Political Markets

Andreas Graefe, LMU Munich

a.graefe@lmu.de

January 1, 2016

Abstract. This chapter summarizes the latest research on prediction markets for political forecasting. After describing the concept of judgmental forecasting, I outline the history of political prediction markets from their predecessors in 16th century Italy to modern day online markets. Then, I describe important aspects of prediction market design, followed by a comparison of features of prediction markets and simple surveys. I also provide evidence on the relative accuracy of prediction markets and alternative methods for forecasting 44 elections in eight countries. Finally, I discuss the findings as well as their implications for future research.

Keywords. prediction markets, betting markets, accuracy, polls, expert judgment, expectation surveys, quantitative models, combining forecasts, wisdom of crowds

Introduction

As long as there have been elections, people have tried to predict their results and, over the past three decades, researchers have made considerable progress in developing and improving methods for how to forecast elections. For example, the accuracy of predictions based on standard vote intention polls has been improved by using statistical techniques to project their results to Election Day (e.g., Erikson and Wlezien, 2008; Campbell, 1996). Researchers have also developed econometric models that provide long-term forecasts by using information that is available months before the election, such as the state of the economy, the popularity of the president, or the time the incumbent party has held the White House. Essentially, such models generate forecasts by comparing the situation in a particular election campaign with what has happened in historical elections. This approach allows for identifying structural factors that influence elections and thus help us to explain and understand election outcomes (Bélanger, this volume).

An alternative—and probably the most straightforward—way for making forecasts is to ask people to think about a situation and to predict what will happen. For example, when it comes to forecasting election results, one may simply ask people how they expect the election to turn out, and use their aggregated judgment as forecasts. In fact, researchers have long shown that the majority of regular citizens are able to correctly predict who will win an election (Lewis-Beck and Skalaban, 1989). Of course, methods that simply reveal and aggregate people's expectations cannot help to explain election results. However, as shown in this chapter, such methods are valuable in quickly producing forecasts that are not only easy to understand, but also quite accurate.

This chapter summarizes the latest research on one popular method for aggregating people's judgment in a forecast, namely so-called prediction markets. The chapter is structured as follows. After describing the concept of judgmental forecasting, I outline the history of political prediction markets from their predecessors in 16th century Italy to modern day online markets. Then, I describe important aspects of prediction market design, followed by a comparison of features of prediction markets and simple surveys. Finally, I provide evidence on the relative accuracy of prediction markets and alternative methods for forecasting elections and discuss the findings as well as implications for future research.

Judgmental forecasting

In contrast to quantitative forecasting methods, which use statistical methods to derive forecasts from data, judgmental forecasting methods primarily rely on people's judgment or qualitative information. Judgmental forecasting is thus valuable if available data are inadequate for quantitative analysis, or if the use of qualitative information can increase the accuracy or acceptability of forecasts (Graefe et al., 2013).

If judgmental forecasts are derived in an unstructured way, the forecasting literature refers to this approach as unaided judgment. Unaided judgment can be useful if people have good knowledge about the situation. When it comes to the outcomes of elections, there is reason to believe that people possess useful knowledge by following the media campaign coverage, reading recent polling results, and talking to their peers about the election. Indeed, in surveys that ask who will an election, 60% to 70% of respondents in the UK and US were able to correctly pick the winner. When the individual responses were combined, these so-called vote expectation surveys predicted the correct winner with a hit rate between 77% and 86% (Graefe, 2014).

In the forecasting literature, combining forecasts is a well-established means to improve accuracy (Armstrong et al., 2015). The combined forecast is usually more accurate than the typical individual forecast and often even more accurate than the most accurate individual forecast (Graefe et al., 2014b). For combinations of judgmental forecasts, this phenomenon has become widely known as "the wisdom of crowds" (Murr, this volume). In vote expectation surveys, the gains from combining forecasts arise from pooling the responses, for example, by simply assigning equal weights to each respondent's prediction.

An alternative means to harness the wisdom of the crowd is to utilize the price system of the market. People commonly use markets to exchange goods and to make profits. For example, if a person knows that the price for a certain good will go down, she can make profit by selling today and buying later at a lower price. Assume she finds a pattern that the price for a good generally decreases on Wednesdays. Then, she would sell on Tuesdays and buy on Thursdays. Through the process of trading, she then reveals her private information to the market and, as a result, the price will not go down on Wednesdays as much any more. The more people compete in looking for information or patterns that predict price movements, the more difficult it gets to find information that is not yet incorporated in the market price. Thus, the market aggregates qualitative information and translates it into a numerical estimate (i.e., the market price).

So-called prediction (or betting) markets utilize this powerful mechanism for aggregating dispersed information not for speculation but for the purpose of making predictions. That is, prediction markets combine human judgment into a single forecast and can thus be classified as a judgmental forecasting method (Graefe et al., 2013). When applied to forecasting political events such as the outcomes of elections, the method is also known as political markets.

History of political markets

Prediction markets are often thought of as a relatively new forecasting method. Yet, such markets have been around for at least half a millennium, as Rhode and Strumpf (2004; 2014) described in their reviews of historical betting markets.

The earliest account of political markets dates back to 1503, when betting on who will be the next pope was already considered old practice. Although Pope Gregory XIV banned betting on papal elections in 1591, such wagers have occasionally reappeared and also gained some press attention, such as the conclaves over the successor to Leo XIII in 1903 as well as Benedict XV in 1922. In Italian city-states such as Venice and Genoa, betting on the outcome of civic elections was common in the 16th and 17th century. In Great Britain, including its former colonies such as Australia, Canada, New Zealand, Singapore, South Africa, and the United States, political betting can be traced as far back as the 18th century. While the main focus of these markets was on election outcomes, betting also occurred on other political events such as the outcome of no-confidence votes, the resignation of leaders, or the outcome of foreign and military missions (Rhode and Strumpf, 2014).

