INFORMATION FLOW BETWEEN PREDICTION MARKETS, POLLS AND MEDIA: EVIDENCE FROM THE 2008 PRESIDENTIAL PRIMARIES

URMEE KHAN AND ROBERT P. LIELI

Abstract. We conduct a comparative analysis of trial-heat opinion polls and election prediction markets as information aggregation mechanisms. We present both cross sectional and time series evidence from the 2008 Democratic Party nomination race between Barack Obama and Hillary Clinton showing that market prices contain information about election outcomes that polls do not, while part of the surprise relative to polls is genuine news to the market as well. We then ask the question whether the apparent information surplus in prices could come from media coverage of these campaigns, and uncover interesting relationships between the pollster's surprise, price movements, and various aspects of media coverage. In particular, prices are shown to anticipate the balance and content of media coverage, but these variables do not Granger cause pollster's surprise. The volume of media coverage, on the other hand, is a significant predictor, but appears not to be incorporated in the price. Taken together, the results are consistent with the mixed evidence on the relative forecasting performance of polls vs. election markets.

Keywords: Prediction markets, opinion polls, information aggregation, media coverage

1. Introduction

With the spread of digital technology prediction markets have emerged as a competitive alternative to traditional polls in producing political forecasts. But at the core, the key issue remains unchanged: collecting and aggregating relevant information and employing it to produce accurate predictions. The efficacy of political forecasting has been front and center in public discourse following the 2016 U.S. presidential election cycle. Widespread failure of forecasting Trump as the winner has set off intense discussion of polling methodologies and

The paper is based on a PhD dissertation chapter of the first author. Many thanks to Joseph Cummins, Gloria Gonzalez-Rivera and Maxwell Stinchcombe for helpful comments and suggestions. All errors are our responsibility.

the possible causes of mismeasurement (Lohr and Singer [2016], Mercer et al. [2016], Bialik and Enten [2016]). In fact, similar questions were raised already after the 2008 primary cycle in light of the fact that the Democratic winner in competitive states was consistently underpolled (Traugott and Wlezien [2009], Traugott et al. [2008]). While the failure of prediction markets has received less public attention, the fact remains that they, too, fared very poorly in forecasting the outcome of the last presidential election. A rather stunning illustration is provided by the daily evolution of vote share and winner-take-all security prices in the Iowa Electronic Markets (IEM). Even a casual glance at these time series shows that, up until the very last moment, markets considered Clinton as the overwhelming favorite.[1]

The theoretical and empirical literature on prediction markets as a forecasting mechanism in general, and as a mechanism for political forecasting in particular, has grown rapidly over the last two decades (see, e.g., Chen and Plott [2002], Arrow et al. [2008] for early endorsements, Tziralis and Tatsiopoulos [2007], Horn et al. [2014] and the references therein for a broad overview). In comparing polls and prediction markets, the empirical literature has typically focused on forecast accuracy. In a series of papers, Berg et al. [2003, 2008] use IEM data to argue that prediction market prices have significantly lower forecast errors both in the short and the long run compared to contemporaneous polls. Results in a similar vein are obtained by Leigh and Wolfers [2006]. Nevertheless, recent events and other concurrent research paint a more complicated picture. As pointed out above, the cautionary tale provided by the 2016 U.S. election cycle applies to markets as well as polls. Jacobsen et al. [2000] present earlier evidence of market based forecasts for European elections not being particularly accurate. Using data from the same time period as Berg et al. [2008], Erikson and Wlezien [2008] construct projections of vote shares and win probabilities based on daily polls and show that these projections can outperform prediction markets. Using a long series of historical data on election betting markets dating back to the early 20th century, Erikson and Wlezien [2012a] present comparative analysis of markets before and after the introduction of scientic

[1]See http://tippie.biz.uiowa.edu/iem/. When accessed on Sept. 18, 2017, these figures were still prominently displayed on the front page.

polling, and provide intriguing evidence that market prices were far better predictors in the period without polls, than when polls were available.

From a theoretical perspective, one could argue that direct comparisons of forecast accuracy across polls and markets are somewhat odd given the fact that the two institutions are designed to answer different questions (who do you support versus who do you think will win). Accordingly, the election forecasting literature has recently proposed to poll *voter expectations* rather than *voter intentions*. As Graefe [2014] and Rothschild and Wolfers [2013] show, the first kind of poll may indeed outperform the second in forecasting election outcomes and vote shares.

Already this brief review of the literature suggests that claims about the superiority of markets as information gathering and processing devices need qualification, and the relative and absolute forecasting capacity of polls and markets warrants further investigation from various angles. Our contribution is precisely to provide a separate, and complementary, approach to comparing polls and prediction markets. The goal is not to analyze forecast accuracy *per se* but rather the dynamics of information assimilation and exchange by these institutions. Ultimately, polls and election markets both aggregate information about the voting public, so there is likely to be information flow from one to the other, inducing correlation in forecasting performance. But when the two methods do perform differently in terms of accuracy, the difference is likely to stem from information that is not common to the two aggregation methods. To what extent does the information content of polls and prediction markets differ at a point in time? Does news relative to polls move markets or do markets anticipate such news at least in part? In the latter case, do market prices really contain information not available elsewhere, or do they contain public information conveyed by the news media, but not (yet) reflected in polls? We investigate these questions through a unique case study—the race between Barack Obama and Hillary Clinton for the 2008 Democratic presidential nomination in the U.S. This historic primary season provides for an unusually long sequence of highly contested elections with rich data on pre-election polls, various prediction market prices and media coverage.

To further motivate and operationalize these ideas, it is helpful to start from the null hypothesis that election markets generally provide more accurate predictions than polls. Any mechanical demonstration of this hypothesis would immediately beg the question why the observed discrepancy in forecasting performance arises. The most natural theoretical explanation, motivated by various forms of the efficient market hypothesis, is that market prices incorporate information that polls do not.[2] How can this explanation be tested empirically? Where could the extra information come from? To answer such questions, we start from a variable called the 'pollster's surprise', defined as the difference between Obama's actual vote share in a state primary and the vote share predicted by the latest round of polls available before the election. We regard this variable as a function of information not contained in polls[3], and we correlate it with various prediction market prices both in cross section (across states) and over time.

Nevertheless, pollster's surprise is still a black box in that it does not reveal what type of information is being missed by polls. There is a widely held view among political scientists that there are 'fundamentals' of the political landscape that can predict the eventual election outcomes, regardless of the variations in trial-heat polling (Gelman and King [1993]). The theoretical role of political campaigns presumably lies in focusing the attention of the voters, and making them aware of the relevant fundamentals. Put more philosophically, the question is how the public actually learns the 'knowables' about the election (Erikson and Wlezien [2012b]). Almost by definition, the source of public information about political campaigns is media coverage. This includes coverage of poll numbers, but potentially other qualitative information that helps reveal systematic biases in polls or simply takes time to be internalized by them. We therefore construct various time series measures of campaign media coverage and correlate it with prediction market prices as well as pollster's surprise over time. We also analyze prices, pollster's surprise and media coverage jointly as a trivariate system.

The use of media variables also motivated by previous empirical work. A growing body of literature at the intersection of political science and communication theory studies media coverage of elections in general, and coverage of election related opinion polls in particular

[2]For example, Kou and Sobel [2004] provide a theoretical model for comparing prediction markets and polls in a time series setting, where the crucial assumption is that the information set of the poll respondents is contained in the information set of the traders.

