International Journal of Forecasting 39 (2023) 956–966

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Aggregating qualitative district-level campaign assessments to forecast election results: Evidence from Japan ☆

Michio UMEDA

Faculty of Global Media Studies, Komazawa University, 1-23-1 Komazawa, Setagaya, Tokyo 154-8525, Japan

ARTICLE INFO

Keywords:
Election forecasting
Poll aggregation
Comparative studies
Japan
Item response theory

ABSTRACT

The poll aggregation is conducted in the USA and European democracies for electoral forecasting. However, this has not been the case in Japan because the news media report on electoral campaigns with qualitative assessments rather than poll numbers, although these assessments are based on extensive polling. Our study developed an approach to aggregate qualitative district-level election campaign coverage in Japan, applied the method to forecast the outcomes of the 2017 general election for Japan's Lower House of the National Diet, and assessed the accuracy of the forecast against the actual results. We integrated the qualitative assessments by using the item response theory, which effectively predicted the electoral results. The method proposed in this paper can be applied to aggregate qualitative assessments by experts in other countries, such as the Cook Political Report in the USA. This would improve the accuracy of election forecasting when combined with existing approaches.

1. Introduction

Poll aggregation has been widely used to integrate information from polls and forecast electoral outcomes. This is due to the proliferation of pre-electoral polls and the improvement in their accuracy over time (Traugott, 2014). It is one of the main approaches for electoral forecasting, along with prediction markets and structural modeling (Lewis-Beck & Stegmaier, 2014). This approach is applied not only in elections in the USA but also in other democratic countries with several pre-electoral polls (e.g., Graefe, 2019).

Poll aggregation is not conducted in Japan for electoral forecasting. This is because the Japanese media outlets conduct polls extensively during the official campaign period. However, the media does not report the electoral races with direct poll figures as those in the other countries, such as "44% of respondents intend to vote for Candidate X" due to the presence of legal regulations for releasing raw poll numbers.[1] Instead, the media report their qualitative evaluations such as "Candidate X is ahead" or "Candidates Y and Z are in a close race" based on their polls and additional information.[2] These qualitative evaluations might be no less informative to

☆ Funding: This work was supported by the JSPS KAKENHI [grant number JP19K01472].

E-mail address: umedam@komazawa-u.ac.jp.

[1] Article 138-3 of the Electoral Law bans the publication of the "popularity poll" of the candidates who are running for office. The clause is usually interpreted as prohibiting the publication of poll results even by news media (e.g., Yasuda & Arakawa, 2009). However, in 1991, the Tokyo High Court (1991) ruled that qualitative media evaluation of the electoral contests does not violate the clause based on another clause of the law – Article 148. Hence, instead of poll figures, the Japanese media usually publish their qualitative assessments for each district.

[2] Some Japanese electoral experts also publish qualitative district-level campaign assessments in magazines a few months before the start of the official campaign (e.g., Predicting Winner, 2021). They could be used to predict the electoral outcome well before the voting day. However, this study does not use them because the experts do not clarify the method and criteria that they have used.

https://doi.org/10.1016/j.ijforecast.2022.03.006

regular voters than numerical evaluations because the media translate the level of certainty into plain language. However, we cannot use a standard averaging method, as in the other countries.

The campaign coverages by these Japanese media outlets would be similar to the expert assessments of the USA district-level campaigns, such as the race rating of the Cook Political Report (2020). Although expert assessments have sometimes been used for electoral forecasting (e.g., Graefe, Armstrong, Jones, & Cuzán, 2014; Sjöberg, 2009), these studies use quantitative rather than qualitative assessments. Thus far, only a few studies have analyzed electoral forecasting in Japan, including those that do not rely on the media campaign coverage or the electoral polls (e.g., Lewis-Beck & Tien, 2012).

This study proposes a new approach of integrating this qualitative assessment of each district in Japan by applying the item response theory (IRT). Moreover, the study utilizes integrated information for electoral forecasting, which predicts the electoral outcome accurately. This approach not only contributes to electoral forecasting in Japan but also presents a method to integrate qualitative assessments of district-level electoral campaigns, which would improve the accuracy of electoral forecasts in combination with the existing approaches aggregating quantitative information.

2. Poll aggregation

Most social surveys, including electoral polls, conduct surveys only on some members of the population, with a few notable exceptions, such as the national census. Instead of surveying the whole population, they select a few hundred or thousand targets as a sample to estimate the value of interest for the population. In theory, the estimates based on a sample become closer to the true values for the population as the sample size increases. However, differences between estimates and true values remain due to random factors during the sampling process (i.e., sampling error). Moreover, individual polling agencies have systematic, non-sampling errors due to the methodological choices (i.e., house effects), such as question wording or sampling frames (Ford, Wlezien, Pickup, & Jennings, 2017).

Poll aggregation, such as averaging the estimates of multiple polls, decreases these sampling and non-sampling errors. Theoretically, the sampling errors of individual polls would cancel each other out through aggregation. Moreover, the aggregation could alleviate non-sampling errors by taking a middle course of various methods, although it is also possible that the entire industry has a systematic bias, which the aggregation cannot reduce.

The US presidential election has a long history (since the 1930s) of using electoral polls to forecast electoral outcomes (e.g., Gallup, 1951). Since the 1980s, the number of polls has significantly increased because of computers and telephones, making polls cheaper and more efficient (Traugott, 2014). Moreover, since the 2000s, several websites such as Real Clear Politics and FiveThirtyEight websites have been collecting polling data, summarizing the "current state" of the election campaign, and predicting the election's outcome by aggregating the electoral polls (Jackson, 2018).

There are various methods for aggregating polling data on such websites. For example, Real Clear Politics, a political news site and polling data collection site founded in 2000, uses the simple averages of voter intentions in the most recent few polls to predict the electoral outcome (Jackson, 2018). In contrast, Pollster.com (now HuffPost Pollster) has developed a system to use a scatterplot standardization with local weighting (LOESS), showing weighted (moving) averages for the most recent survey (HuffPost Pollster). In addition, FiveThirtyEight website uses a complex forecasting model using polls with basic information such as each state's voter composition (e.g., Silver, 2020).

There are other popular websites that collect polling data outside the USA. For example, in Germany, Wahlumfrage.de presents a simple unweighted average from seven prominent pollsters, and Signal & Rauschen publishes both unweighted and (sample size and survey date) weighted averages of the polls (Graefe, 2019). Moreover, POLITICO Europe tracks polling data for all European elections and countries. POLITICO Poll of Polls uses Kalman filtering (Green, Gerber, & De Boef, 1999) to combine a survey series, which allocates more weight to poll results with a larger sample size and that have been conducted more recently (POLITICO).