In the US, political betting was common in the 19th, where public bets on a candidate were considered a sign of support. However, the heyday of election betting was between 1884 and 1940, when large-scale markets on presidential elections were operated (Rhode and Strumpf, 2014). These markets not only provided accurate forecasts of the election outcome in an era before scientific polling (Erikson and Wlezien, 2012), they were also widely popular among investors and journalists alike. At certain times, the trading volume in these markets exceeded that in the stock exchanges on Wall Street and major news outlets such as the *New York Times*, *Sun*, and *World* reported the betting odds as forecasts of the election outcome on a nearly daily

basis. However, political markets began to disappear after 1940 due to a number of reasons. First, moral concerns associated with betting on political outcomes arose, second, opinion polls became popular as an alternative method for predicting election results, and, third, other gambling opportunities such as horse race betting became available (Rhode and Strumpf, 2004).

Modern day online prediction markets first appeared in 1988, when the University of Iowa launched the Iowa Electronic Markets (IEM) to predict the US presidential election of the same year (Forsythe et al., 1992). The accuracy of the market's forecasts ignited the interest of researchers in different fields. Since the mid-1990s, various studies have tested the predictive validity of prediction markets in areas other than political forecasting, for example, for forecasting sports events or business figures. A second boost in prediction markets' popularity can be traced back to two events in the early-2000s. The first refers to the cancellation of the Policy Analysis Market in 2003, a project funded by the Unites States' Defense Advanced Research Project Agency (DARPA), whose goal was to improve existing intelligence institutions by predicting military and political instability around the world. Ironically, it was the decision to cancel the project due to ethical concerns with betting on political events (e.g., terrorist attacks), which attracted broad media coverage and thus made a broad public aware of prediction markets (Hanson, 2007). The second event was the publication of James Surowiecki's (2004) bestselling book *The Wisdom of Crowds*, in which he describes prediction markets as an efficient means to harness collective intelligence. Shortly after, prediction markets were listed on the *Gartner Hype Cycle* and soon major business consultancies saw them as an emerging trend. However, empirical studies found little evidence that prediction markets outperform alternative methods for business forecasting (Graefe, 2011), which is one of the reasons why prediction markets have not made it to regular adoption in companies (Wolfram, 2015). In comparison, prediction markets have

become an established method in modern day election forecasting, most likely due to their ability to provide more accurate predictions than vote intention polls.

Market design

This section describes two types of contracts that are most common for forecasting political events (winner-take-all and index contracts) and discusses critical questions of market design, such as how to define the prediction event and how to motivate participation.

Contract types

Prediction markets enable participants to buy and sell contracts whose payoff depends on the outcome of a future event. Once the outcome is known, participants are paid off in exchange for the contracts they hold. The choice of a specific market type depends on what is being forecast, for example, whether the goal is to predict one-off events or continuous outcomes.

Winner-take-all contracts allow for obtaining a probability estimate of whether or not an outcome will occur. Such contracts are the most popular type of prediction market and are, for example, used to predict election winners. Imagine a contract that pays off $1 if candidate A wins the election and $0 otherwise. Then, a contract price of $0.55 for 'candidate A' means that the market predicts that the candidate has a 55% chance to win the election. If a participant believes that candidate A's chance to win is higher, say 70%, she would buy contracts for any price less than $0.70. Assume she bought 10 shares at a price of $0.55 per share and candidate A eventually won the election. Then, she would have made a profit of $4.5 (=10*[$1-$0.55]).

Index contracts allow for predicting numerical outcomes, such as vote-shares in an election. The payoff of index contracts is not known a priori but depends on the final value of the target criterion. Imagine a contract pays off $0.01 times the final vote share received by candidate A

and the contract currently trades at $0.51 per share. This means that the market predicts candidate A to receive 51% of the vote. If a participant believes that candidate A's final vote share will be higher, he would buy that contract. Assume he bought 100 shares at a price of $0.51 per share. Furthermore, assume candidate A eventually received 48% of the vote and the contract thus paid off $0.48 per share. In this case, the market participant would have lost $3 (=100*[$0.48-$0.51]).

Both types of contracts are implemented in the IEM's U.S. presidential election markets: a winner-take-all market to predict the winner of the popular vote and a vote-share market to predict the popular two-party vote shares of the Republican and Democratic candidates.

Contract definition

Regardless the implemented contract types, it is important to clearly define the rules for settling a contract. For instance, when it comes to US presidential elections, it has to be clear whether the winner will be judged based on the popular (as done by the IEM) or the electoral (As done by Betfair) vote.

The following example shows how poorly designed contracts may alienate market participants. In 2006, Tradesports.com offered a contract that aimed to predict whether North Korea would launch a long-range missile that leaves its airspace by July 31st. When, on July 4th, news that North Korean missiles dropped in the Sea of Japan made worldwide headlines, market participants who had bet that this would happen thought they had won. There was a problem, however. The contract had specified the U.S. Department of Defense as the official source of information for judging the event's outcome, which never published a formal statement on the missile tests. As a result, Tradesports rewarded traders who bet *against* the missile launch, a decision that frustrated those who made correct predictions (Graefe, 2008).

Another important issue when setting up a contract is whether or not participants actually possess valid information. Like any other judgmental forecasting method, prediction markets will fail if no relevant public information exists that could be aggregated. Sunstein (2006: 134-137) summarized several cases in which prediction markets failed to provide accurate forecasts, such whether people would resign from (and be nominated to) the US Supreme Court or whether weapons of mass destruction would be found in Iraq. Vaughan Williams and Paton (2015) provided additional evidence in their analysis of betting odds from papal conclaves from 1503 to 2013. In seven of the ten conclaves for which odds were available, the markets failed to predict the next pope.

Incentive scheme

Prediction markets provide performance-based incentives and thus are expected to motivate participation and truthful information revelation. The most straightforward way is to allow participants to invest real money. However, the law in many countries prohibits such real-money markets. An alterative is to use play money and to award prizes to the best performing participants based on rankings of, for example, their portfolio value. Research on the relative performance of play-money and real-money markets is limited and inconclusive. While one study found no differences in accuracy for sports events (Servan-Schreiber et al., 2004), two studies found real-money markets to be more accurate than play-money markets (Rosenbloom and Notz, 2006; Deimer and Poblete, 2011). One concern with play-money markets is that participants cannot lose money, which may open doors to market manipulation, which I discuss in the following section.