[3]In Section 4.1 we show that pollster's surprise is not systematically related to the polled level of support.

(see, e.g., Holtz-Bacha and Strömbäck [2012], Searles et al. [2016] and references therein). An important observation is that media coverage of the universe of polls produced during an election cycle is subject to editorial choices and selection bias (see Searles et al. [2016] for evidence from the 2008 U.S. elections). This implies that in addition to the published poll numbers themselves, information about the nature of media coverage or the editorial process may well be useful in predicting results. That the raw frequency of media mentions of a candidate can be predictive of election outcomes is shown by Veronis [2007][4] for newspapers and French presidential elections, and by Tumasjan et al. [2010] for Twitter posts and German federal elections. As for markets and media, Lerman et al. [2008] study the relationship between media coverage and prediction markets for the 2004 U.S. presidential election cycle. They show that information obtained from automated text analysis of daily newspaper articles can be used to construct trading strategies that are more profitable than purely market based ones. This result contradicts the efficient market hypothesis, and suggests that the extent to which markets incorporate information from media coverage is a pertinent empirical question.

The bulk of our statistical analysis consists of bivariate tests using cross sectional as well as time series data. The cross sectional data set takes each state level election as a separate data point; we consider only those states where (i) poll data are available in the week before the election (allowing the construction of the state level pollster's surprise for Obama); and (ii) there is a state level prediction market that trades the 'Obama wins state' winner-take-all security.[5] We find significant positive correlation between the state level pollster's surprise and the market price the day before the election, which suggests that prices eventually contain information about election results beyond polls. We also construct a daily time series of pollster's surprises, where the surprise is zero for days with no primaries, and a

[4]Veronis passed away in 2013. The (unpublished) original paper, titled "Citations dans la presse et résultats du premier tour de la présidentielle 2007", is no longer available online but is discussed in several articles that cite it, e.g, Tumasjan et al. [2010].

[5]These markets were run by INTRADE, a now defunct electronic market maker. We thank a referee for pointing out that data from these markets are still available online.

weighted sum of the state level surprises for days with primaries.[6] We then take the daily price changes of the 'Obama wins overall nomination' Arrow-Debreu security on IEM, and conduct bivariate Granger causality tests between the two variables.[7] We find significant causality from pollster's surprise to price changes, i.e., part of what appears to be a surprise from the pollster's standpoint is also a genuine surprise for the market. Though somewhat less robust to specification and technical issues[8], we also find evidence that price changes in the overall nomination market anticipate the aggregate surprise time series, albeit this market is further removed from any individual state primary. With some caveats, this again suggests that markets possess information not available in polls.

The bivariate results between price and pollster's surprise provide the final empirical justification for including media coverage into the analysis. If prices do contain information beyond polls, it is natural to ask whether this 'information surplus' is private or, at least partly, reflected by media coverage. While it is easy to talk about media coverage in the abstract, in reality it is a complex and multi-faceted process. In order to measure various aspects of it, we draw on a unique and rich data set, created during the 2008 election cycle by the Pew Research Center Project for Excellence in Journalism (PEJ). We construct three daily times series of media coverage indices from several variables describing the media coverage received by the candidates. One captures the raw *volume* of media coverage of Obama and Clinton in various media outlets; one captures the *balance* of media coverage received by Obama versus Clinton; and, lastly, a proxy that attempts to capture how favorable the *content* of media coverage is from Obama's standpoint. The media content index classifies each day during the sample period into one of three categories: (i) days on which positive coverage of Obama dominates negative coverage (a positive index value); (ii) days on which

[6]The sample period is 1/1/2008 through 6/15/2008, 167 days altogether. The weight we attach to each state reflects the percentage of total delegates at stake in the state primary.

[7]In intuitive terms, the variable X Granger causes Y, if the history of X helps predict future values of Y even after taking the own history of Y into account; see Section 3.

[8]A pertinent issue is that the pollster's surprise time series is zero on most days with a handful of nonzero (mostly positive) spikes. Using such a markedly nonlinear process as the dependent variable in a linear autoregressive model is inherently problematic. Thus, in testing the null hypothesis that pollster's surprise is Granger caused by some other variable X, we employ the Sims test version, which uses X as the dependent variable, and lags and *leads* of surprise as regressors (in addition to lags of X). See Section 3 for details.

negative coverage of Obama dominates positive coverage (a negative index value); and (iii) days on which no such determination can be made (zero index value, about two thirds of the sample period).

We first employ the media coverage indices in pairwise Granger causality tests with pollster's surprise and the price changes of the 'Obama wins nomination' IEM security. Regarding pollster's surprise, we find strong causality from media volume to surprise, and weaker, less robust causality from surprise to media volume. The former result could be somewhat mechanical as media coverage likely intensifies leading up to state elections, and Obama's surprises are predominantly positive. The latter result indicates that election surprises generate extra media coverage over the following days. There is only weak evidence of causality from media balance to surprise (and none the other way), and no evidence of causality involving the media content index. Regarding price movements, media volume is not a significant predictor, nor is it predicted by past price changes. Nevertheless, both price changes and media volume cause surprise in pairwise tests, which suggests that their information content about future surprises is orthogonal. Media balance or content do not cause price changes either, but there is weak causality from price movements to media balance. Most interestingly, we present some evidence (both formally and visually) that prices anticipate the media content index, i.e., price changes tend to be above average (and positive) already before a 'good media day' for Obama, and below average (or even slightly negative) before 'bad media days'. No such discrepancy in price changes can be observed *after* bad/good news days. However, as mentioned above, the number of identifiable good/bad media days is rather small in the sample, so we consider these findings suggestive rather than conclusive. At the very least, the results can serve as a pointer for future investigations.

Finally, we conduct some trivariate Granger causality tests between price changes, pollster's surprise and media coverage (one index at a time) to refine our understanding of the dynamic relationships between these variables. The highlights from the trivariate exercise are as follows. The Granger causality tests from price to surprise and surprise to price barely change when controlling for the history of any of the media variables. This suggests that price movements contain information about future election surprises above and beyond the media variables. Furthermore, causality from media volume to pollster's surprise also persists when controlling for past price movements, meaning the volume of media coverage contains information about pollster's surprise above and beyond price movements. This finding is in line with the bivariate results, and in contradiction with the efficient market hypothesis. In contrast, the already weaker causality from media balance to pollster's surprise completely disappears when controlling for past price movements, i.e., prices do seem to internalize this aspect of the media coverage. Finally, the evidence of causality from surprise to media volume becomes stronger when controlling for prices.

These empirical results are consistent with markets potentially giving more accurate short-term election forecasts than polls alone due to *some* type of extra information. While prices seem to incorporate information ahead of time about the balance and content of media coverage, but these variables do not help predict pollster's surprise. (Of course, media balance and content may very well affect public opinion in a way that *is* reflected in polls.) Nevertheless, there is an aspect of media coverage (volume) that appears informative about pollster's surprise but which prices do not incorporate. Going against market efficiency, this result suggests that poll results combined with suitable public information can potentially outperform markets. These are strong claims, and are subject to some technical caveats, but are consistent with the mixed results in the literature regarding the performance of market based vs. poll based forecasts.