In contrast, as discussed above, these types of poll aggregation have not been conducted in Japan due to legal concerns, although a few studies in Japan have discussed survey methodologies, predictions, classification criteria, and the accuracies of the media's electoral coverage (e.g., Fukushima, 2015, 2018; Iida, 2007; Miharu, 2019). A characteristic of election coverage in Japan is that assessments are not expressed using direct poll results, such as "44% of respondents intend to vote for Candidate X", but with qualitative evaluations such as "Candidate X is ahead" or "Candidates Y and Z are in a close race".[3]

The following is an example of a media report:

In District 1, Tsushima is ahead, with Masuda in pursuit. Tsushima has largely consolidated his support among the LDP supporters and is also trying to win over the independent voters. Masuda has consolidated the JIP supporters, but he has not fully captured the DPJ supporters (Aomori District 1, Mainichi Shimbun, December 8, 2014).

These evaluations are comparable to the ordinal scale evaluation used by expert assessments of district-level races in the USA. The Rothenberg Political Report, a widely cited political newsletter in the USA, uses the terms "favored", "lean", or "toss-up" (Rothenberg, 2014, pp. 336,

[3] For example, Jiji Press, one of the major news agencies in Japan, describes their method in an article as follows: *First, we analyze the relationship between the poll support ratings of candidates who ran in past LH [Lower House, i.e., the House of Representatives of the National Diet] elections and the vote shares that they won. Next, we prepare forecasting formulas according to the characteristics of the districts and candidates. The support ratings of each candidate obtained from this poll are assigned to the forecasting formulas. Finally, we conduct simulations for each electoral district to calculate the estimated vote share and probability of winning for each candidate* (October 19, 2017).

now Inside Elections). Similarly, the Cook Political Report, another newsletter, rates the electoral races as "likely", "lean", or "toss-up" (Cook Political Report, 2020).

Nevertheless, there is no definitive order in the expressions used in the Japanese media. It is unclear whether different media use the same terms to describe a similar situation. In the following sections, this study discusses a method to integrate the information from these reports for electoral forecasting, which is applicable beyond Japan to contexts that utilize the qualitative assessments provided in an ordinal scale, such as in the USA.

3. Media election coverage in Japan in the 2014 and 2017 Lower House elections

This section briefly discusses the Japanese political system, the background of the Lower House (LH, the House of Representatives) elections in 2014 and 2017, the official campaign period, and the media election coverage. Japan is a parliamentary democracy with a bicameral legislature consisting of the LH and the Upper House (UH, the House of Councilors). Both houses use a mixed electoral system that combines majoritarian and proportional representation (PR) electoral systems. The LH general elections of 2014 and 2017 used 295 (in 2014) or 289 (in 2017) single-member districts (SMDs), in addition to 11 regional PR districts that elected 180 (in 2014) or 176 (in 2017) representatives.

The 2014 LH election was held on December 14. Prime Minister Shinzo Abe of the conservative, rural-based Liberal Democratic Party (LDP) returned to power in 2012 with his party's long-standing coalition partner, the Buddhist, centrist Clean Government Party (CGP, also known as Komeito). In November 2014, PM Abe dissolved the LH and called for a snap election in order to postpone an increase in the value-added tax rate. The center-left Democratic Party of Japan (DPJ), the main opposition that had been in power from 2009 to 2012, had not fully recovered from their unpopularity and breakup at the end of their time in power. Before the election, the DPJ coordinated the nomination of candidates with another opposition party, the liberal and Osaka-based Japan Innovation Party (JIP). Consequently, the SMD campaigns were mainly fought between the governing coalition candidates (either LDP or CGP, henceforth GC) and candidates from either the DPJ or JIP, although a few other small parties also fielded their candidates despite little chance of victory.[4] The GC achieved another landslide victory in the election and retained a significant majority in the LH.

The next LH election was held on October 22, 2017. PM Abe again led the LDP-CGP coalition. At the same time, opposition parties from the center to the left failed to unite before the election. Instead, they ran divided campaigns as the center-left Party of Hope (POH) and the leftist Constitutional Democratic Party of Japan (CDPJ). The POH and JIP coordinated candidate nominations, while the CDPJ coordinated with the JCP and other left-wing parties. However, most of the SMDs were contested de facto between the GC and one of the opposition camps. The top two candidates of each district achieved 75% or more votes in nine out of the ten districts. Therefore, it can be assumed that at the district level, it was essentially a two-horse race.

In Japan, the formal campaign period begins 12 days before the voting day of the LH election, and candidates can officially run a campaign only during this period. Therefore, the news media usually focus on campaign coverage during this short period. In addition, because of the high cost of conducting electoral polls in all districts, most media conduct only one poll in the early or middle stages of the campaign. Some conduct a second survey, but it often covers only some of the districts.

This study used the election coverage of the GC candidates running in SMDs of seven leading media organizations in Japan during the 2014 and 2017 LH elections (five national newspapers: Yomiuri Shimbun, Nikkei Shimbun, Asahi Shimbun, Mainichi Shimbun, and Sankei Shimbun and two news agencies: Kyodo News and Jiji Press)[5]. The reason for limiting the analysis to the GC candidates is that the fates of the GC and the opposition candidates in each electoral district are mutually determined — they cannot be handled separately. We excluded a few districts without GC candidates from the analysis (three in both 2014 and 2017).

Tables 1a and 1b summarize the survey particulars of the media electoral polls. They used a telephone survey with random digit dialing. The number of respondents varied from around 40,000 for Sankei to 130,000 for Asahi in 2014. Moreover, Yomiuri and Nikkei jointly conducted a telephone survey, although they published their articles independently. Yomiuri and Nikkei conducted a survey again in the later phase of the campaign, but it covered only about half of the districts.

As mentioned above, Japanese media do not report specific figures of the situation in each district, such as "44% of voters plan to vote for Candidate X in district A, 40% for Candidate Y, and 16% are undecided". Instead, their articles use qualitative expressions such as "Candidate X has the advantage" and "Candidates Y and Z are neck and neck".

Roughly speaking, they use expressions such as "dominant", "stable", and "leading" for candidates with an advantage, "catching up" and "struggling" for those with a disadvantage, and "neck and neck" for those in a close race. It is also known that when they use an expression for a close race, the candidate mentioned first has a slight advantage over the one mentioned second (Iida, 2007; Miharu, 2019).

[4] For example, the Japanese Communist Party (JCP), a small left-wing party in Japan, nominated candidates in 292 out of 295 SMDs in 2014 and in 206 out of 289 SMDs in 2017. However, they had little chance of winning; the party's candidates won in only one SMD (Okinawa 1) in both elections thanks to a local alliance with other opposition parties.

[5] These media outlets also predicted each party's achievable PR seats at their respective PR districts in numbers, in addition to their national seat share with the range. In general, their predictions in the PR tier are more accurate than those in the SMD tier due to the fact that they have a much larger size of respondents for each PR district than they do for each SMD district. Moroever, a small survey bias would not change the conclusions because the seat allocation rule is more proportional to their vote share in the tier.