Markets vs. expectation surveys

Prediction markets use the price system of the market to aggregate people's expectations regarding the outcome of a future event. Thus, the method is closely related to expectation surveys, which simply ask people to make a prediction of what will happen. This section discusses important similarities and differences between both methods, which are summarized in Table 1.

Table 1: Comparison of prediction markets and expectation surveys

	Prediction Market	Expectation Survey
Information aggregation		
Timing	Continuous	One-shot
Aggregation mechanism	Price system	Statistics (Mean)
Incentives		
Participation	Yes	No
Information seeking	Yes	No
Truthful information revelation	Yes	No
Participants		
Entry	Open	By invitation
Selection	Self-selected	Random
Sample	Non-representative	Representative
Vulnerable to manipulation	Yes	No

Information aggregation

Surveys are one-off activities that need to be triggered and that provide results at a certain point in time. Thus, they cannot aggregate information in real-time and are therefore less useful in situations where new information may become available frequently or suddenly.

In comparison, prediction markets use the price mechanism of the market to aggregate information. Thereby, the market price reflects the combined expectations of all participants at any time. Since markets are continuously open for participation, they have the potential to instantly and automatically incorporate new information into the forecast.

This feature is often proclaimed as one of the method's major advantages. Snowberg et al. (2013) describe the reaction of a contract traded at Intrade to the killing of Osama bin Laden as an illustration of how fast prediction markets can incorporate new information. This contract tracked the probability that bin Laden would be captured or killed by December 31st, 2011. When Keith Urbahn, former chief of staff to Defense Secretary Donald Rumsfeld, posted on his Twitter account that he heard bin Laden was killed, the market forecast rose within 25 minutes from 7% to nearly 99%; eight minutes later, the first media outlets announced the news. However, in their analysis of betting on the 2013 papal conclave, Vaughan Williams and Paton (2015) concluded that the Betfair prediction market failed to sufficiently and quickly incorporate information that became available over the course of the conclave.

Incentives

Traditional surveys provide no incentives for participation. While some people might feel an intrinsic motivation or moral obligation to participate in surveys, many people may regard participation as a loss of time. Furthermore, surveys do not motivate participants to actively search for information or to reveal their true beliefs. As a result, respondents' expectations may be influenced by their preferences. For example, when being asked to predict who will win an election, people may be inclined to name their preferred candidate. This bias, which is known as wishful thinking, is long known to be common and occurs in all types of elections, from local

referenda to national elections, and across various countries; see Miller et al. (2012) for an overview of recent research on wishful thinking.

In comparison, prediction markets offer several theoretical advantages over traditional surveys, as they provide incentives for participation, information seeking, and truthful information revelation. Since market participants can win or lose money based on the accuracy of their individual predictions, they should only become active if they think that the current prediction is incorrect. For the same reasons, market participants have an incentive to actively look for information, which may help them to improve the market forecast, and to reveal their true beliefs. For example, regardless of which candidate they support, participants should truly reveal whom they expect to win.

However, empirical evidence suggests that the incentives provided by prediction markets may not always be sufficient to overcome wishful thinking. Forsythe et al. (1999) analyzed trading behavior in the 1988 U.S. presidential election vote-share market and the 1993 Canadian House of Commons Market, both operated by the IEM. They found that participants in both markets exhibited wishful thinking. In particular, participants bought more shares of their preferred candidates and sold more shares of candidates that they did not support. Rothschild and Sethi (2015) provided additional evidence in their analysis of individual trading data in Intrade's 2012 US presidential election winner market. The authors found that 94% of traders, who accounted for 69% of the total trading volume, (almost) exclusively held shares of either Obama or Romney. While the data did not provide information about which candidate a trader supported, the results provide a strong indication for the existence of wishful thinking.

Participants

Participants in prediction markets are often referred to as "self-selected experts." The underlying assumption is that the incentive mechanism should make sure that only people who think that they know better than the market as a whole become active. While it is technically possible to restrict participation to a certain group of people, most markets are open for anyone to participate. Not surprisingly then, participants do not form a representative sample of the population, as not all people are equally likely to participate in prediction markets. For example, a study on the 1988 IEM found that participants were predominantly white, male, well educated, and belonged to middle and upper income categories. In addition, participants tended to be more Republican and less independent in their partisan leanings, and were more politically active than the general public (Forsythe et al., 1992).

The likely reason for the bias towards higher educated participants is that many people lack the understanding of how markets work and, in particular, how to translate their expectations into a market price. One laboratory experiment asked participants how satisfied they were with participating in one of four judgmental forecasting methods: prediction markets, the Delphi method, nominal groups and traditional face-to-face meetings. Prediction markets were rated least favorable. In particular, market participants were least satisfied with the process and rated the method highest in terms of difficulty of participation (Graefe and Armstrong, 2011). While this finding suggests that prediction markets are unsuitable for involving a representative sample of participants, more research on how people perceive participation in prediction markets would be valuable.

In comparison, expectation surveys aim to obtain responses from a random and representative sample of the population. However, the fact that response rates in traditional phone surveys have

decreased below ten percent in recent years (Kohut et al., 2012) has raised doubts about the representativeness of survey samples (Wang et al., 2015). For example, some people may not answer calls from unknown numbers and may not be reachable by phone at all. Or, people who did answer the phone may be unwilling to participate, for example, if they support unpopular parties or candidates (Noelle-Neumann, 1974). As a result, those who chose to participate may not form a representative sample.

Manipulation

In contrast to surveys, where each respondent achieves equal weight, the weight of an individual participant's opinion in prediction markets depends on his budget and can thus vary widely across traders. For example, Rothschild and Sethi (2015) found that in Intrade's 2012 US presidential market, a single trader accounted for one quarter of the total money on Romney. The possibility for a single participant to invest large amounts of money brings along the potential to influence market prices, which is concerning with regards to market manipulation. That is, certain (groups of) participants may attempt to manipulate the market forecast in order to affect perceptions of the state of the race and thus to influence the election outcome. For example, in times where prediction markets increasingly gain the media's attention, manipulators could try to increase a candidate's predicted chance of winning in order to motivate support, increase turnout, or alter voting preferences. The underlying logic is that voters may change their preferences in favor of the candidate who is most likely to win, which is known as the so-called bandwagon effect (Simon, 1954). That said, research to date is inconclusive as to how forecasts affect turnout and voting preferences.

