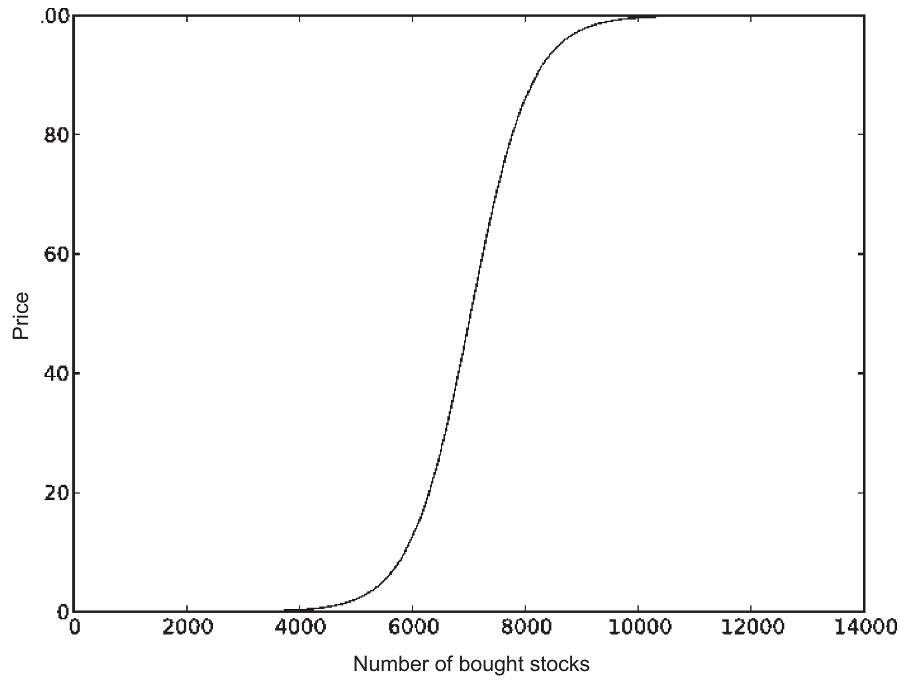

Figure 4: Example Graph of Logarithmic Market Scoring Rule Price Function

better explanation for the observed prediction biases. That said, this analysis encompasses the largest collection of prediction market contracts from real European elections and referendums, collected from three different countries and over the period of five years. Our strategy here is to make a simple overview analysis of what those data indicate, while leaving for future research the task to pursue other methodological approaches such as experiments to single out potential causes of bias and measure their effects.

So, what do the data reveal? Evaluating the horse-race hypothesis first, we expect to see market predictions being closer to the 50% mark than the actual outcomes, but only for the contracts where this threshold concurs with the winning threshold. For the referendums the predictions were on average too high for contracts below the midpoint of 50%, and too low for the contracts predicted to be higher than 50%. The 20 contracts above 50% underestimated the yes-vote share on average by 3.4 percentage points. The 23 contracts on referendums below 50% overestimated the yes-vote share on average by 10.1 percentage points. The result applies also in relative terms: high prices underestimated the outcome by around 5% of their own value; low prices by as much as 36%.

In contrast to the referendums, according to the horse-race logic we wouldn't expect a bias in the case of the party vote shares. However, the party vote shares were also systematically biased towards the midpoint and not significantly less so than the referendum vote shares. While the relative bias towards the means among the referendums (21%) is somewhat higher than among the elections (19%) the t-test reveals that the difference is not significant (Pr = 0.37). Hence, although parts of the bias concur with the expectations derived from the horse-race hypothesis, the bias does not disappear in the markets where we expect them to. In sum, there is weak support for this political explanation.

Next, following the hypothesis based on the winner's curse in combination with cash constraint, we expect cheap contracts to be overpredicted and expensive contracts to be underpredicted. To check whether the data supports the model, we categorize the contracts within each market as cheap or expensive relative to the price of the other contracts in the market.

When the contracts are classified in this manner, we find that both the expensive and the cheap contracts are overpredicted on average. Contrary to the winner's curse model expectations, the cheap contracts are somewhat less overpredicted, with an average of 2.7 percentage points, while the expensive contracts have an average prediction error of 3.3 percentage points.

Finally, we analyse the data with respect to the hypothesis that it is the automated market maker called the logarithmic market scoring rule (LMSR) that introduces biases into the predictions. As previously argued, if the LMSR is the cause of a bias in the predictions, then we should observe that the predictions in general are closer to the midpoint of the min-max scale than the outcomes are. For 54 out of the 62 predicted election and referendum outcomes, the min-max scale ranged from 0 to 100. As we have seen from Table 2, the predictions were on average too high for contracts below the midpoint of 50% and too low for the contracts predicted to be higher than 50%. For the remaining eight contracts, seven had a min-max range of 0-60 (the 2013 German election), and one had a min-max range of 0-10 (Liberal party in the 2009 Norwegian election). All but one of these contracts revealed the same pattern as observed in the other 51 contracts: of those contracts that were lower than the midpoint of 30 in the 2013 German election, five out of six were overpredicted. The only contract that was above 30 – the CDU/CSU – was underpredicted. The Liberal party in Norway was predicted to receive a vote share of 4.9%, an overprediction of one percentage point, also in line with the pattern from the other observations.

