Acta Politica (2023) 58:461–482
https://doi.org/10.1057/s41269-022-00250-x

RESEARCH NOTE

The predictive value of polls in a fragmented multi-party system: the Netherlands (1998–2021)

Tom W. G. van der Meer[1] · **Lisa Janssen**[1] · **Tom Louwerse**[2]

Accepted: 23 May 2022 / Published online: 23 June 2022

Abstract

Although vote intention polls are often used in the public debate as forecasts of future election outcomes, their predictive value has been subjected to scholarly inquiry. This research note contributes to the literature by assessing the predictive value of vote intention polls simultaneously at the macro-level (polls), meso-level (parties), and micro-level (voters). We analyse polls presented by the main polling agencies in the Netherlands (covering seven election cycles between 1998 and 2021), as well as micro-level panel data (covering 27,572 respondents and 46 polls between 2006 and 2010, and 35,574 respondents and 31 polls between 2010 and 2012). We reach three main conclusions. First, vote intention polls in the Netherlands generally do not provide more information than the previous election outcome, until the last few weeks of an election cycle. Second, the predictive accuracy of vote intention polls is lower for challenger parties than for non-challenger parties, particularly midway through the election cycle. Third, the predictive value for individual voters is generally very low, until the last few months before the election.

Keywords Opinion polls · Elections · Voters · Forecast · Volatility · Party preference

Introduction

The debate on the extent to which vote intention polls can and should be used to forecast election outcomes goes back to at least the 1940s (Seymour et al. 1949). It flared up in 1948 with the notoriously incorrect headline prediction that Thomas Dewey (R) would beat Harry Truman (D) in the presidential election. More recently, the predicted outcomes of the Brexit vote in the United Kingdom, and the presidential election victory of Donald Trump (R) in the United States have sparked similar

✉ Tom W. G. van der Meer
t.w.g.vandermeer@uva.nl

[1] Department of Political Science, University of Amsterdam, REC-B 10.09, Nieuwe Achtergracht 166, 1018 WV Amsterdam, The Netherlands

[2] Department of Political Science, Leiden University, Leiden, The Netherlands

debates. Yet, the debate is not merely concerned with forecasts made on the eve of election day. Rather, vote intention polls are presented and used as forecasts of future election outcomes by voters, media, and politicians, even months before the elections (Irwin and Van Holsteyn 2002; Brettschneider 1997; Toff 2019; Jaworski and Fitzgerald 2008).

Recent scholarly research has delved into the predictive value of vote intention polls (e.g. Caballe et al. 2013; Jennings and Wlezien 2016, 2018; Goot 2021). In their encompassing study of elections across the globe, Jennings and Wlezien (2016, p. 231) conclude: "very early polls predict the vote, at least to some extent. This largely reflects differences in the equilibrium support of parties and candidates. Polls do become increasingly informative over time, however, pointing to real evolution of preferences". Their analysis of vote intention polls also shows that the overall predictive power of polls has *not* deteriorated over the last few decades.

This research note provides an empirical contribution to the comparative literature. Most international-comparative analyses of the predictive power of vote intention polls focus on the *overall* net poll outcomes as the unit of analysis, either dichotomously (as predictions of an electoral win) or as vote shares (as predictions of the distribution). Our analysis of one country allows us to look beyond the predictive value of polls at the macro-level, by examining the meso-level (parties and pollsters) and the micro-level (voters) as well. This is essential to understand the extent to which a lack of predictive accuracy finds its roots in differences between (types of) parties, (types of) pollsters, and (types of) voters.

At the macro-level, we assess the predictive accuracy of all regular nationwide polls in a radically proportional and fragmented country, the Netherlands. At the meso-level, we assess the predictive accuracy of these polls for specific (types of) parties and analyse whether some pollsters pick up on trends in party support before others. Ultimately, the net outcomes at the macro-level are based on preferences at the micro-levels: the extent to which voters' intended vote is identical to the vote that they actually cast in the following election. Using extensive panel data on two election cycles (2006–2010 and 2010–2012) this paper assesses the share of respondents whose vote intention is identical (with or without interruption) to their actual future vote, and explains variation in the timing of this alignment.