Early evidence on whether prediction markets are vulnerable to manipulation is mixed. While some studies showed that manipulation has not been successful historically (Rhode and Strumpf,

2004), in the laboratory (Hanson et al., 2006), or in the field (Camerer, 1998), one study reported successful manipulation attempts (Hansen et al., 2004). In their review of studies of price manipulation, Wolfers and Zitzewitz (2004) concluded that, besides a short transition phase, none of the known attacks had a noticeable influence on the prices.

However, recent experience suggests that manipulation is a concern. Rothschild and Sethi (2015) showed that the single trader who invested large amounts of money to back Romney had a lasting effect on the contract's market price. Over the last months before the election, the price of Intrade's Obama contract was consistently between five and ten percentage points lower than the respective contract on a competing exchange. While this presumed manipulation came at a price (the trader lost close to seven million dollars), the loss is negligible compared to the cost of a US presidential campaign. As put by Rothschild and Sethi (2015: 24), "a media narrative could be manipulated at a cost less than that of a primetime television commercial."

Manipulation appears to be an even bigger concern in play-money markets, in which there is no monetary cost to manipulation, as participants cannot lose money. A market that aimed to predict the parties' vote shares in the 2013 German election provides an example. During the four months leading up to the election, this market's forecast for the Alternative for Germany (AfD), a right-wing euroskeptic party that was just founded a few months earlier, was unrealistically high. At times, the market predicted that the AfD would gain more than 25% of the vote, which was more than ten times the party's polling numbers. On Election Eve, the market predicted the AfD to receive nearly 14%. In comparison, the party polled on average a bit above 3% and the average of five other prediction markets forecast a vote share of 4.7%, which was exactly the party's final vote share achieved in the election. Interestingly, of the six available prediction markets, this market that predicted extraordinary high vote shares for the AfD was the only one whose

forecasts were reported by a prominent news outlet. The Handelsblatt, a leading German business newspaper, sponsored the market and regularly reported its forecasts on their website. It thus appears likely that AfD supporters hijacked the market to create media attention and to influence public perception of the party's strength.

Market accuracy

This section reviews the relative accuracy of prediction markets for forecasting numerical election outcomes, such as the vote shares or number of seats gained by parties in an election. As shown in Appendix I, data were collected for 44 elections in eight countries: Austria (N=10), Canada (2), Germany (20), Netherlands (1), Norway (1), Sweden (1), Switzerland (1), and the US (8). To allow for a fair comparison, I only compared forecasts that were released around the same time. When forecasts were made for different forecast horizons, I averaged across the horizons. For each method, forecast error was measured in terms of the mean absolute error (MAE). The MAE is the absolute deviation of each party's predicted and final vote share, averaged across all parties. I then calculated the relative absolute error (RAE) of the prediction market forecast compared to the forecast error of the respective benchmark method. Values below 1 mean that the prediction market provided more accurate forecasts, values above 1 mean that the benchmark method was more accurate. The benchmark methods were established methods for election forecasting, namely polls, expert judgment, expectation surveys, econometric models, and method combination.

vs. polls

The most common approach to forecast elections is to ask individual respondents for whom they intend to vote if the election were held today and to then use the aggregated response as a forecast for whom the electorate as a whole will vote. Such intention polls do not provide

predictions but rather snapshots of public opinion at a certain time; yet, polling results are commonly interpreted as forecasts of what will happen on Election Day (Hillygus, 2011). Research in forecasting shows that intentions can provide valid forecasts if respondents form a representative sample of the target population, have no reason to lie, and are unlikely to change their intention over time (Graefe et al., 2013). However, these conditions may not always hold when using intention polls for forecasting elections. First, response rates in traditional phone surveys have decreased below ten percent in recent years (Kohut et al., 2012), which undermines the assumption of random and representative samples. Second, respondents may not be willing to reveal their true intentions, or may abstain from participating in the survey in the first place, for example, if they support unpopular parties or candidates (Noelle-Neumann, 1974). Third, prior research suggests that people's vote intention often varies widely over the course of a campaign (Campbell, 1996; but see Gelman et al., 2016).

Individual polls

As a result, single polls often provide poor predictions, in particular if the election is still far away. Forecasts derived from single polls should thus be considered a weak benchmark to judge the accuracy of prediction markets. Table 2 shows the results of accuracy comparisons between prediction markets and individual polls from a total of 43 elections conducted in Austria, Canada, Germany, the Netherlands, Norway, Sweden, Switzerland, and the US.

Prediction markets achieved their best performance for forecasting the US presidential elections. In each of the seven presidential elections from 1988 to 2012, the markets outperformed the polls. On average across the seven elections, the market's RAE compared to the typical poll was 0.52. That is, the error of prediction markets was 48% (=1-0.52) lower than the respective error of the

typical poll. On average, prediction markets also outperformed polls in all other countries, except for Austria.