Table 1a
Media poll report for the LH election in 2014 (December 14, 2014).

Election: 2014	Article published	Survey conducted	# of respondents	Lean GC	Toss-Up: Tilt GC	Toss-Up: Tilt Opp.	Lean Opp.
Yomiuri	4 Dec	2–3 Dec	81,381	204	42	27	18
Nikkei	4 Dec			226	31	9	25
Kyodo	5 Dec	2–3 Dec	121,762	228	15	13	35
Mainichi	8 Dec	5–7 Dec	75,258	231	11	12	37
Jiji	8 Dec	NA	NA	205	37	24	25
Sankei	9 Dec	4–7 Dec	40,936	235	14	7	35
Asahi	11 Dec	6–9 Dec	129,467	218	24	18	31

Table 1b
Media poll report for LH election in 2017 (October 22, 2017).

Election: 2017	Article published	Survey conducted	# of respondents	Lean GC	Toss-Up: Tilt GC	Toss-Up: Tilt Opp.	Lean Opp.
Yomiuri	12 Oct	10–11 Oct	78,285	156	64	38	28
Nikkei	12–13 Oct			188	25	31	42
Kyodo	13 Oct	10–11 Oct	90,261	213	17	25	31
Asahi	14–15 Oct	10–13 Oct	88,152,	200	32	26	28
Mainichi	16 Oct	13–15 Oct	73,087	204	27	20	35
Jiji	16 Oct	NA	NA	189	32	36	12
Sankei	17 Oct	12–15 Oct	39,944	209	17	12	48

Note. LH = Lower House; GC = governing coalition; Opp. = opposition. We used newspaper article databases of Yomiuri, Nikkei, Mainichi, Sankei, and Asahi for the information. As for the two news agencies, the results of the Kyodo News survey are taken from articles published in the morning edition of Chunichi Shimbun on December 5, 2014 and October 13, 2017, which are cited in the Chunichi Shimbun database.. For Jiji Press, we used an article provided by G-Search, a database service company, but we failed to find a methodology for Jiji's surveys. Fukushima (2015, 2018) summarizes the electoral polls of these news media, but also does not include a description of Jiji's survey methodology.

Table 2a
The closeness and the accuracy of media forecasts in 2014.

Election: 2014	Total (292)	Within 4% (61)	4%–8% (55)	Over 8% (176)	# of GCs who won (231)
Yomiuri	90%	59%	95%	100%	247
Nikkei	89%	56%	93%	100%	258
Kyodo	89%	61%	91%	98%	244
Asahi	92%	67%	95%	100%	242
Mainichi	94%	77%	95%	100%	242
Jiji	91%	59%	98%	99%	242
Sankei	91%	67%	91%	99%	250

Tables 1a and 1b categorize the media description of the campaign into four categories, Lean GC, Toss-Up/Tilt GC (GC mentioned first), Toss-Up/Tilt Opposition (GC mentioned second), and Lean Opposition.[6] The media seem to use different criteria to describe a particular situation. For example, in 2017, Yomiuri concluded that the 156 districts are Lean GC and 28 are Lean Opposition, while Sankei summarized that 209 are Lean GC and 48 are Lean Opposition. In other words, Yomiuri described that 184 districts favor either the one or the other camp, while Sankei did this for 257 districts.

To what extent were these coverages able to predict the electoral outcomes? Tables 2a and 2b show the accuracy of the media reports according to the closeness of the electoral outcomes. The prediction was considered accurate if the district was described as either Lean GC or Toss-Up/Tilt GC and GC won, or if the district was described as either Lean Opposition or Toss-Up/Tilt Opposition and GC lost.

The media seem to have predicted the district winners well. They succeeded in forecasting the victors in around 90% of the districts. Nevertheless, the high accuracy could be thanks to a landslide victory by the GC in these elections, making the predictions easy in a number of districts. Therefore, we also examined the accuracy depending on the margin of victory, that is, whether the margin of victory between the GC and the opposition candidates with the highest number of votes was less than 4%, 4%–8%, or greater than 8%.[7]

It seems less easy to predict the outcome when the result was close. Among the districts where the margin of victory was less than 4%, the media's prediction was correct in 56%–77% of the cases in 2014 and 62%–75% in 2017. In contrast, when the margin was 4%–8%, they were accurate around 90% of the time and almost perfect when the margin was over 8%. The average number of respondents in each district would have been around 140 to 440, given the sample size of the surveys. Therefore, there is a possibility that the sampling error alone would

[6] See Table A.1 in the Appendix for wording in the articles and their classifications.

[7] In this study, the margin of victory is defined as the difference in the vote shares between the GC and the opposition candidates with the highest number of votes, divided by the sum of their vote share. Hence, if a GC candidate won 41% of the votes, POH won 35%, and CDPJ won 24%, the GC's margin of victory would be $(0.41-0.35)/(0.41+0.35) = 7.9\%$.

Table 2b
The closeness and the accuracy of media forecasts in 2017.

Election: 2017	Total (286)	Within 4% (71)	4%–8% (63)	Over 8% (152)	# of GCs who won (223)
Yomiuri	91%	70%	94%	100%	222
Nikkei	91%	73%	92%	99%	214
Kyodo	87%	62%	87%	99%	230
Asahi	90%	68%	90%	100%	232
Mainichi	91%	73%	92%	99%	231
Jiji	93%	75%	98%	99%	219
Sankei	90%	69%	90%	100%	227

Note: GC = governing coalition.

lead to an incorrect assessment of which candidate is ahead in a close electoral race. Moreover, there should be a non-sampling error due to, for example, refusal/non-response, sampling frame, and a bias in estimating who would or would not cast their votes.[8]

In addition, some voters appeared to decide or change their voting intentions in the final phase of the campaign. In 2014, all companies significantly overestimated the number of GC seats. It seems that there was a national swing against the GC in the final week of the campaign, which made it difficult to predict the outcome, especially for close races.

4. Poll aggregation: method and results

This section discusses the method that was used to aggregate the district assessments of the Japanese media. As described above, the Japanese media qualitatively evaluate the race in each district. Therefore, we could not use a (weighted) average of the electoral polls, as in other countries. Moreover, as Table 1a and 1b suggest, the news media use heterogeneous criteria and expressions for their classification. Hence, it would be inefficient to assign numerical values to their expressions; that is, for example, 1 for Lean GC and 2 for Toss-Up/Tilt GC, and then calculate the (weighted) average of their expressions to forecast the electoral outcomes.

To handle this available information appropriately, this study applied the IRT. The IRT is a latent variable model in which a small number of latent variables are assumed to explain the relationship between the observed responses. Concepts that are challenging to measure directly with conventional methods, such as political attitudes, can be measured by assuming these latent variables.