The paper proceeds as follows. In Section 2, we describe the Democratic primary process and the construction of the three sets of variables (price changes, pollster's surprise, media indices) in some detail. Section 3 gives a technical overview of the estimation and testing strategy. Section 4 presents the results and Section 5 concludes. An Online Appendix contains additional information about the data sets, variable definitions, and results not reported in the main text due to space constraints.

2. Institutional background, data sources and variable definitions

In broad terms, the contest for the Democratic presidential nomination in the U.S. is conducted as follows. There is a sequence of primary elections and caucuses, where registered

Democratic voters in each state and U.S. territory vote on the set of candidates on the ballot. The primaries take place between January and June, with a large number of elections concentrated on what is called 'Super Tuesday' in February or early March. In each election, there are a number of 'pledged delegates' at stake, who are picked up by the candidates in proportion to their vote share. The number of delegates varies from state to state and is roughly proportional to the size of the Democratic base. The pledged delegates formally cast their vote at the Democratic National Convention, along with a number of 'superdelegates', whose votes are not bound by election results. Of course, by the time the Convention actually takes place, candidates with an insufficient number of delegates are expected to concede.

While the 2008 primary season started out with several candidates, it quickly turned into a highly contested two-way race between Barack Obama and Hillary Clinton. Thus, each primary had a real stake throughout the election cycle; Clinton formally conceded only after the last couple of primaries on June 3. The length and tightness of this race gives a unique opportunity to collect relatively long and meaningful time series on election outcomes, polls, prediction market prices, and media coverage.

Market data. We analyze price data from two types of election markets operating during the 2008 primary cycle. First, there were state level markets, run by INTRADE, in which traders could essentially bet on the outcome of the Democratic primary in that state. By design, these markets ceased to function after the election had taken place, and their overall depth depended on the prominence of the state in the primary process, with the bulk of the trading activity typically taking place in the last few days before the election. We specifically focus on the binary (winner-take-all) market for Obama carrying the state, and record the closing price of the corresponding security on the last day before the primary. We use these prices in a cross-sectional analysis of state markets.

Second, IEM ran markets for the ultimate Democratic presidential nomination.[9] Here we also focus on the 'Obama wins nomination' winner-take-all security, and collect the daily time series of closing prices between January 1, 2008 and June 15, 2008 (167 days overall). The

[9]See http://tippie.biz.uiowa.edu/iem/closed/Nomination08.html; accessed on September 18, 2017.

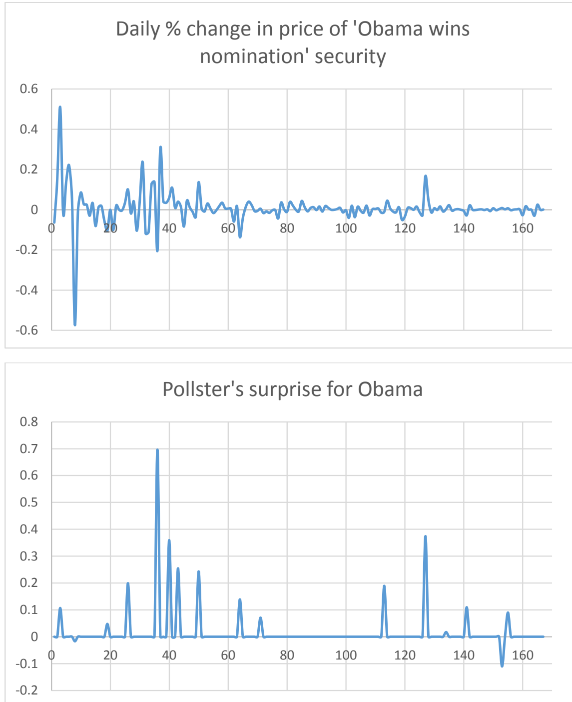

FIGURE 1. The price and surprise variables

price series is rendered stationary by taking its log-difference, yielding day-to-day percentage changes. The data are displayed on the top panel of Figure 1.

Poll data and pollster's surprise. The data on state polls and primary outcomes come from the Real Clear Politics (RCP) website and Wikipedia.[10] The polls in the respective states (or territories) are conducted on likely voters, asking who they would support in the upcoming primary. For each primary election, we consider the difference between Obama's actual vote share and his predicted vote share as measured by the latest poll data available before the

[10]The relevant web addresses are (last accessed on February 16, 2017):

http://www.realclearpolitics.com/epolls/2008/president/democratic_delegate_count.html

https://en.wikipedia.org/wiki/Statewide_opinion_polling_for_the_Democratic_Party_presidential_primaries,_2008 election in that state.[11] These raw differences are used directly in the cross sectional analysis of state markets, and are further aggregated before relating them to the evolution of prices in the market for overall nomination. In particular, we multiply the raw differences by the fraction of total delegates at stake in each state; hence, a one percentage point difference in actual versus predicted vote share in a 'big' state counts as a larger surprise than the same difference in a 'small' state. To compute total surprise for a given day, we add up the weighted surprise measures for all elections held that day. If there are no elections on the day in question, we set the surprise measure to zero. Therefore, the daily surprise time series consists of a sequence of spikes concentrated on election days; see Figure 1.

There are some data availability and interpretation issues that affect the construction of the pollster's surprise. In most states the last available poll was held just a couple of days before the election, but there are some (typically small) states where the last poll came months earlier or there was no poll data available at all. To ensure a consistent interpretation of the surprise figures across elections, states with old or no polls were dropped from the analysis. Two large states, Michigan and Florida, were also dropped due to the controversies surrounding the Democratic primaries in those states.[12] Overall, the elections for which sufficiently recent poll data was available account for over 90% of all delegates at stake (not counting Florida and Michigan delegates). The election data and the construction of the surprise index is described in more detail in Online Appendix A.

Media data. We construct three indices capturing different aspects of media coverage, using data compiled by the Pew Research Center Project for Excellence in Journalism (PEJ).[13] During the campaign period, PEJ monitored and sampled various media outlets on a daily

[11] There are well-established stylized facts about opinion polls being less informative earlier in the election cycle (Erikson and Wlezien [2012b]). By using the latest polls, we maximize their information content.

[12] In short, delegates from these states were originally stripped of voting rights at the Democratic National Convention as a penalty for these states not complying with party directives on the timing of primaries. While the rights of Michigan and Florida delegates were eventually restored, this happened only months after the primary date, and for a long time it was not clear that the primaries held in these states would count at all. Obama did not even appear on the Michigan ballot. Links on the Wikipedia page cited in footnote 10 contain further details.

[13] See `http://www.journalism.org`; accessed on February 17, 2017.

basis and compiled a database of news stories. Stories were classified according to topic, lead newsmaker and placement prominence. Word counts for print and web stories and duration-in-seconds for broadcast/cable stories were also recorded. For convenience, we provide a brief summary of PEJ's sampling process in Online Appendix B.1.