Table 2: Forecast accuracy of prediction markets vs. alternative methods

Benchmark		Country	No. of elections	Relative absolute error (RAE)	Market accuracy compared to benchmark	
					Higher	Lower
Intentions	**Typical poll**	**Sum / Weighted average**	**43**	**0.89**	**31**	**11**
		Germany	20	0.91	15	4
		Austria	10	1.21	3	7
		US	7	0.52	7	0
		Canada	2	0.74	2	0
		Netherlands	1	0.96	1	0
		Norway	1	0.57	1	0
		Sweden	1	0.90	1	0
		Switzerland	1	0.77	1	0
	Combined polls	**Sum / Weighted average**	**18**	**0.82**	**12**	**6**
		Germany	9	1.03	4	5
		US	7	0.57	6	1
		Canada	1	0.55	1	0
		Sweden	1	0.94	1	0
	Combined and projected polls	**US**	**7**	**0.83**	**4**	**3**
Expectations	**Typical expert**	**Sum / Weighted average**	**8**	**0.83**	**4**	**4**
		US	6	0.74	4	2
		Germany	1	1.09	0	1
		Canada	1	1.07	0	1
	Combined experts	**Sum / Weighted average**	**7**	**0.87**	**4**	**3**
		US	6	0.79	4	2
		Germany	1	1.40	0	1
	Expectation surveys	**Sum / Weighted average**	**8**	**1.08**	**5**	**3**
		US	7	1.06	5	2
		Germany	1	1.25	0	1
Models	**Typical model**	**Sum / Weighted average**	**8**	**0.76**	**6**	**2**
		US	7	0.66	6	1
		Germany	1	1.48	0	1
	Combined models	**Sum / Weighted average**	**8**	**0.89**	**4**	**4**
		US	7	0.72	4	3
		Germany	1	2.04	0	1
Method combination	**PollyVote**	**Sum / Weighted average**	**7**	**1.28**	**2**	**5**
		US	6	1.21	2	4
		Germany	1	1.70	0	1

Across all 43 elections, prediction markets were more accurate than polls in 31 elections, whereas polls provided more accurate forecasts in 11 elections (in one German election, prediction markets and the typical poll performed similarly). On average across the eight countries, weighted by the number of elections, prediction markets reduced the error of a typical poll by 11% (RAE:0.89).

Combined polls

There is often high variance in the results of polls by different survey organizations, even if these polls were conducted at around the same time. Such variance can be caused by sampling problems, non-responses, and faulty processing (Erikson and Wlezien, 1999). One way to deal with this problem and to increase the accuracy of poll-based forecasts is to combine polls that were conducted at around the same time. The reason is that the systematic (and random) errors that are associated with individual polls tend to cancel out in the aggregate (Graefe et al., 2014b). The good news is that combining has impacted how people nowadays consume polls and online polling aggregators such as realclearpolitics.com and pollster.com have become increasingly popular (Blumenthal, 2014).

Table 2 shows evidence from empirical comparisons of combined polls and prediction markets. In 12 of the 18 elections for which data are available, prediction markets outperformed combined polls, whereas combined polls were more accurate in the three remaining six elections, five of which were conducted in Germany. On average, the error of prediction markets was 18% lower than the respective error of combined polls.

Poll projections

Polls conducted by the same survey organization, and by the polling industry as a whole, can fluctuate widely across the course of a campaign. The reason is that people's response behavior, in particular in early polls, is influenced by campaign events such as conventions (Campbell et al., 1992) and debates (Benoit et al., 2003). The effects of such events on the outcome of high-visibility elections such as U.S. presidential elections are limited, however. As the election nears, people are less influenced by the latest campaign events and have formed stable vote intentions based on a combination of information they have learnt during the campaign, such as the state of the economy, and their basic predispositions, such as ideology and party identification (Gelman and King, 1993). Therefore, it is not until shortly before Election Day that polls provide accurate forecasts.

However, researchers found ways for how to harness early polls for forecasting by calculating poll projections, as they are termed hereafter. Poll projections take into account the historical record of polls in order to make a forecast. For example, assume that the incumbent leads the polls by fifteen points in August. In analyzing historical polls conducted around the same time along with the respective election outcomes, one can derive a formula for translating August polling figures into an estimate of the incumbent's final vote share in November. This is commonly done by regressing the incumbent's share of the vote on his polling results during certain time periods before the election. Prior research found that such poll projections are more accurate than treating raw polls as forecasts (e.g., Erikson and Wlezien, 2008; Campbell, 1996).

One can also combine both strategies (i.e., combining polls and calculating poll projections) to generate poll-based forecasts. One study first calculated rolling averages of all polls that were published in a one-week period and then used these results to calculate poll projections (Graefe,

2014). Table 2 compares the forecasts of such combined poll projections to prediction markets. Prediction markets were more accurate in four of the seven elections and reduced the error of combined poll projections on average by 17%.

vs. expectations

Expert judgment

Judgment of political insiders and experienced election observers were used to forecast elections long before the emergence of scientific polling (Kernell, 2000); and they still are. The common assumption is that experts have experience in reading and interpreting polls, assessing their significance during campaigns, and estimating the effects of recent or expected events on the aggregate vote. Given their omnipresence, surprisingly little is known about the relative accuracy of experts' judgment and prediction markets.

Table 2 compares the accuracy of prediction markets and individual and combined expert judgment. The relative performance of markets and individual experts is mixed in that each method provided more accurate forecasts in four of the eight elections for which data are available. However, only in the US did the prediction markets outperform individual experts. On average, the error of prediction markets was 17% lower than the respective error of the typical expert.

The comparison of prediction markets and combined expert judgment is limited to one German and six US elections, since only a single expert forecast was available for the Canadian case. The results are mixed. Prediction markets outperformed combined experts in four of the six US elections and thereby reduced error by 21%. In comparison, the markets performed poorly in the one German election, with an error 40% higher than the corresponding error of combined experts.

On average across both countries, prediction markets reduced the error of combined experts by 13%.

Expectation surveys

Rather than utilizing responses to the traditional vote intention question, direct forecasts of the election outcome can be derived from responses to the vote expectation question, which asks respondents how they expect the election to turn out (Murr, this volume). The expectation question is usually kept simple by framing the election outcome as a selection problem. While the exact phrasing depends on the specifics of the particular electoral system, citizens are commonly asked to predict the candidate (or party) that will lead the government after the election. For example, the question in the American National Election Studies (ANES) asks respondents which candidate they expect to be elected president or who will win the election in their home state. The question in the British General Election Studies asks which party will get the most MPs or, alternatively, which party will win. The question in the German Longitudinal Election Study asks which coalition of parties will form a government.