The IRT was initially developed in psychology and education to measure ability and proficiency through examinations (Paek & Cole, 2019, p. 2). In recent years, this method has also been used in the political science field, for example, to measure political knowledge (Delli Carpini & Keeter, 1993), the policy positions of individual legislators (Clinton, 2006; Hirano, Imai, Shiraito, & Taniguchi, 2011; Shor & McCarty, 2011), the policy positions of districts and municipalities (Tausanovitch & Warshaw, 2013), and the political positions of the US Supreme Court judges (Martin & Quinn, 2002). The method is also applied to aggregate expert judgments in the Varieties of Democracy (V-Dem) Research Project (e.g., Pemstein, Marquardt, Tzelgov, Wang, Medzihorsky, Krusell, Miri, & Römer, 2020).

In this study, we used the graded response model of the IRT, which can employ three or more levels of graded response. However, we start with the binomial, two-parameter model. The binominal model is used when the response is binary, such as the answer to a test question: correct answer (1) or wrong answer (0). The response is a function of the parameter θ_i (e.g., the ability of examinee i), the difficulty parameter β_j of item j, and the discriminant parameter α_j, which are not directly observed.

$$\mathrm{p}(X_{i,j}=1|\theta_i,\beta_j,\alpha_j)=\frac{\exp\left[\alpha_j\left(\theta_i-\beta_j\right)\right]}{1+\exp\left[\alpha_j\left(\theta_i-\beta_j\right)\right]}$$
$$s.t.\ i=1,2,3,\ldots,n; j=1,2,3,\ldots,m$$

Due to the difficulty parameter of each item, the examinee with the same ability answers individual items with varying accuracy rates. In addition, due to differences in the discrimination parameters, the relationship between ability and accuracy rate changes, even for questions with the same difficulty level. For example, an examinee with a high ability will almost certainly get the correct answer for an item with a high discriminant parameter, while an examinee with a low ability will likely answer the item incorrectly. On the other hand, the relationship between ability and accuracy rate becomes less clear if the discriminant parameter is low.

The graded response model extends the binomial model for situations in which the responses are given on an ordinal scale of three or more levels. Suppose that the situation θ_i in district i is not directly observable to the researcher, and the news media j classifies the district into one of four categories, such as Lean GC (4), Toss-Up/Tilt GC (3), Toss-Up/Tilt Opposition (2), or Lean Opposition (1), according to its survey and classification criteria (with a certain degree of error).

In this case, $p_{ij1}(\theta)$, $p_{ij2}(\theta)$, $p_{ij3}(\theta)$, and $p_{ij4}(\theta)$ are the probabilities of the districts being classified into the respective categories given θ_i. On the other hand, we estimated situation θ_i of each district and the differences in each media outlet criteria (α_j and β_j) from their reports. Hence, the formula is:

$$p^*_{ijk}(\theta)=\frac{\exp\left[\alpha_j\left(\theta_i-\beta_{jk}\right)\right]}{1+\exp\left[\alpha_j\left(\theta_i-\beta_{jk}\right)\right]}$$
$$s.t.\ i=1,2,3,\ldots,n; j=1,2,3,\ldots,m; k=1,2,3,4$$

[8] For example, the turnout in the 2014 election was a record low (52.7%), and 6.6% points lower than the 2012 election (59.3%), which was the lowest at that time.

$$p_{ij1}(\theta) = 1 - p^*_{ij2}(\theta)$$
$$p_{ij2}(\theta) = p^*_{i2}(\theta) - p^*_{ij3}(\theta)$$
$$p_{ij3}(\theta) = p^*_{i3}(\theta) - p^*_{ij4}(\theta)$$
$$p_{ij4}(\theta) = p^*_{i4}(\theta)$$

There are some advantages to using the graded response model of the IRT. First, we can estimate situation θ_i in each district on an interval scale by integrating the media outlets' qualitative assessments after considering their criteria. In the graded response model, the discrimination and difficulty parameters represent the differences in their criteria. This study used the "mirt" package in R to estimate these parameters (Chalmers, 2020).

After applying the method to the data, we made the following assumptions[9]. First, we assumed that media companies used the same criteria to classify SMD campaign situations for both 2014 and 2017 LH elections. Hence, they would use the same expression (or, more precisely, will use one of the expressions with the same probability) to report whether the situations θ_i are the same in 2014 and 2017.

Next, we did not weigh the media reports according to the sample size or survey date. In theory, a report based on a larger sample size should be more reliable. However, Tables 2a and 2b suggest that surveys with larger sample sizes and/or closer to the voting day are not necessarily more accurate, at least in these elections. This could be due to a coincidence in the districts with small margins. This could also be due to non-sampling errors because these news media use different survey methods. Moreover, all surveys were conducted within no more than one week because the media intensively conduct electoral surveys during the short official campaign period (12 days). Therefore, the weight based on the proximity to voting day was expected not to improve the result by much.

Based on these assumptions, we first ran the graded-response IRT model using the media assessments for the 2014 LH election data. Next, we estimated situation θ_i in each district with the parameters and the media's assessments in 2014 and 2017. Table 3 shows the difficulty parameter β_{jk} and the discriminant parameter α_j of each company, estimated with the graded-response IRT model. The difficulty parameter β_{j1} is the value at which media outlet j is more likely to assess the district as Toss-Up/Tilt Opposition rather than Lean Opposition if the situation θ_i is larger than this. If the situation θ_i is larger than β_{j2}, the media assesses the district as Toss-Up/Tilt GC rather than Toss-Up/Tilt Opposition. If it is beyond β_{j3}, the media outlet is more likely to evaluate the district as Lean GC rather than Toss-Up/Tilt GC.

[9] As mentioned above, recently Yomiuri and Nikkei write their articles separately but share the survey errors since they conduct the survey jointly. To incorporate their common error in the model, we used a two-dimensional IRT model, which assumed a common factor η_i for the two companies: η_i affects only the judgments of Yomiuri and Nikkei, while θ_i affects the judgment of all companies. However, the correlation of θ_i estimated with one- and two-dimensional models are almost perfect at .998. On the other hand, the standard errors of the parameters for the two companies become very large in a two-dimensional model. Therefore, this study used the one-dimensional model of the IRT, although there is some violation of the local independence assumption.

Table 3
The discrimination and difficulty parameters of media companies.

Parameters	α_j	β_{j1}	β_{j2}	β_{j3}
Yomiuri	8.123	−1.590	−1.027	−0.509
Nikkei	9.303	−1.409	−1.233	−0.739
Kyodo	6.603	−1.428	−0.969	−0.522
Asahi	4.105	−1.307	−1.094	−0.852
Mainichi	4.044	−1.339	−1.015	−0.694
Jiji	4.480	−1.220	−1.020	−0.860
Sankei	4.310	−1.280	−1.150	−0.910

These parameters quantify the differences in the evaluation criteria between the news media. For example, the difficulty parameters imply that Yomiuri evaluates more districts as toss-ups than the other companies do; the company evaluates a district as Lean Opposition rather than Toss-Up/Tilt Opposition when θ_i is smaller than −1.590, while Sankei evaluates it as Lean Opposition when θ_i is smaller than −1.280. Similarly, Yomiuri is more likely to evaluate a district as Lean GC than Toss-Up/Tilt GC when θ_i is greater than −0.509, but Sankei evaluates it as Lean GC when θ_i is greater than −0.910. The interpretation of the discriminant parameter α_j is more complicated than that of the difficulty parameter. It is possible to interpret that a news medium with a large α_j, such as Yomiuri, tends to synchronize with the other companies.