For our study, we select the stories that were categorized under the topic 'campaign' in the relevant date range (January 1, 2008 to June 15, 2008), had Barack Obama or Hillary Clinton as the lead newsmaker, were news items rather than opinion pieces, and had relatively high prominence in their placements. (The precise definition of the applied filter is given in Online Appendix B.2.) Separately for the two candidates, we construct time series variables representing the (i) number of newspaper (print) stories per day; (ii) number of web stories per day; (iii) their combined word count; (iv) number of TV and radio stories per day; and (v) the total duration of TV and Radio stories (in seconds). For a given candidate, we combine the five component measures of media exposure into a single variable by extracting their first principal component. Our first media coverage index is the sum of the principal components for the two candidates, interpreted as a measure of the raw *volume* of media coverage pertaining to the race. Our second index is the difference between the same two principal components (Obama minus Hillary), interpreted as the *balance* of media coverage across the two candidates. See Online Appendix B.2. for more information on the interpretation of the principal components and the derived indices.

The third and final media index uses additional data and attempts to capture the qualitative content of media coverage. Using the 'substoryline' codes assigned by PEJ, we attempt to identify TV and Radio stories that can be classified as 'good' or 'bad' from Obama's standpoint with reasonable certainty. The substorylines refer to specific events or news items such as 'Clinton, Obama spar over MLK comments', 'Kennedy family endorses Obama', etc., but not the primary results themselves.[14] Where necessary, we googled these events for their broader context to make a (subjective) call on their classification, or whether to drop them altogether. The final list of the employed substorylines and their classification is given in Online Appendix B.3. On each day with some storylines identified as 'good' or 'bad', we

[14]General campaign and election coverage is lumped into the 'other' substoryline, which we exclude from the analysis, as its precise content and classification is ambiguous on any given day.

add up the duration of stories good for Obama vs. bad for Obama, and take the difference. The days on which no such stories can be identified are assigned the value zero (110 days altogether). The resulting variable is our 'media content' index, and it consists of 57 non-zero of spikes varied in sign and size.

3. Time series methodology: bivaraite and trivariate Granger causality

Granger causality, since its inception by Granger [1969], has become a widely used concept to assess whether a time series is Y contains unique information about the future of another time series X. Here we restate some basic definitions and provide a brief description of the test variants we employ. We then discuss implementation issues and limitations.

Let $W_t = (X_t, Y_t)'$ be a bivariate covariance stationary process. A fairly narrow (but practical) definition of bivariate Granger causality states that Y does not Granger cause X at horizon $h \geq 1$ if the linear projection of X_{t+h} on the information set $I_X(t) = \{1, X_t, X_{t-1}, \ldots\}$ coincides with the linear projection of X_{t+h} on the larger information set $I_W(t) = \{1, X_t, X_{t-1}, \ldots, Y_t, Y_{t-1}, \ldots\}$. Assuming that the linear projections involved in the definition can be well approximated by a VAR(K) model with a finite number of lags K, one can readily test for causality going from Y to X by estimating the regression equation

$$X_t = \alpha_0 + \alpha_1 X_{t-1} + \ldots + \alpha_K X_{t-K} + \beta_1 Y_{t-1} + \ldots + \beta_K Y_{t-K} + \epsilon_t, \tag{1}$$

and conducting, say, a Wald test of the hypothesis $H_0 : \beta_1 = \ldots = \beta_K = 0$. The horizon h does not play a crucial role in this bivariate setting as it can be shown that Y causes X at horizon $h = 1$ in a VAR(K) model if and only if Y causes X at all horizons. To test for causality in the reverse direction, i.e. going from X to Y, one can simply interchange the role of Y and X in equation (1).

A technical problem with the test based on model (1) is that in our application some of the variables of interest, namely, the pollster's surprise and the media content index, are highly nonlinear (a sequence of spikes with zeros in between). In general, if X is such a variable, and causality from Y to X is being tested, it is unlikely that a finite order autoregression provides a good approximation to the linear projection of X_t on $I_W(t)$, meaning that the error term ϵ_t is severely heteroskedastic and/or autocorrelated for any finite K even under the null. While using a robust Wald test potentially mitigates the problem, there is another test strategy that avoids the use of such X as the dependent variable altogether. The method is due to Sims [1972] and Geweke et al. [1983], and consists of regressing the potential causing variable Y on its own history as well as past, current, and *future* values of X:

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \ldots + \alpha_K Y_{t-K} + \beta_1 X_{t-1} + \ldots + \beta_K X_{t-K} + \gamma_0 X_t + \gamma_1 X_{t+1} + \ldots \gamma_K X_{t+K} + \epsilon_t. \tag{2}$$

Lack of Granger causality from Y to X corresponds to the condition $H_0 : \gamma_1 = \ldots = \gamma_K = 0$. This approach to testing Granger causality allows us to avoid using pollster's surprise as the dependent variable in a linear model.

The definition of Granger causality readily extends to trivariate settings with the (covariance stationary) process $W_t = (X_t, Y_t, Z_t)'$ now including a 'control' or auxiliary variable Z. Let $I_W(t) = \{1, X_t, Y_t, Z_t, X_{t-1}, Y_{t-1}, Z_{t-1}, \ldots\}$ and, similarly, let $I_{XZ}(t)$ denote the smaller information set that consists of the history of X and Z only. In this trivariate setting we say that Y does not (linearly) cause X at horizon $h > 0$ in the presence of Z, denoted $Y \overset{h}{\nrightarrow} X|Z$, if the linear projection of X_{t+h} on $I_W(t)$ coincides with the linear projection of X_{t+h} on $I_{XZ}(t)$. Furthermore, Y does not (linearly) cause X up to horizon $h > 0$ in the presence of Z, denoted $Y \overset{(h)}{\nrightarrow} X|Z$ if $Y \overset{j}{\nrightarrow} X|Z$ for $j = 1, \ldots, h$.

A natural starting point for testing lack of causality from Y to X in the presence Z is to augment the regressions (1) and (2) with the explanatory variables $Z_{t-1}, \ldots, Z_{t-K}$. Nevertheless, the trivariate setting is complicated by the fact that acceptance of $H_0 : \beta_1 = \ldots = \beta_K = 0$ or $H_0 : \gamma_1 = \ldots = \gamma_K = 0$ only implies non-causality at horizon $h = 1$; causality may still exist at horizons $h > 1$, which needs to be tested for separately.[15] As pointed out by Hill [2007], non-causality from Y to X at horizon h can be followed by causation at horizon $h + 1$ only if an indirect causality chain $Y \overset{1}{\rightarrow} Z|X$, $Z \overset{1}{\rightarrow} X|Y$ exists. Hence, testing for such a chain can be an intermediate step in testing for longer term causality. If a chain exists, Hill [2007] proposes further joint tests for $Y \overset{(h)}{\nrightarrow} X|Z$, $h \geq 2$. Alternatively, one can conduct direct tests of long term causality $Y \overset{h}{\nrightarrow} X|Z$ individually for each horizon $h \geq 2$, as proposed by Dufour and Renault [1998]. This last test is easily implemented by replacing the dependent variable X_t in regression (1) with X_{t+h-1} (and adding $Z_{t-1}, \ldots, Z_{t-K}$ as explanatory variables).

[15] Existence of causality at horizon $h = 1$ generally implies causality at further horizons.