The use of the expectation question in pre-election surveys goes back before the emergence of intention polling (Hayes, 1936) and scholars have long studied the question why certain people provide more accurate forecasts than others (Lewis-Beck and Skalaban, 1989; Lewis-Beck and Tien, 1999; Dolan and Holbrook, 2001). However, only recently, scholars have begun to study the value of vote expectation surveys as a method to forecast elections in countries such as Germany, Sweden, UK, and US (Lewis-Beck and Stegmaier, 2011; Murr, 2011; Murr, 2015; Rothschild and Wolfers, 2012; Graefe, 2014; Sjöberg, 2009; Graefe, 2015a). For example, one study compared the accuracy of the expectation question to polls, prediction markets, quantitative models, and expert judgment for predicting election winners and vote

shares in the seven US presidential elections from 1988 to 2012. Across the last 100 days

preceding each election, responses to the expectation question correctly predicted the election

winner with a hit rate of 92%, which was more accurate than the corresponding hit rate of polls

(79% correct), prediction markets (79%), expert judgment (66%), and quantitative models (86%).

When predicting vote shares, expectations were again most accurate on average (Graefe, 2014).

Table 2 shows the results of this study for the relative performance of prediction markets and

expectations. With an MAE of 1.7 percentage points, the error of prediction markets was on

average 6% higher than the respective error of expectation surveys, even though markets

outperformed expectation surveys in five of the seven elections. Another study compared the

accuracy of prediction markets and expectation surveys for forecasting the 2013 German election.

In this case, the error of prediction markets was 25% higher than the corresponding error of

expectation surveys (Graefe, 2015a). Based on the weighted average across both countries, the

error of prediction markets was 8% higher than the error of expectation surveys.

vs. quantitative models

A common theory of electoral behavior is that elections are referenda on the incumbent's

performance. That is, voters are expected to reward the government for good performance and

punish the incumbent party otherwise. Since the late 1970s, economists and political scientists

tested this theory by developing quantitative models to predict election results. Most models are

based on multiple regression analysis of two to five predictor variables, which typically capture

economic conditions, the incumbent's popularity, and how long the incumbent (party) controlled

the government (Bélanger, this volume).

The development and testing of these models has become a well-established sub-discipline of

political science and the models' forecasts are regularly published about two months prior to

Election Day in scientific journals. For example, forecasts from established models of US presidential election outcomes were published in special symposiums in *Political Methodologist* 5(2), *American Politics Research* 24(4) and *PS: Political Science and Politics* 34(1), 37(4), 41(4), and 45(4). These models predict the correct election winner most of the time. Across the six elections from 1992 to 2012, 34 of 39 forecasts of seven well-known models correctly predicted the winner. However, the models' performance in predicting vote shares is mixed. Their mean absolute error (MAE) was three percentage points, and ranged from zero to ten points (Graefe, 2014).

As shown in Table 2, prediction markets outperformed both individual and combined models in forecasting the US elections for which data were available, whereas in the 2013 German election, markets provided less accurate forecasts than both benchmarks (Graefe, 2015b). On average, prediction markets reduced the error of the typical model by 24% and the error of the combined model forecast by 11%.

vs. method combination

As shown for combinations of forecasts from polls, expert judgment, and quantitative models, combining individual forecasts is an effective means to increase accuracy. Combining is beneficial because, first, it allows incorporating the different information sets included in the component forecasts. As a result, the combined forecast includes more information. Second, if the systematic and random errors of individual forecasts are uncorrelated, they cancel out in the aggregate. Therefore, combining is particularly valuable if one combines forecasts that use different methods and draw upon different information, as the underlying forecasts are likely uncorrelated (Armstrong, 2001).

The PollyVote project demonstrates the benefits of combining forecasts across different methods for predicting election outcomes. In averaging forecasts within and across four methods (polls, prediction markets, expert judgment, and quantitative models), the PollyVote has created highly accurate forecasts for the six US presidential elections from 1992 to 2012 (Graefe et al., 2014b; Graefe et al., 2014a) as well as for the 2013 German federal election (Graefe, 2015b).

Table 2 shows the relative performance of the PollyVote and prediction markets. The PollyVote outperformed prediction markets in four of the six US elections and in the 2013 German election. On average, the error of prediction markets was 28% higher than the corresponding error of the PollyVote.

Discussion

Prediction markets are an effective means to forecast election. Evidence to date shows that markets often provide more accurate forecasts than established benchmark methods such as polls, quantitative models, and expert judgment. With data from 43 elections, the evidence is particular strong for the markets' relative accuracy compared to single polls. Markets outperform single polls in about three out of four elections, with an average error reduction of 11%. Most of these comparisons are based on the markets' final forecasts and the final pre-election polls from different pollsters. Thus, the comparisons focus on a time period when polls tend to provide fairly accurate forecasts. For longer-term forecasts, the advantage of prediction markets should increase further.

However, as noted above, single polls are generally a weak benchmark to assess the accuracy of a forecasting method. When comparing the relative accuracy of markets and polls, researchers should thus consider more sophisticated poll-based forecasts such as combined and projected polls. The case of the seven US presidential elections—for which comparisons to single polls,

combined polls, and combined poll projections are available over a 90-day forecast horizon—demonstrates how such approaches increase the relative accuracy of polls compared to prediction markets. Prediction markets outperformed single polls in each of the seven elections, with an average error reduction of 48%. Compared to combined polls, prediction markets were more accurate in six of the seven elections, with an average error reduction of 43%. Compared to combined and projected polls, prediction markets were more accurate in only four of the seven elections, and the average error reduction decreased to 17%.

While the evidence for the relative performance of markets and polls is strong, more studies are necessary that compare the markets' accuracy to alternative forecasting methods. Due to their similarity, the logical approach would be to compare the accuracy of markets and expectation surveys. However, only two studies provide evidence, which is mixed and inconclusive. While prediction markets provided more accurate forecasts than expectation surveys in five of the eight elections for which data were available, the markets' average error across these elections was 8% higher than the corresponding error of expectation surveys. These results suggest that it is not necessary to aggregate information through the price system of the market. Simply asking people about their expectations and pooling the responses to generate a forecast might be enough. Future research should address whether these results hold in other elections in other countries as well as for other types of problems.