So, to what extent did the estimated situation θ_i correlate with the actual electoral results? Fig. 1 shows θ_i on the X-axis and the margin of victory of the GC candidate to the opposition party candidate with the largest number of votes on the Y-axis for the 2014 and 2017 LH elections. Many observations have a value of 0.643 or −2.147 in the X-axis. These are the districts that have been reported by all the media as Lean GC or Lean Opposition.

Fig. 1 shows that the estimated situation θ_i is strongly correlated with the margin of victory not only in 2014 but also in 2017 ($r = 0.806$ in 2014 and 0.855 in 2017). Nevertheless, the correlations might not be strong enough to forecast the winner in a closely contested race.

On the other hand, θ_i estimated with the IRT method is a relative indicator. With it, we can infer the relative strength of a governing party candidate in a district as compared to those in other districts. However, we cannot determine precisely when we should start forecasting the victory of the GC rather than their defeat in the districts. We know that every survey has a systematic bias due to the coverage or non-response error, but we do not know which one is correct without seeing the actual values. In other words, we need a zero-point adjustment in order to apply the indicator for the purpose of forecasting. In the next section, we discuss a method for forecasting the electoral outcome with θ_i.

5. Forecasting the electoral outcome: method and results

This section describes the method used to forecast the electoral outcome with the integrated information in the previous section. To be more specific, we first used the relationship between the media reports (and the estimated

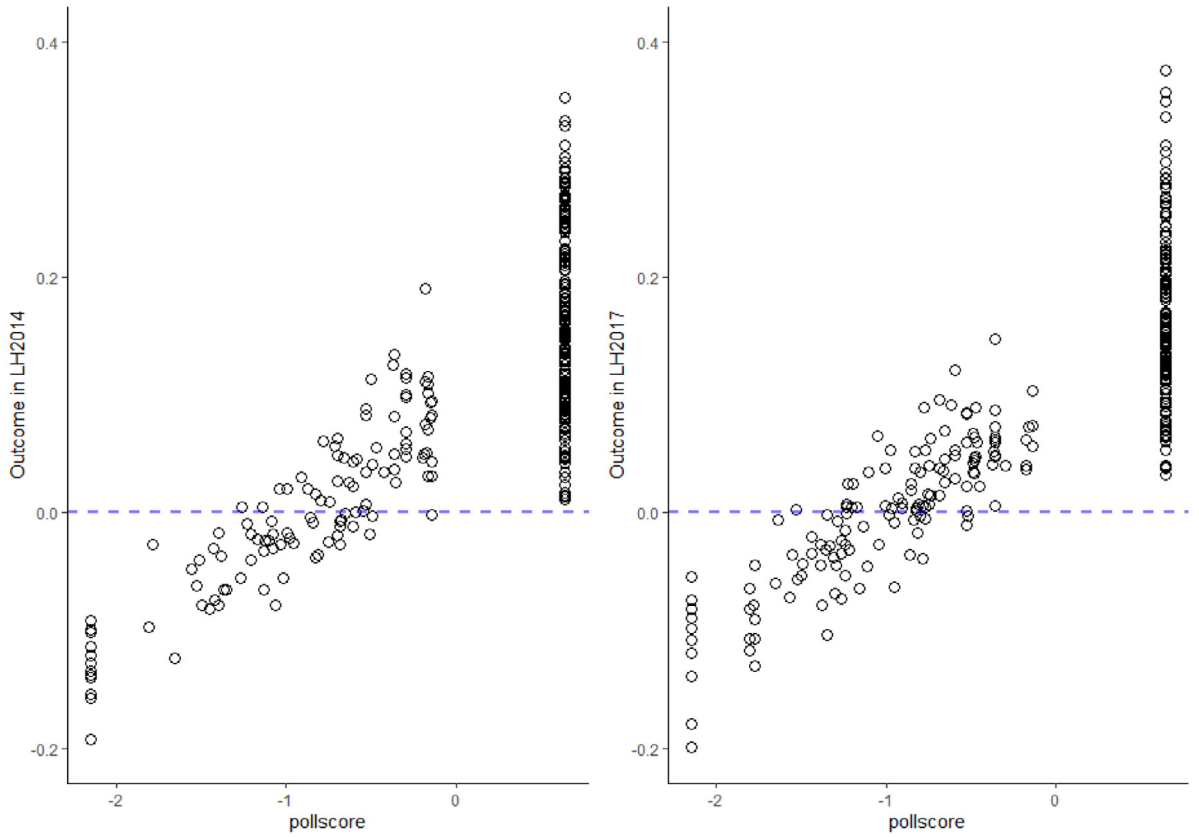

Fig. 1. Estimated situation (X-axes) vs. actual election results for the 2014 and the 2017 LH election (Y-axes).

situation θ_i based on the report) and the electoral results in the 2014 LH election to "train" the model. Next, we tested the model by the following: (1) forecasting the winners of individual electoral districts, (2) the margin of victory for the ruling and opposition candidates, and finally, (3) the overall number of electoral districts that the GC won in the 2017 LH elections.

First, as discussed in the previous section, we ran the IRT model with the media assessments in 2014 and then estimated situation θ_i in each district with the parameters and the media's assessments in both 2014 and 2017. However, θ_i is the relative indicator of the GC's popularity in a district. Hence, it does not indicate the cutoff where we can start forecasting their victory rather than their defeat.

As a convenience measure, we predicted the winners of individual districts by using the mean of the difficulty parameter (MDP) β_{j2} (-1.073, MDP as follows) of each media outlet as the threshold: if θ_i is larger than MDP, we predict that the GC candidate wins but otherwise loses.[10]

Table 4 shows that this simple method correctly forecasts the winner of 93% (266 out of 286) seats in the 2017 LH election. The method correctly forecasts 73% of the districts with a margin of less than 4%, 98% of districts with a 4%–8% margin, and all of those with a margin of 8% or more. The forecast is almost as accurate as that of Jiji Press, which turned out to be the best predictor of the district winners in 2017 among seven media outlets.

Next, we forecast the margin of victory in each electoral district with the following method. First, we analyzed the relationship between situation θ_i and the margin of victory Y_{i14} of the GC and the opposition candidates (with the largest vote among them) in 2014 with an OLS regression analysis. Hereafter, we refer to this regression analysis as the 2014 regression model. Using this model, we forecast the margin in 2017, $\hat{Y}_{i17}$, based on situation θ_i estimated from the media coverage in 2017.