Limitations. The definition of Granger causality given above is based on the linear projection of Y_{t+h} on $I_W(t)$, which is a rather narrow aspect of the joint distribution of these two objects. Under fairly mild additional assumptions, e.g., that the error ϵ_t is a martingale difference sequence w.r.t. $I_W(t)$, the tests discussed above can be reinterpreted as tests of causality in the conditional mean. More generally, Granger causality from Y to X could (also) be conveyed by higher moments, which will not be detected by these tests. Nevertheless, with the limited amount of data available the implementation of more demanding higher order causality tests is rather pointless, even though it would be an interesting exercise given the obviously nonlinear nature of some of our variables.[16]

Another concern is that using conventional nominal significance levels for individual tests might lead to overrejection due to specification search and the large number of tests performed. We do not formally deal with this issue but note that our strongest results are based on p-values that very close to zero and reasonably robust across choices of K.

4. Empirical results

4.1. **Cross sectional evidence.** Table 1 displays the coefficients from three linear regressions. In each regression, the dependent variable is the (unweighted) state level pollster's surprise for Obama, measured in percentage points. In column (a) the sole explanatory variable is Obama's vote share in the last available poll. The slope coefficient is not statistically different from zero, supporting the interpretation that the pollster's surprise is a function of information not available in polls. Column (b) shows the regression where the explanatory variable is the closing price of the 'Obama wins' Arrow-Debreu security on the state level prediction market the day before the election (the price in cents). The slope coefficient is positive and highly significant; in fact, the R-squared statistic indicates that as much as 20 percent of the variation in the pollster's surprise is eventually picked up by the market price. The data and the fitted regression line are depicted in Figure 2. Replacing the independent

[16] See, for example, Comte and Lieberman [2000] for very careful definitions of causality up to second order and their relationships. The literature on nonlinear and nonparametric Granger causality tests is also huge; a relatively early example is Hiemstra and Jones [1994], and a relatively recent one is Taamouti et al. [2014].

TABLE 1. Regressions of state level pollster's surprise on last poll and last price

Expl. variable	(a)	(b)	(c)
Last poll	0.0877		-0.1903
s.e.	(0.0952)		(0.1188)
t-stat	[0.92]		[-1.60]
p-val	[0.363]		[0.118]
Closing price day before		0.0706	0.1070
s.e.		(0.0233)	(0.0322)
t-stat		[3.02]	[3.32]
p-val		[0.005]	[0.002]
Constant	0.8941	0.7815	7.188
Sample size	38	38	38
R-squared	0.023	0.203	0.215

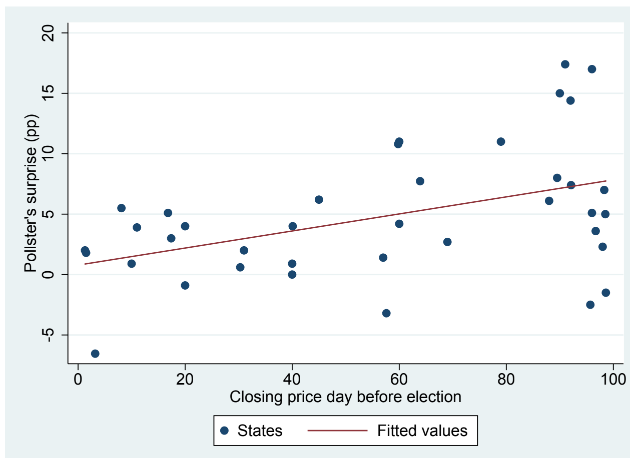

FIGURE 2. Regression of pollster's surprise on closing prices

variable with closing prices two days before the election produces very similar results. Finally, column (c) includes both explanatory variables as a robustness check; the significant predictive power of prices is maintained.

Thus, the cross sectional results strongly indicate that by the election date market prices will have incorporated information not available in polls.

4.2. **Time series evidence: bivariate Granger causality.** We now present evidence on the dynamic relationships between our time series variables. In what follows, 'price' will mean the daily log-price changes in the 'Obama wins Democratic nomination' winner-take-all security and 'surprise' will mean the daily time series of aggregate election surprises.

The bivariate Granger causality results are displayed in Tables 2 through 6; Table 7 provides a quick summary. For each null hypothesis of the form 'Y does *not* Granger cause X', we report results from four different tests: (i) Wald test based on equation (1) with a conventional covariance matrix; (ii) Wald test based on equation (1) with a robust covariance matrix; (iii) Wald test based on equation (2) with a conventional covariance matrix; (iv) Wald test based on equation (2) with a robust covariance matrix. We will refer to tests (i) and (ii) as a 'Granger test' and tests (iii) and (iv) as a 'Sims test'; see Section 3 for a more detailed description. Here we only note that if X is a sequence of spikes (specifically, the pollster's surprise or the media content index), then we consider the robust Sims test as the most reliable. Otherwise the robust Granger test is the benchmark procedure. We implement each type of test for a range of lag specifications, noting the preferred lag orders picked by various information criteria (IC). This allows for direct assessment of sensitivity to lag choice. Given our relatively short sample, and the daily frequency, we consider $K \leq 14$.

Pollster's surprise and prices. The top panel of Table 2 examines whether past election surprises are helpful for predicting future prices (taking the history of prices into account). The conventional and robust Granger tests are both highly significant and very robust to the number of lags included in the test equation. Only the robust Sims tests are unable to detect causality, but we discount these results as in these regressions the dependent variable is the (highly nonlinear) surprise time series. We interpret these results as strong and robust evidence that part of the pollster's surprise is genuine surprise for the market as well; in other words, prices react to election outcomes.

Conversely, causality from price to surprise is being tested in bottom panel of Table 2. Here the Sims tests are more appropriate, and they strongly reject in models that are at least moderately large (5 lags and up), with all the IC supporting these models. For the robust Sims test even the smaller models are not that far from significant rejection (with p-values of 7 and 14 percent). We discount the Granger tests as they use surprise as the dependent variable, and interpret the results as reasonably robust and strong evidence that prices anticipate part of the pollster's surprise. That is, price movements contain some information about future election outcomes not available in polls.

TABLE 2. Granger causality tests between Surprise and Price (p-values)

	H_0 : [Surprise$\nRightarrow$Price]			
	Granger test:		Sims test:	
Lag order	Non-robust var.	Robust var.	Non-robust var.	Robust var.
1	0.001	0.004	0.000	0.208
3	0.000	0.012	0.000	0.428
5	0.000	0.018	0.000	0.525
7	0.000	0.000	0.000	0.582
9	0.000	0.005	0.001	0.802
14	0.010	0.011	0.000	0.411
	AIC=9; BIC=8; HQ=8		AIC=14; BIC=1; HQ=1	
	H_0 : [Price $\nRightarrow$ Surprise]			
1	0.272	0.354	0.181	0.069
3	0.343	0.640	0.387	0.138
5	0.116	0.626	0.039	0.003
7	0.028	0.385	0.000	0.000
9	0.006	0.391	0.000	0.000
14	0.000	0.011	0.000	0.000
	AIC=14; BIC=1; HQ=1		AIC=9; BIC=8; HQ=9	

TABLE 3. Granger causality tests between Surprise and Media Volume (p-values)