Prediction markets have the potential to continuously aggregate information once it becomes available, which can make the method particularly valuable in situation where new information is likely to arise frequently and unexpectedly. In addition, the ability to continuously incorporate information allows for using prediction markets to assess the impact of events, although few such studies are available (e.g., Daron and Roberts, 2000; Arnesen, 2011; Vaughan Williams and

Paton, 2015). Further research is necessary to assess the potential of prediction markets for event studies as well as the conditions under which markets are able to effectively aggregate new information. For example, an interesting question would be to study how the trustworthiness of a source or medium (e.g., newspaper, Twitter) influences the market's reaction to news.

One concern with prediction markets is that the forecasts are subject to manipulation. Recent experience with two markets in the US and in Germany suggests that certain (groups of) traders were successful to influence market prices over a longer period. If the media increasingly include market forecasts in their campaign coverage, the potential gains for manipulators—and thus the threat of manipulation attempts—might increase even further. Future research should thus assess the impact of media coverage of prediction markets forecasts on the accuracy of the forecasts. Also, it is crucial to compare a market's prediction to benchmark forecasts. One way to do this is to look at forecasts from other prediction markets. In both the US and the German case, the manipulated market's predictions differed substantially from those of other markets, which made it possible to spot anomalies. Another way is to compare the market predictions to forecasts from other methods. For example, the PollyVote collects and combines forecasts from different methods and thus provides a useful source for assessing whether a certain prediction should be classified as an outlier. If (a) prediction markets are the only available forecasting method and thus cannot be compared to a benchmark and (b) the market forecasts are covered in the media, one should be cautious about potential market manipulation.

References

Armstrong JS. (2001) Combining forecasts. In: Armstrong JS (ed) *Principles of Forecasting: A Handbook for Researchers and Practitioners.* New York: Springer, 417-439.
Armstrong JS, Green KC and Graefe A. (2015) Golden Rule of Forecasting: Be conservative. *Journal of Business Research* 68: 1717-1731.

Arnesen S. (2011) How prediction markets help us understand events' impact on the vote in US Presidential Elections. *Journal of Prediction Markets* 5: 42-63.

Bélanger E. (this volume) Econometric approaches to forecasting. In: Arzheimer K, Evans J and Lewis-Beck MS (eds) *The SAGE Handbook of Electoral Behaviour.*

Benoit WL, Hansen GJ and Verser RM. (2003) A meta-analysis of the effects of viewing US presidential debates. *Communication Monographs* 70: 335-350.

Blumenthal M. (2014) Polls, forecasts, and aggregators. *PS: Political Science & Politics* 47: 297-300.

Camerer Colin F. (1998) Can Asset Markets Be Manipulated? A Field Experiment with Racetrack Betting. *Journal of Political Economy* 106: 457-482.

Campbell JE. (1996) Polls and Votes The Trial-Heat Presidential Election Forecasting Model, Certainty, and Political Campaigns. *American Politics Research* 24: 408-433.

Campbell JE, Cherry LL and Wink KA. (1992) The convention bump. *American Politics Research* 20: 287-307.

Daron RS and Roberts BE. (2000) Campaign Events, the Media and the Prospects of Victory: The 1992 and 1996 US Presidential Elections. *British Journal of Political Science* 30: 259-289.

Deimer S and Poblete J. (2011) Real-Money Vs. Play-Money Forecasting Accuracy In Online Prediction Markets – Empirical Insights From Ipredict. *Journal of Prediction Markets* 4: 21-58.

Dolan KA and Holbrook TM. (2001) Knowing versus caring: The role of affect and cognition in political perceptions. *Political Psychology* 22: 27-44.

Erikson RS and Wlezien C. (1999) Presidential polls as a time series: the case of 1996. *Public opinion quarterly* 63: 163-177.

Erikson RS and Wlezien C. (2008) Are political markets really superior to polls as election predictors? *Public opinion quarterly* 72: 190-215.

Erikson RS and Wlezien C. (2012) Markets vs. polls as election predictors: An historical assessment. *Electoral Studies* 31: 532-539.

Forsythe R, Nelson F, Neumann GR, et al. (1992) Anatomy of an experimental political stock market. *The American Economic Review* 82: 1142-1161.

Forsythe R, Rietz TA and Ross TW. (1999) Wishes, expectations and actions: a survey on price formation in election stock markets. *Journal of Economic Behavior & Organization* 39: 83-110.

Gelman A, Goel S, Rivers, D, et al. (2016) The mythical swing voter. *Quarterly Journal of Political Science (forthcoming)*.

Gelman A and King G. (1993) Why are American presidential election campaign polls so variable when votes are so predictable? *British Journal of Political Science* 23: 409-451.

Graefe A. (2008) Prediction markets - Defining events and motivating participation. *Foresight: The International Journal of Applied Forecasting* 9: 30-32.

Graefe A. (2011) Prediction market accuracy for business forecasting. In: Vaughan Williams L (ed) *Prediction Markets: Theory and Applications.* New York: Routledge, 87-95.

Graefe A. (2014) Accuracy of vote expectation surveys in forecasting elections. *Public Opinion Quarterly* 78: 204-232.

Graefe A. (2015a) Forecasting proportional representation elections from non-representative expectation surveys. *Electoral Studies (under review)*.

Graefe A. (2015b) German election forecasting: Comparing and combining methods for 2013. *German Politics* 24: 195-204.

Graefe A and Armstrong JS. (2011) Comparing face-to-face meetings, nominal groups, Delphi and prediction markets on an estimation task. *International Journal of Forecasting* 27: 183-195.

Graefe A, Armstrong JS, Jones RJJ, et al. (2014a) Accuracy of Combined Forecasts for the 2012 Presidential Election: The PollyVote. *PS: Political Science & Politics* 47: 427-431.

Graefe A, Armstrong JS, Jones RJJ, et al. (2014b) Combining forecasts: An application to elections. *International Journal of Forecasting* 30: 43-54.

Graefe A, Green KC and Armstrong JS. (2013) Forecasting. In: Gass SI and Fu MC (eds) *Encyclopedia of Operations Research and Management Science.* 3 ed. New York: Springer, 539-604.