In this process, it is necessary to consider the media's systematic overestimation of the GC victory in 2014. Table 2a shows that the news media predicted that the GC would win 242 to 258 districts, whereas they won 231. Thus, θ_i based on these evaluations also overestimated their chances of victory without proper adjustment. If we

[10] We admit that the method is relatively arbitrary and not theoretically rigorous. Another possible solution for the cutoff point is to regress the electoral outcome to θ_i, and then calculate the value of θ_i that, on average, yields a tie between the GC and the opposition candidates to use as the cutoff point. The value is -0.889 in the model that used the 2014 data. If there was no systematic bias among the media outlets (or national swing after the surveys) in that election, the value could have been used for this purpose. However, there was a non-ignorable bias among media reports in 2014. The value is smaller than difficulty parameters β_{j2} of any media outlets in 2014, ranging from -0.969 (Kyodo) and -1.233 (Nikkei). Therefore, we use MDP instead as a surrogate for the cutoff for binary predictions.

Table 4
The closeness and the accuracy of combined media forecasts in 2017.

Election: 2017	Total (286)	Within 4% (71)	4%–8% (63)	Over 8% (152)	# of GCs' win (223)
Combined	93%	73%	98%	100%	227

Note: of GCs who won win is based on binary forecast: the number of districts where the estimated situation θ_i is larger than the MDP, the mean of the difficulty parameters β_{j2} of five media outlets (-1.073).

assumed that the exact relationship holds between the media assessments and the electoral outcomes in 2017 as it did in 2014, we would, in turn, significantly underestimate the margin of victory of the GC candidates in 2017 (unless the media had the same systematic bias that they had in 2014).

We modified the forecasting model to remove the bias so that the GC and the opposition candidates are tied (i.e., the margin of victory is zero) when θ_i is the MDP. With this adjustment, the revised forecasting model of the margin of victory is $\hat{Y}_{i17} \sim 0.1115 + 0.1040\theta_i$.[11]

We applied the same approach to forecast the winning probability of the GC candidate in each district. We used probit analysis to analyze the relationship between situation θ_i and the victory or defeat of the GC candidates in 2014 (the 2014 probit model). Using the coefficients of the probit model, we forecast the winning probability of the GC candidates in 2017 based on θ_i estimated from the media coverage before the election. We also had to adjust for the bias in 2014 so that the winning chance of the GC candidate is 0.50 when θ_i is the MDP.[12] The revised forecasting model is $\hat{p}(W_{i17} = 1) \sim F(3.298 + 3.074\theta_i)$. The winning probability of the GC candidates as estimated by the probit model, in turn, can be summed up to forecast the total number of districts won by the GC. This probability method is better than the binary model which was discussed earlier because it incorporates prediction uncertainty.

How well were these models able to forecast the electoral results? Fig. 2 shows the relationship between the forecasts and the electoral results for the 2017 LH election, using the forecasting model described above. The figure also shows the Y=X line, which is the theoretical relationship between the two, and the probit curve, which shows the winning probability corresponding to the forecasted margin of victory. The *Y*-axis scale on the left side shows a scale for the margin of victory, while that on the right side indicates a scale for the winning probability.

This figure shows that the model forecast the 2017 electoral results well. If the results are regressed to the forecasts, the intercept is -0.012, and the coefficient is 0.981, which is very close to the respective theoretical values, zero and one. There are no statistically significant differences between the theoretical and the actual relationships. Regarding the forecast accuracy, the root mean square error (RMSE) is 5.6%, and the mean absolute error (MAE) is 4.3%. Moreover, RMSE and MAE decrease to 3.2% and 2.6%, respectively, when the districts at both ends of θ_i are excluded. These are the districts that all media assessed as leaning to either camp and that are likely to be affected by the floor and ceiling effects.

Moreover, the sum of the GC's winning probability for all districts was 225 seats, which is quite close to the total number of districts that the GC actually won in 2017 (223 seats). We ran 10,000 simulations with the forecasted winning probability in each district to show the expected range of the GC seats at the national level. The results suggest that the GC seats would be roughly between 219 and 232 (5th and 95th percentiles). However, the forecast assumed no systematic bias in the electoral polls of the media nor a national swing after their polls. As in the 2014 LH election, a national swing might have happened after the polls, or the media polls could have had systematic errors. Consequently, the chances that the result would fall outside this range might be more significant than those suggested by the simulations.

6. Discussion and conclusion

This paper presents a method for aggregating qualitative media campaign assessments by using the Japanese LH elections as a case study. Further, we used this aggregated information to forecast the electoral results at the district and national levels. More specifically, we integrated the media coverage in the Japanese (2014 and 2017) LH elections by using the graded model of the IRT. In addition, we applied the relationship between the estimated situations in each district and the electoral results in 2014 to create a forecasting model of the margin of victory and winning probability of the GC candidates in 2017.

The results show that the models can predict the electoral outcome with a high level of accuracy from the media coverage available a few days before the voting day. At the national level, our model forecast that the GC candidates would win in a total of 225 districts, quite close to the 223 that they actually won. At the individual district level, the model correctly predicted the winner in more than 99% of the districts where the final margin of victory was 4% or more, and in approximately three-quarters of the toss-up districts where the margin was

[11] If we assign value of MDP (-1.073) to θ_i, the right-hand side of the 2014 regression model equation $0.0924 + 0.1040\theta_i$ becomes 0.0191. This number indicates when the race should be tied in theory; in practice, the GC on average lost by a 1.91 percentage point margin. Therefore, the forecasting model needed an adjustment for 0.0191 such that the right-hand side is zero if the MDP is assigned to θ_i. The revised model for the 2017 LH election is now $\sim 0.1115 + 0.1040\theta_i$, where $0.1115 = 0.0924 + 0.0191$.

[12] In the 2014 probit model, the relationship between the GC candidate's win (1) or loss (0), W_{i14}, and θ_i was estimated as $\sim F(2.450 + 3.074\theta_i)$. If we assign the MDP for θ_i in this model, the right-hand side of the equation should be $F(0)$, such that $= F(0) = 0.5$: the governing and the opposition party candidates have an equal chance to get elected. However, in practice, it was $F(-0.848)$. To remove the bias of -0.848 in the original model, we used a forecasting model with an intercept of $3.298 (2.450 + 0.848)$. Hence, the forecasting model of the winning probability of GC is $\sim F(3.298 + 3.074\theta_i)$.