	H_0 : [Surprise$\nRightarrow$Media], Media=VOLUME			
	Granger test:		Sims test:	
Lag order	Non-robust var.	Robust var.	Non-robust var.	Robust var.
1	0.008	0.012	0.004	0.010
3	0.101	0.088	0.104	0.058
5	0.029	0.054	0.120	0.028
7	0.058	0.031	0.152	0.061
9	0.326	0.220	0.196	0.061
14	0.139	0.017	0.557	0.506
	AIC=13; BIC=1; HQ=1		AIC=4; BIC=1; HQ=1	
	H_0 : [Media$\nRightarrow$Surprise], Media=VOLUME			
1	0.051	0.068	0.003	0.029
3	0.003	0.100	0.000	0.000
5	0.000	0.034	0.000	0.000
7	0.002	0.212	0.002	0.000
9	0.003	0.233	0.001	0.000
14	0.025	0.712	0.063	0.000
	AIC=4; BIC=1; HQ=4		AIC=13; BIC=1; HQ=3	

Pollster's surprise and media variables. Table 3 examines the dynamic relationship between pollster's surprise and media volume. The evidence that past election surprises predict the volume of media coverage is mixed and sensitive to model specification; see the robust Granger tests in the top panel of Table 3. Nevertheless, the test does reject well below 5% at or around the lag order picked by BIC/HQ (=1) as well as AIC (=13). Conversely, the Sims tests in the bottom panel show strong and robust causality from media volume to surprise. A possible interpretation of this result is rather mechanical—it is natural for media coverage to intensify leading up to the state primaries, and, at the same time, the aggregate surprise time series for Obama is a sequence of overwhelmingly positive spikes.

Table 4 shows the corresponding tests for balance of media coverage. There is no evidence of surprises causing balance (top panel), while the robust Sims test gives mixed results for balance causing surprise (bottom panel). Most models reject at about the 5% level, but the smallest model, picked by all three information criteria, does not. The conventional Sims tests do not reject at all. Hence, the evidence that media balance anticipates surprise is weak at best.

Finally, there is no evidence of causality between the media content index and pollster's surprise either way (tables not shown to conserve space). As discussed in Section 2, the number of identified good/bad news days is rather small, so lack of power may very well play a role in these negative results.

Prices and media variables. There is no evidence of Granger causality between prices and volume of media coverage in either direction (tables not shown to save space). Nevertheless, as shown above, prices and media volume are both reasonably strong predictors of pollster's surprise. Taken together, these results suggest that the information content of the two variables about future surprises is orthogonal. We will investigate this issue further using trivariate tests.

Table 5 presents the relationship between prices and balance of media coverage. There is rather questionable evidence of media balance anticipating prices; see the robust Granger test results in the top panel. Formally, rejections do occur for the larger models, but it is hard to rationalize why week old media coverage would matter for prices while recent

TABLE 4. Granger causality tests between Surprise and Media Balance (p-values)

	H_0 : [Surprise$\not\Rightarrow$Media], Media=BALANCE			
	Granger test:		Sims test:	
Lag order	Non-robust var.	Robust var.	Non-robust var.	Robust var.
1	0.488	0.613	0.466	0.647
3	0.615	0.630	0.628	0.607
5	0.766	0.778	0.797	0.594
7	0.855	0.852	0.732	0.347
9	0.749	0.611	0.612	0.208
14	0.452	0.105	0.570	0.713
	AIC=1; BIC=1; HQ=1		AIC=1; BIC=1; HQ=1	
	H_0 : [Media$\not\Rightarrow$Surprise], Media=BALANCE			
1	0.736	0.488	0.832	0.631
3	0.853	0.420	0.606	0.026
5	0.684	0.211	0.619	0.051
7	0.689	0.132	0.624	0.067
9	0.623	0.237	0.729	0.023
14	0.872	0.425	0.704	0.053
	AIC=1; BIC=1; HQ=1		AIC=1; BIC=1; HQ=1	

TABLE 5. Granger causality tests between Price and Media Balance (p-values)

	H_0 : [Media$\not\Rightarrow$Price], Media=BALANCE			
	Granger test:		Sims test:	
Lag order	Non-robust var.	Robust var.	Non-robust var.	Robust var.
1	0.728	0.563	0.546	0.275
3	0.749	0.307	0.815	0.179
5	0.984	0.885	0.901	0.425
7	0.366	0.017	0.494	0.149
9	0.221	0.024	0.271	0.028
14	0.269	0.007	0.369	0.105
	AIC=13; BIC=9; HQ=9		AIC=1; BIC=1; HQ=1	
	H_0 : [Price$\not\Rightarrow$Media], Media=BALANCE			
1	0.009	0.003	0.008	0.082
3	0.033	0.004	0.071	0.044
5	0.100	0.012	0.144	0.083
7	0.084	0.001	0.141	0.018
9	0.542	0.210	0.194	0.058
14	0.597	0.001	0.250	0.138
	AIC=1; BIC=1; HQ=1		AIC=9; BIC=8; HQ=9	

coverage apparently does not. (Moreover, the conventional Granger tests do not reject at all.) Conversely, the robust Granger test shows reasonably strong causality from prices to balance; see the lower panel of Table 5. The smallest model, in particular, is backed by all the IC, and rejects unambiguously.

Finally, the relationship between prices and the media content index is examined in Table 6. As indicated by the top panel, there is no evidence that the media content index causes prices (the caveat about power applies). As for prices anticipating content, the robust Sims test in the bottom panel is the benchmark, as the media content index is a sequence of spikes. The test rejects non-causality for the larger models picked by the IC, but at shorter lags as well. Again, rejection at short lags is important because it is hard to explain how price movements could anticipate, say, a random controversial remark made by Obama a week later. On the other hand, rejections at short lags could be interpreted as Obama making the controversial remark, prices reacting immediately or at least within the day, and negative media coverage gradually building up over the following few days.

TABLE 6. Granger causality tests between Price and Media Content (p-values)

	H_0 : [Media$\not\Rightarrow$Price], Media=CONTENT			
	Granger test:		Sims test:	
Lag order	Non-robust var.	Robust var.	Non-robust var.	Robust var.
1	0.850	0.445	0.979	0.933
3	0.999	0.939	0.999	0.993
5	1.000	0.975	0.962	0.401
7	0.845	0.470	0.913	0.770
9	0.617	0.074	0.644	0.751
14	0.894	0.022	0.769	0.791
	AIC=11; BIC=9; HQ=9		AIC=1; BIC=1; HQ=1	
	H_0 : [Price$\not\Rightarrow$Media], Media=CONTENT			
1	0.315	0.171	0.319	0.018
3	0.754	0.403	0.675	0.044
5	0.912	0.582	0.923	0.248
7	0.957	0.597	0.982	0.096
9	0.952	0.266	0.968	0.010
14	0.997	0.701	0.965	0.237
	AIC=1; BIC=1; HQ=1		AIC=9; BIC=8; HQ=9	