It seems, then, that among the three hypothesized explanations for the observed prediction bias, the data supports the LMSR hypothesis. Unlike what we would expect from the horse-race hypothesis, the bias is not confined to referendums only, and as opposed to the winner's curse with cash constraint hypothesis, the cheaper contracts in a market are less overpredicted than the larger ones. The pattern of bias does however fit well with the LMSR hypothesis: almost all contracts are overpredicted when they are below the midpoint of the scale and underpredicted when they are above it. The mechanism works so that as the prices approach zero, the LMSR algorithm makes it increasingly expensive for traders to push down the prices of the vote share contracts. The same goes for buying contracts as their prices approach the maximum scale value (normally 100). Like a rubber band, the predictions are stretched and "want" to return to their normal condition, which is at the midpoint. Unless traders spend a disproportionately large sum of money to withstand this force – which they are unlikely to do due to cash contraints – the predictions will reflect this "rubber band" bias that the logarithmic market maker has introduced. In multiparty elections, almost all contracts are below 50%, and the majority are below 20%. On average, these contracts will be overpredicted. Therefore, unlike probability markets, the logarithmic-based algorithm seems not to represent vote share contracts very well.

4. CONCLUSION

We have in this paper presented 62 prediction market vote share contracts for elections and referendums in Switzerland, Germany, and Norway. We are aware of no larger sample of market predictions and their outcomes in real European elections and referendums. We observe an aggregate bias in the predictions, whereby the actual outcomes tend to have more extreme values than what was predicted. Small-sized vote share contracts tend to be overpredicted and large-sized vote share contracts tend to be underpredicted. We discuss three explanations that may account for this pattern. The results portrayed in this paper indicate that using the logarithmic market scoring rule (LMSR) as automated market maker in vote share markets may be problematic. However, these real world observations may conceal compounded effects, and we cannot infer with certainty from our study that the horse-race logic or the winner's curse is irrelevant as an explanatory factor. We recommend that researchers conduct experimental studies where the effects of the potential causes can be measured in isolation. A fruitful experiment would be to compare predictions from prediction markets that use the LMSR with markets utilizing the continuous double auction (i.e., "manual" market makers), or other kinds of automated market makers such as those based on a linear function. There is no denying that the LMSR has served the prediction market community very well, especially in overcoming the problem of illiquid markets. If what our data indicate is true, that is, that this market maker causes biased vote share predictions, then we should aim to refine the algorithm to overcome bias rather than entirely discard it.

REFERENCES

Ansolabehere, S. and Iyengar, S. (1994). Of horseshoes and horse races: Experimental studies of the impact of poll results on electoral behavior. *Political Communication,* 11: 413-430.

Arnesen, S. (2012). *Leaping into the Unknown: Comparing, Testing, and Applying Methods of Predicting Elections.* University of Bergen, Bergen, Norway.

Arnesen, S. and Bergfjord, O. J. (2014). Prediction markets vs polls – an examination of accuracy for the 2008 and 2012 US presidential elections. In *The Journal of Prediction Markets,* 8: 24-33.

Berg, H. and Proebsting, T. A. (2009). Hanson's automated market maker. In *The Journal of Prediction Markets,* 3: 45-59.

Berg, J., Forsythe, R., Nelson, F., and Rietz, T. (2008a). Results from a dozen years of election futures markets research. In *Handbook of Experimental Economic Results* (Vol. 1). Amsterdam: Elsevier.

Berg, J., Nelson, F. and Rietz, T. (2008b). Prediction market accuracy in the long run. In *International Journal of Forecasting,* 24: 285-300.

Broh, C.A. (1980). Horse-race journalism: Reporting the polls in the 1976 presidential election. In *Public Opinion Quarterly:* 44, 514-529.

Erikson, R.S. and Wlezien, C. (2008). Are political markets really superior to polls as election predictors? In *Public Opinion Quarterly:* 72: 1-24.

Erikson, R.S., & Wlezien, C. (2012). Markets vs. polls as election predictors: An historical assessment. In *Electoral Studies,* 31(3): 532-9.

Hanson, R. (2003). Combinatorial information markets. In *Information Systems Frontiers,* 5: 107-119.

Hanson, R. (2007). Logarithmic market scoring rules for modular combinatorial information aggregation. In *Journal of Prediction Markets,* 1: 1-15.

Jacobsen, B., Potters, J., Schram, A., van Winden, F. and Wit, J. (2000). (In) accuracy of a European political stock market: The influence of common value structures. In *European Economic Review,* 44: 205-230.

Kahneman, D., and Tversky, A. (1973). On the psychology of prediction. In *Psychological Review,* 80: 237-251.

Othman, A., Pennock, D. M., Reeves, D. M., and Sandholm, T. (2013). A practical liquidity-sensitive automated market maker. In *ACM Transactions on Economics and Computation, 1*(14).

Othman, A., and Sandholm, T. (2013). The Gates Hillman prediction market. In *Review of Economic Design,* 17, 95-128.

Pennock, D. (2004). A dynamic pari-mutuel market for hedging, wagering, and information aggregation. In *Proceedings of 5th Annual Conference on Electronic Commerce,* 170-179. ACM Press, New York.

Pennock, D. (2006). *Implementing Hanson's Market Maker.* http://blog.oddhead.com/2006/10/30/implementing-hansons-market-maker/. Last access: June 11, 2015.

Pennock, D., and Sami, R. (2007). Computational Aspects of Prediction Markets. In N. Nisan, editor, *Algorithmic Game Theory*. Cambridge: Cambridge University Press.

Potters, J., and Wit, J. (1996). Bets and Bids: Favorite-Longshot Bias and Winner's Curse. Tilburg University, Center for Economic Research, Tilburg.

Spann, M., and Skiera, B. (2003). Internet-based virtual stock markets for business forecasting. In *Management Science,* 49: 1310-1326.

Strijbis, O., and Schnapp, K.-U. (2015). *Aktivierung und Überzeugung im Bundestagswahlkampf 2013.*Springer, Wiesbaden.