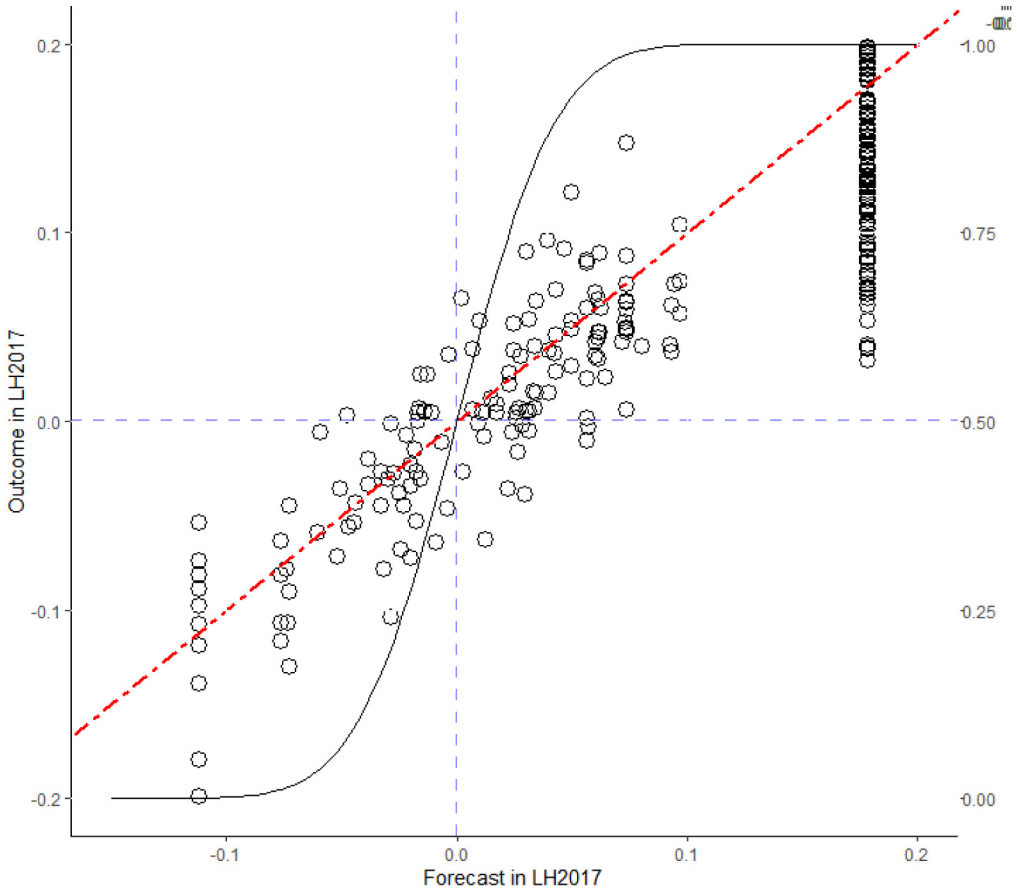

Fig. 2. Relationship between the forecasts and the electoral results for the 2017 LH election.
Note. LH = Lower House. The figure does not include the districts where the margin of victory of governing coalition candidates was more than 0.20, for ease of viewing the results. There were 42 such districts in the 2017 LH election.

less than 4%. This study also quantified the classification criteria of each news medium as the parameters of the IRT model.[13] The forecast would be of interest not only to the media but also to political parties that are planning their campaign resource allocation, and interested voters who seek a summary of the media reports in a single indicator.

The method performs as precisely as the best predictor among the media outlets, which was Jiji Press in the year 2017. However, it does not mean that the method is not practical. If we know which media outlet will have the best forecast before the election, we can use its report. Unfortunately, it is difficult to know this in advance. The method used in this study allows us to make predictions as accurate as the one provided by the media outlet, which turned out to be the best after the election. Nevertheless, the accuracy of the electoral forecast with our method still depends on the quality of the news media reports, which are not available until less than a week before the voting day (in Japan at least).

This paper proposes a new way to forecast the election results by the information provided on an ordinal scale to forecast electoral outcomes by taking Japan as its case study using the Japanese media's qualitative assessments during the election campaign period. However, the method is not limited to the Japanese context. The method proposed in this paper is applicable to aggregating these evaluations, even when they use non-standardized classification criteria, such as the Japanese media do. By enabling the integration and application of qualitative assessments, this approach can improve forecast accuracy when it is combined with existing approaches.

[13] The method is robust to different electoral settings. For example, Japan had another LH election in 2021 and we examined the 2017 and 2021 LH election and media evaluation data to forecast the electoral results in 2021. We used basically the same approach with a few adjustments in response to the change in the media's survey method. The correlations between the electoral outcome and θ_i are stronger in 2021 (r = .877). However, in 2021, the media on average underestimated the GC's popularity, and consequently the model also underestimated their vote share and chance of winning. For example, the sum of estimated winning probabilities for all districts was 184, and the binary model predicted the GC would win in 186 districts, while the governing parties won in 198 single-member districts. The binary model nevertheless performs better in forecasting district-level winners than any individual media reports (89%, 255 out of 286, while Mainichi, which made the best district-level forecasts in 2021, correctly forecasted the winner in 253 districts).

Table A.1
Classification and expressions in news articles.

Classification	Expressions in news articles (Japanese)
Lean GC	*stable (antei), superior (yui), dominating (yusei, dokuso), solid (tegatai, banjyaku), overwhelming (attou), outdistancing (tsukihanasu), widely supported (habahiroi shiji), dominant (yuryoku), leading (senkou), outpacing (nukedashita)*, and/or any disadvantageous expressions for the opponents.
Toss-Up	*close match (sessen), rivalry (hakuchu), melee (konsen), fierce contest (hageshiku arasou, seriau), neck and neck (yokonarabi), evenly matched (gokaku),*and *fight to the death (shito)*
Lean Oppisition	*hard (kenmei), in pursuit/chasing (ou), chasing hard (moutsui), closing in (nikuhaku), catching up (oiage), trying to increase support (shiji kakudai), lagging (deokure), sluggish (nobinayamu), regaining ground (makikaeshi),* and/or any advantageous expressions for the opponents.

Note. GC = governing coalition. Some words, such as "stable" and "dominating," are expected to be stronger expressions to indicate the degree of the advantage than "leading." However, these differences would hardly influence the forecast of victory. Therefore, we classified all these expressions into the same category, Lean GC, in this study.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix

As for the expressions in Table A.1. Toss-Up is reclassified into "Toss-Up: Tilt GC" and "Toss-Up: Tilt Opposition", based on which candidates were mentioned earlier in the article's description. Suppose that candidate X runs from GC and candidate Y belongs to the opposition. If a media outlet reported that "candidate X and candidate Y are in a close match", the district is categorized as "Toss-Up: Tilt GC", whereas if it is reported as "candidate Y and candidate X are neck and neck", the district is coded as "Toss-Up: Tilt Opposition". In general, there are statistically significant differences in the vote-share margin in the electoral outcome between the districts described earlier and later expressions.