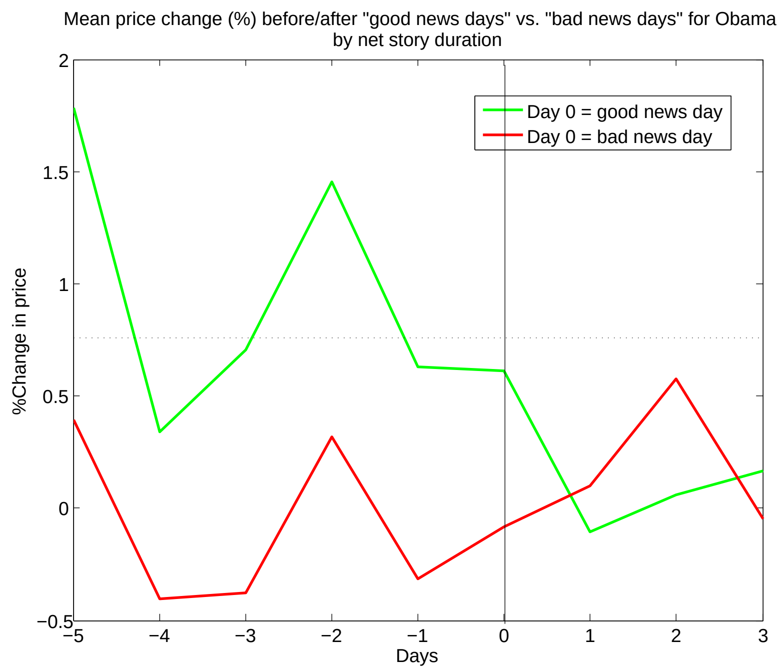

FIGURE 3. Price changes before/after bad news days vs. good news days

The result that prices anticipate the content index is further illustrated in Figure 3. A 'good news day' for Obama is when the media content index is positive, i.e., the duration of TV and Radio stories favorable for Obama exceeds the duration of unfavorable stories. 'Bad news days' are defined analogously. The identified bad/good news days tend to come in bunches as media coverage builds up gradually after a triggering event. Collectively, these informative days constitute 'Day 0' on Figure 3. The corresponding value of the red line shows the average price change on bad news days, while the green line shows the average price change on good news days. The label 'Day -1' denotes the set of all days preceding an informative day (which includes some informative days as well). The corresponding value of the red line shows the average price change on days preceding a bad news day, and the green line shows the average price change on days preceding a good news day, etc. The average price change over all days in the sample period is shown as a horizontal dotted black line.

(As the price for Obama's nomination was gradually increasing over the sample period, the overall average is positive.)

The pattern seen in Figure 3 is very suggestive. On average, price changes *before* good news days are positive and often exceed the unconditional average, while price changes *before* bad news days are below the unconditional average and often negative. The difference disappears after Day 0. While the two lines are likely not statistically distinguishable at conventional significance levels, this pattern suggests that the market anticipates and adjusts for favorable and unfavorable media coverage. While traders may actually receive private information ahead of time about certain planned events (e.g., a public figure endorsing Obama), this clearly cannot be the case for other events (random controversial remarks made in a speech). The interpretation in this case is that markets react to triggering events quicker, while conventional media coverage is slower to build up and persists for several days.

As a convenient summary, Table 7 shows all the conclusions drawn from the bivariate Granger causality tests.

Table 7. Qualitative summary of bivariate Granger causality results

	TO:				
FROM:	Surprise	Price	Volume	Balance	Content
Surprise		+ + +	+	−	−
Price	++		−	++	+
Volume	+ + +	−			
Balance	+/−	+/−			
Content	−	−			

Note: + + +: strong and robust evidence; ++: fairly strong and robust evidence; +: some evidence; +/−: weak or questionable evidence; −: no evidence.

4.3. **Time series evidence: trivariate Granger causality.** In order to corroborate and refine the bivariate result, we will briefly discuss some findings from trivariate Granger causality tests. See Section 3 for the additional technical issues involved in trivariate Granger causality testing. To save space, all tables forming the basis of the results presented here are relegated to the Online Appendix.

One step ahead tests. The first question to ask is whether the two-way Granger causality between price and pollster's surprise persists (one step ahead) when the history of media coverage is controlled for in the test equations. The short answer is yes; none of the media indices affect the price-surprise relationship appreciably. For media balance and content this is very much expected based on the bivariate results: there is at most weak evidence of media balance causing pollster's surprise on its own (none for content), and price movements already anticipate both of these variables; see Table 7. Indeed, looking at the same test equations from another perspective, the already weak causality from balance to surprise completely disappears in the presence of price. Hence, if media balance carries any information at all about election outcomes that is not reflected in polls, it is already incorporated in the price.

In contrast to balance and content, media volume (strongly) causes surprise in a bivariate setting, but adding volume as a control to the price-surprise equations still does not affect causality in either direction. Moreover, volume to surprise causality persists in the presence of price. These findings point toward prices and media volume containing *independent* information about election surprises relative to polls. This interpretation is further corroborated by the fact that price and volume do *not* Granger cause each other in a bivariate setting (see again Table 7). Of course, these results contradict the efficient market hypothesis, and imply that short term election forecasts based on polls can be enhanced by adding information about media volume in some form. These enhanced forecasts may potentially outperform markets as well, as volume information is not incorporated in the price.

A caveat, as mentioned earlier, is that the apparent causality from volume to surprise may just be an ex-post mechanical relationship due to media coverage intensifying before elections and Obama's surprises being mostly positive for independent reasons. Hence, the external validity of these findings needs further investigation outside the context of the present case study.

Another noteworthy result from the one-step-ahead trivariate tests is that the evidence of causality from surprise to media volume becomes considerably stronger in the presence of price history as a control.

Indirect causality chains and longer horizon causality. In testing direct causality between price and media volume, no relationship is found in the bivariate setting (Table 7), and one-step-ahead non-causality persists when controling for the presence of surprise. However, the results discussed in the previous subsection imply the existence of the causality chains [volume $\rightarrow_1$surprise $\rightarrow_1$price] and [price $\rightarrow_1$surprise $\rightarrow_1$volume]. These indirect channels with surprise as the mediator open the door for Granger causality between price and media volume over longer time horizons, despite the fact that price and media volume do not cause each other one step ahead in the presence of surprise. Using Step 3 of the Hill [2007] procedure supports long term (indirect) causality from media volume to price, jointly over all horizons up to a week ahead. The Dufour and Renault [1998] horizon-specific direct tests produce less conclusive results. Regarding long term (indirect) causality from price to media volume, Step 3 of the Hill [2007] procedure gives somewhat mixed evidence, but there is some support for causality.

Similarly, there is strong evidence of a [surprise $\rightarrow_1$price $\rightarrow_1$balance] causality chain based on the one step ahead results. This implies that [surprise$\rightarrow_h$balance|price] is possible for $h > 1$ despite the fact that surprise does not cause balance one step ahead (with or without price as a control). Indirect causality jointly over all horizons up to a week ahead is confirmed by the Step 3 of the Hill [2007] procedure; the horizon-specific direct tests yield mostly negative results.

5. Discussion and conclusion

We started with a set of questions about the information content of opinion polls and prediction market prices:

> To what extent does the information content of polls and prediction markets differ at a point in time? Does news relative to polls move markets or do markets anticipate such news at least in part?
>
> In the latter case, do market prices really contain information not available elsewhere, or do they contain public information conveyed by the news media, but not (yet) reflected in polls?

Regarding the first two questions, evidence from our case study suggests that both state level and national prediction market prices contain information about primary outcomes not available in raw poll numbers. On the other hand, pollster's surprise Granger causes price movements in the national market, i.e., part of the pollster's surprise seems to be genuine news to the market as well.