Hansen J, Schmidt C and Strobel M. (2004) Manipulation in political stock markets – preconditions and evidence. *Applied Economics Letters* 11: 459-463.

Hanson R. (2007) The Policy Analysis Market (A thwarted experiment in the use of prediction markets for public policy). *Innovations: Technology, Governance, Globalization* 2: 73-88.

Hanson R, Oprea R and Porter D. (2006) Information aggregation and manipulation in an experimental market. *Journal of Economic Behavior & Organization* 60: 449-459.

Hayes SPJ. (1936) The predictive ability of voters. *Journal of Social Psychology* 7: 183-191.

Hillygus DS. (2011) The evolution of election polling in the United States. *Public Opinion Quarterly* 75: 962-981.

Kernell S. (2000) Life before polls: Ohio politicians predict the 1828 presidential vote. *PS: Political Science and Politics* 33: 569-574.

Kohut A, Keeter S, Doherty C, et al. (2012) Assessing the representativeness of public opinion surveys. In: Pew Research Center for the People & The Press (ed). Washington, D.C.

Lewis-Beck MS and Skalaban A. (1989) Citizen forecasting: can voters see into the future? *British Journal of Political Science* 19: 146-153.

Lewis-Beck MS and Stegmaier M. (2011) Citizen forecasting: Can UK voters see the future? *Electoral Studies* 30: 264-268.

Lewis-Beck MS and Tien C. (1999) Voters as forecasters: a micromodel of election prediction. *International Journal of Forecasting* 15: 175-184.

Miller MK, Wang G, Kulkarni SR, et al. (2012) Citizen Forecasts of the 2008 US Presidential Election. *Politics & Policy* 40: 1019-1052.

Murr AE. (2011) "Wisdom of crowds"? A decentralised election forecasting model that uses citizens' local expectations. *Electoral Studies* 30: 771-783.

Murr AE. (2015) The wisdom of crowds: Applying Condorcet's jury theorem to forecasting US presidential elections. *International Journal of Forecasting* 31: 916-929.

Murr AE. (this volume) Wisdom of crowds. In: Arzheimer K, Evans J and Lewis-Beck MS (eds) *The SAGE Handbook of Electoral Behaviour.*

Noelle-Neumann E. (1974) The Spiral of Silence A theory of public opinion. *Journal of Communication* 24: 43-51.

Rhode PW and Strumpf KS. (2004) Historical presidential betting markets. *Journal of Economic Perspectives* 18: 127-141.

Rhode PW and Strumpf KS. (2014) The long history of political betting. In: Vaughan Williams L and Siegel DS (eds) *The Oxford Handbook of the Economics of Gambling.* Oxford: Oxford University Press, 560-586.

Rosenbloom ES and Notz W. (2006) Statistical tests of real-money versus play-money prediction markets. *Electronic markets* 16: 63-69.

Rothschild D and Sethi R. (2015) Wishful thinking, manipulation, and the wisdom of crowds: Evidence from a political betting market. Available at: dx.doi.org/10.2139/ssrn.2322420.

Rothschild D and Wolfers J. (2012) Forecasting elections: voter intentions versus expectations. *Working paper*: Available at: ssrn.com/abstract=1884644.

Servan-Schreiber E, Wolfers J, Pennock DM, et al. (2004) Prediction markets: Does money matter? *Electronic markets* 14: 243-251.

Simon HA. (1954) Bandwagon and underdog effects and the possibility of election predictions. *Public Opinion Quarterly* 18: 245-253.

Sjöberg L. (2009) Are all crowds equally wise? A comparison of political election forecasts by experts and the public. *Journal of Forecasting* 28: 1-18.

Snowberg E, Wolfers J and Zitzewitz E. (2013) Prediction Markets for Economic Forecasting. In: Graham E and Timmermann A (eds) *Handbook of Economic Forecasting.* Elsevier.

Sunstein CR. (2006) *Infotopia: How Many Minds Produce Knowledge,* New York: Oxford University Press.

Surowiecki J. (2004) *The Wisdom of Crowds,* London: Little, Brown Book Group.

Vaughan Williams L and Paton D. (2015) Forecasting the outcome of closed-door decisions: Evidence from 500 years of betting on papal conclaves. *Journal of Forecasting* 34: 391-404.

Wang W, Rothschild D, Goel S, et al. (2015) Forecasting elections with non-representative polls. *International Journal of Forecasting* 31: 980-991.

Wolfers J and Zitzewitz E. (2004) Prediction markets. *Journal of Economic Perspectives* 18: 107-126.

Wolfram T. (2015) Have corporate prediction markets had their heyday? *Foresight: The International Journal of Applied Forecasting* 37: 29-36.

Appendix I: Overview of elections

Country	N	Elections	Sources
Germany	20	Federal (1990; 1994; 1998; 2002; 2005; 2009; 2013), European Parliament (2009; 2014), Regional (Hesse 1991; 1999; Bavaria 1994; Berlin 1999; Brandenburg 1999; Baden-Württemberg 2001; North Rhine-Westphalia 2000; Rhineland-Palatinate 2001; Saxony 1999; Saxony-Anhalt 1998; Schleswig-Holstein 1999)	(Brüggelambert, 2004; Graefe, 2015b; Berlemann and Schmidt, 2001; Hoops, 2015)
Austria	10	Federal (1995; 1995; 1999; 2002), Regional (Styria 1995; Vienna 1996, 2001), President (1998), EU parliament (1996; 1999)	(Hofinger and Ogris, 2002; Filzmaier et al., 2003)
US	8	President (1988; 1992; 1996; 2000; 2004; 2008; 2012), Senate (2010)	(Cuzán, 2011; Graefe, 2014; Graefe et al., 2014b)
Canada	2	Federal (1993; 1997)	(Forsythe et al., 1995; Antweiler and Ross, 1998)
Netherlands	1	Second chamber (1994)	(Jacobsen et al., 2000)
Norway	1	Federal (2009)	(Arnesen, 2011b)
Sweden	1	EU referendum (1994)	(Bohm and Sonnegard, 1999)
Switzerland	1	Federal (2011)	(Rau, 2011)