References

Chalmers, P. (2020). Package mirt. Retrieved from: https://cran.r-project.org/web/packages/mirt/mirt.pdf. (Accessed 11 March 2021).

Clinton, J. D. (2006). Representation in Congress: Constituents and roll calls in the 106th House. *Journal of Politics*, *68*(2), 397–409. http://dx.doi.org/10.1111/j.1468-2508.2006.00415.x.

Cook Political Report (2020). House race ratings. Retrieved from: https://cookpolitical.com/ratings/house-race-ratings. November 2, 2020. (Accessed 3 July 2021).

Delli Carpini, M. X. D., & Keeter, S. (1993). Measuring political knowledge: Putting first things first. *American Journal of Political Science*, *37*(4), 1179–1206. http://dx.doi.org/10.2307/2111549.

Ford, R., Wlezien, C., Pickup, M., & Jennings, W. (2017). Poll and votes. In K. Arzheimer, J. Evans, & M. S. Lewis-Beck (Eds.), *The SAGE handbook of electoral behaviour* (pp. 787–812). London: Sage Publication.

Fukushima, Y. (2015). The 47th LH election: how newspapers and TV reported (Dai 47 kai syugiingiin senkyo: shimbun terebi wa dou tsutaetaka). *Yoron*, *116*, 7–17.

Fukushima, Y. (2018). The 48th LH election: How newspapers and TV reported (Dai 48 kai Syugiingiin Senkyo: Shimbun Terebi wa dou tsutaetaka). *Yoron*, *121*, 68–81.

Gallup, G. (1951). The Gallup poll and the 1950 election. *Public Opinion Quarterly*, *15*(1), 16–22. http://dx.doi.org/10.1086/266275.

Graefe, A. (2019). Accuracy of German federal election forecasts, 2013 and 2017. *International Journal of Forecasting*, *35*(3), 868–877. http://dx.doi.org/10.1016/j.ijforecast.2019.01.004.

Graefe, A., Armstrong, J. S., Jones, R. J., & Cuzán, A. G. (2014). Combining forecasts: An application to elections. *International Journal of Forecasting*, *30*(1), 43–54. http://dx.doi.org/10.1016/j.ijforecast.2013.02.005.

Green, D. P., Gerber, A. S., & De Boef, S. L. D. (1999). Tracking opinion over time: A method for reducing sampling error. *Public Opinion Quarterly*, *63*(2), 178–192. http://dx.doi.org/10.1086/297710.

Hirano, S., Imai, K., Shiraito, Y., & Taniguchi, M. (2011). Policy positions in mixed member electoral systems: Evidence from Japan (unpublished manuscript). Retrieved from: https://imai.fas.harvard.edu/research/japan.html. (Accessed 18 July 2018).

HuffPost Pollster (n.d.), The Huffington Post pollster policy for including polls in charts. Retrieved from: https://elections.huffingtonpost.com/pollster/faq. (Accessed 20 April 2021).

Iida, Y. (2007). The analysis of the media electoral coverage—the 44th general election as a case (shimbun no senkyo jyousei hodo no bunseki—dai 44 kai sousenkyo wo jirei to shite). *Jissen Jyoshidai Ningensyakai Gakubu Kiyo*, *3*, 19–42.

Jackson, N. (2018). The rise of poll aggregation and election forecasting. In L. R. Atkeson, & R. M. Alvarez (Eds.), *The Oxford handbook of polling and survey methods* (pp. 609–632). Oxford: Oxford University Press.

Lewis-Beck, M. S., & Stegmaier, M. (2014). US presidential election forecasting. *PS: Political Science and Politics*, *47*(2), 284–288. http://dx.doi.org/10.1017/s104909651400002.x.

Lewis-Beck, M. S., & Tien, C. (2012). Japanese election forecasting: Classic tests of a hard case. *International Journal of Forecasting*, *28*(4), 797–803. http://dx.doi.org/10.1016/j.ijforecast.2012.04.005.

Martin, A. D., & Quinn, K. M. (2002). Dynamic ideal point estimation via Markov chain Monte Carlo for the US supreme court 1953–1999. *Political Analysis*, *10*(2), 134–153. http://dx.doi.org/10.1093/pan/10.2.134.

Miharu, M. (2019). *Opinion poll as a weapon (Buki to shiteno yoronchosa)*. Tokyo: Chikuma Shobo.

Paek, I., & Cole, K. (2019). *Using R for item response theory model applications*. Taylor & Francis.

Pemstein, D., Marquardt, K. L., Tzelgov, E., Wang, Y., Medzihorsky, J., Krusell, J., et al. (2020). *The V-Dem measurement model: Latent variable analysis for cross-national and cross-temporal expert-coded data*: *Working Papers 21, 5th edition*, Göteborg: University of Gothenburg, Varieties of Democracy Institute, Retrieved from: https://www.v-dem.net/static/website/files/wp/wp_21_5th.pdf. (Accessed 5 February 2022).

POLITICO (n.d.). Poll of polls. Retrieved from: https://www.politico.eu/europe-poll-of-polls. (Accessed 27 June 2021).

Predicting Winner (2021). Predicting the winners and the losers of all 289 electoral districts in the lower house election. *Syukan Asahi, 126(41)*.

Rothenberg, S. (2014). Election forecasting and public opinion polls. *PS: Political Science and Politics*, *47*(2), 336–338. http://dx.doi.org/10.1017/s1049096514000158.

Shor, B., & McCarty, N. (2011). The ideological mapping of American legislatures. *American Political Science Review*, *105*(3), 530–551. http://dx.doi.org/10.1017/s0003055411000153.

Silver, N. (2020). Our new polling averages show Biden leads Trump by 9 points nationally. Retrieved from: https://fivethirtyeight.com/features/our-new-polling-averages-show-biden-leads-trump-by-9-points-nationally. (Accessed 20 April 2021).

Sjöberg, L. (2009). Are all crowds equally wise? A comparison of political election forecasts by experts and the public. *Journal of Forecasting*, *28*(1), 1–18. http://dx.doi.org/10.1002/for.1083.

Tausanovitch, C., & Warshaw, C. (2013). Measuring constituent policy preferences in Congress, state legislatures, and cities. *Journal of Politics*, *75*(2), 330–342. http://dx.doi.org/10.1017/s0022381613000042.

Tokyo High Court (1991). Judgment, February 8. Retrieved from: https://www.courts.go.jp/app/files/hanrei_jp/648/016648_hanrei.pdf. (Accessed 14 July 2021).

Traugott, M. W. (2014). Public opinion polls and election forecasting. *PS: Political Science and Politics*, *47*(2), 342–344. http://dx.doi.org/10.1017/s1049096514000171.

Yasuda, M., & Arakawa, A. (2009). *Article-by-article commentary of the Electoral Law (Chuikujo kaisetsu koushoku senkyo ho)*. Tokyo: Gyosei.