Regarding the last question, our analysis yields several insights. The extra information possessed by the market about election surprises is *not* accounted for by the three media coverage indices constructed here (volume, balance and content). However, the reasons differ. While prices seem to incorporate ahead of time information about media balance and content, these variables do not Granger cause election surprises (balance and content can still have an effect on election outcomes, but the effect then goes through polls). Volume of media coverage does Granger cause election surprises, but this information does not seem to be reflected in price movements.

What conclusions can we draw from our results? Overall, the picture that emerges is of a fairly complex relationship between markets, polls, and media. Specifically, the information flow into markets has efficient and inefficient features at the same time. It seems some election-relevant information implicit in media coverage is ignored by traders, but they do draw from sources other than the public media. These sources provide valuable information not available in polls. The ignored election-relevant information is likely related simply to the raw volume of media coverage. This is not intuitive, but consistent with Tumasjan et al. [2010] (volume being informative) and Lerman et al. [2008] (prediction markets ignoring information in media). In sum, these results lend further credence to the idea that combining forecasts from various methods can improve the quality of election forecasting (e.g., Rothschild [2015]).

There are of course a number caveats to our analysis and results. First of all, this is just one case study. External validity (overfitting) is naturally a concern, especially in relation to the findings about market inefficiency. But in exchange for limiting our analysis to this particular case study, we have been able to create and utilize unique datasets, both cross sectional and time series, and ask interesting questions about the role of three diverse mechanisms (markets, polls and media) in aggregating information. At the very least, we have generated interesting hypotheses which can be tested in other contexts as well.

References

Arrow, K. J., R. Forsythe, M. Gorham, R. Hahn, R. Hanson, J. O. Ledyard, S. Levmore, R. Litan, P. Milgrom, and F. D. Nelson (2008). Economics: The promise of prediction markets. *Science 320*(5878), 877.

Berg, J., F. Nelson, and T. Rietz (2003). Accuracy and forecast standard error of prediction markets. *Tippie College of Business Administration, University of Iowa*.

Berg, J. E., F. D. Nelson, and T. A. Rietz (2008). Prediction market accuracy in the long run. *International Journal of Forecasting 24*(2), 285–300.

Bialik, C. and H. Enten (2016). The polls missed trump. we asked pollsters why. https://fivethirtyeight.com/features/the-polls-missed-trump-we-asked-pollsters-why/. Accessed: 2017-09-18.

Chen, K. Y. and C. R. Plott (2002). Information aggregation mechanisms: concept, design and field implementation. *California Institute of Technology Social Science Working Paper 1131*.

Comte, F. and O. Lieberman (2000). Second-order noncausality in multivariate garch processes. *Journal of Time Series Analysis 21*(5), 535–557.

Dufour, J.-M. and E. Renault (1998). Short run and long run causality in time series: Theory. *Econometrica 66*(5), 1099–1125.

Erikson, R. S. and C. Wlezien (2008). Are political markets really superior to polls as election predictors? *Public Opinion Quarterly 72*(2), 190–215.

Erikson, R. S. and C. Wlezien (2012a). Markets vs. polls as election predictors: An historical assessment. *Electoral Studies 31*(3), 532–539.

Erikson, R. S. and C. Wlezien (2012b). *The timeline of presidential elections: How campaigns do (and do not) matter*. University of Chicago Press.

Gelman, A. and G. King (1993). Why are american presidential election campaign polls so variable when votes are so predictable? *British Journal of Political Science 23*(4), 409–451.

Geweke, J., R. Meese, and W. Dent (1983). Comparing alternative tests of causality in temporal systems analytic results and experimental evidence. *Journal of Econometrics 21*(2), 161–194.

Graefe, A. (2014). Accuracy of vote expectation surveys in forecasting elections. *Public Opinion Quarterly 78*(S1), 204–232.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica 37*, 424–438.

Hiemstra, C. and J. D. Jones (1994). Testing for linear and nonlinear granger causality in the stock price-volume relation. *The Journal of Finance 49*(5), 1639–1664.

Hill, J. B. (2007). Efficient tests of long-run causation in trivariate VAR processes with a rolling window study of the money–income relationship. *Journal of Appl. Econometrics 22*(4), 747–765.

Holtz-Bacha, C. and J. Strömbäck (2012). *Opinion polls and the media: Reflecting and Shaping Public Opinion*. Springer.

Horn, C. F., B. S. Ivens, M. Ohneberg, and A. Brem (2014). Prediction markets–a literature review 2014. *The Journal of Prediction Markets 8*(2), 89–126.

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

Kou, S. G. and M. E. Sobel (2004). Forecasting the vote: a theoretical comparison of election markets and public opinion polls. *Political Analysis 12*(3), 277–295.

Leigh, A. and J. Wolfers (2006). Competing approaches to forecasting elections: Economic models, opinion polling and prediction markets. *Economic Record 82*(258), 325–340.

Lerman, K., A. Gilder, M. Dredze, and F. Pereira (2008). Reading the markets: Forecasting public opinion of political candidates by news analysis. In *Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1*, pp. 473–480. Association for Computational Linguistics.

Lohr, S. and N. Singer (2016). How data failed us in calling an election. https://www.nytimes.com/2016/11/10/technology/the-data-said-clinton-would-win-why-you-shouldnt-have-believed-it.html. Accessed: 2017-09-18.

Mercer, A., C. Deane, and K. McGeeney (2016). Why 2016 election polls missed their mark. http://www.pewresearch.org/fact-tank/2016/11/09/why-2016-election-polls-missed-their-mark/. Accessed: 2017-09-18.

Rothschild, D. (2015). Combining forecasts for elections: Accurate, relevant, and timely. *International Journal of Forecasting 31*(3), 952–964.

Rothschild, D. and J. Wolfers (2013). Forecasting elections: Voter intentions versus expectations. *Technical Report*.

Searles, K., M. H. Ginn, and J. Nickens (2016). For whom the poll airs: Comparing poll results to television poll coverage. *Public Opinion Quarterly 80*(4), 943–963.

Sims, C. A. (1972). Money, income and causality. *American Economic Review 62*, 540–552.

Taamouti, A., T. Bouezmarni, and A. El Ghouch (2014). Nonparametric estimation and inference for conditional density based granger causality measures. *Journal of Econometrics 180*(2), 251–264.

Traugott, M., G. Bolger, D. W. Davis, C. Franklin, R. M. Groves, P. J. Lavrakas, M. S. Mellman, P. Meyer, K. Olson, J. A. Selzer, et al. (2008). An evaluation of the methodology of the 2008 pre-election primary polls. *AAPOR Ad Hoc Committee on the 2008 Presidential Primary Polling*.

Traugott, M. W. and C. Wlezien (2009). The dynamics of poll performance during the 2008 presidential nomination contest. *Public Opinion Quarterly 73*(5), 866–894.

Tumasjan, A., T. O. Sprenger, P. G. Sandner, and I. M. Welpe (2010). Predicting elections with twitter: What 140 characters reveal about political sentiment. *Icwsm 10*(1), 178–185.

Tziralis, G. and I. Tatsiopoulos (2007). Prediction markets: An extended literature review. *The Journal of Prediction Markets 1*(1), 75–91.

Department of Economics, University of California, Riverside

E-mail address: urmee.khan@@ucr.edu

Department of Economics, Central European University, Budapest

E-mail address: lielir@@ceu.edu