Our analysis focuses on the vote intention polls of five main polling agencies in the Netherlands between 1998 and 2021. The Netherlands has a highly proportional electoral system and an increasingly fragmented multiparty system. Although Jennings and Wlezien (2016) find no evidence that fragmented multiparty systems are different in the predictive power of polls (in the last 200 days before the elections), the level of fragmentation in the Netherlands (where the effective number of parties[1] grew from 5 in 1998 to 8 in 2021) is much larger than the cut-off point in their analyses (3.5). The high level of fragmentation in the Netherlands has implications for the volatility and predictability of electoral preferences: "In system where there are fewer parties, voters face less choice, and this may make preferences more stable

[1] The effective number of parties is essentially an inverted Hirschmann–Herfindahl index of concentration, i.e., 1 divided by the sum of each party's squared fraction of parliamentary seats.

and also more predictive of the final vote" (Jennings and Wlezien 2016, p 222). Consequently, polls are expected to have less predictive power in the Netherlands until the final stages of the electoral cycle (Jennings and Wlezien 2016). Previous work has examined the predictive value and the quality of vote intention polls in the Netherlands mainly in the context of a specific election, such as the 'polling discrepancy' in the 1986 elections (Andeweg 1986) and the very volatile polls during the 2012 election campaign (De Lange 2013). Moreover, these analyses focus mostly on the differences between the final polls before an election and the outcome. Our analysis adds to this literature by focusing on the predictive value of opinion polls over the entire electoral cycle, and by analysing data for a long time period, including the most recent, highly fragmented elections.

This research note's aims to offer an empirical contribution primarily: complementing macro-level evidence of polling data with meso- and micro-level data. Therefore, after a short discussion of previous research on this topic, we discuss the macro, meso and micro levels, outlining our expectations and presenting our findings for each level in turn.

The predictive value of vote intention polls

Polls reporting voters' intended votes receive widespread media attention (Brettschneider 1997; Patterson 2005). Even though these polls represent a snapshot of vote intentions at a particular point in time, they are often used as de facto forecasts of the upcoming election result (Weimann 1990), or they are at the very least used to speculate about it. The reports of most polls are not designed to offer a good description of public opinion at any given time. Unlike actual opinion poll reports, vote intention polls often do not report the raw data. Rather, the undecideds and non-responses are "added to one of the parties or to the non-voters, using past experience, specific house procedures and weighting factors specific to election forecasts" (Donsbach and Hartung 2008, p. 433). Some polls generally do not even report percentages but rather the supposed parliamentary seat distribution (or likelihood of a win), as the outcome of the distribution of respondents' vote intentions.

Because of the way vote intention poll reports are designed, it comes as no surprise that they are used as forecasts by voters, media, and politicians. Voters employ polls to forecast election outcomes, a.o. to inform strategic voting behaviour (Irwin and Van Holsteyn 2002; Meffert and Gschwend 2011; see also Zerback et al. 2015). While media are the gatekeepers of polling data (Toff 2019), they tend to prefer polls that predict clear winners and losers (Goot 2021) in order to speculate about the outcomes (Weimann 1990; Brettschneider 1997; Toff 2019), even months before the elections take place. Media are particularly influential in framing dry figures as a win or a loss (Van der Meer et al. 2016). Politicians, finally, tend to assume the predictive value of polls in their campaign narratives (Jaworski and Fitzgerald 2008).

Yet, vote intention polls face various challenges that have made registering, let alone forecasting, electoral behaviour more difficult. Societal cleavages tend to structure party preferences less strongly than they did decades ago (Evans and Tilley 2012). Party systems in many western democracies have undergone a dimensional

transformation (Kriesi et al. 2008; Pellikaan et al. 2007). The share of volatile, floating voters and late deciders is on the rise. Rising levels of fragmentation, particularly high in the Netherlands, have particularly stimulated volatility within (idiosyncratic) blocks of ideologically similar parties (Van der Meer et al. 2012) that form so-called consideration sets (Oscarsson and Rosema 2019). Moreover, the quality of opinion polls has come under pressure, as traditional sampling frames such as phone registers have become less complete, and response rates have declined (Keeter et al. 2017). Nevertheless, there are no indications that election outcomes have become structurally less predictable over time (Jennings and Wlezien 2016, 2018; Nadeau et al. 2020).

Macro-level analysis: overall predictive accuracy of vote intention polls

To assess the predictive accuracy of vote intention polls, scholars, polling agencies, and journalists implicitly work with benchmarks. Evidently, the upper benchmark is perfect alignment. But the lower benchmark is up for debate. Often, the predictive accuracy of polls is assessed in comparison to total error, implying that the pre-electoral polls are informative when a forecast is more than 0% accurate. From that perspective, "the final outcome is fairly clear in the polls before the election campaign really begins" (Jennings and Wlezien 2016). However, polls are not organized in a zero-information environment. One benchmark is formed by the result of the previous election. Vote intention polls then become informative forecasts when their predictive accuracy is higher than that of the previous election result.

To assess the predictive value of vote intention polls in vote shares at the macro-level, we gathered the poll results from the five main polling agencies in The Netherlands between 1998 and 2021. We restricted our analysis to this timeframe for reasons of data availability and the low number of active polling agencies before 1998.[2] The data were retrieved from the reports published on the corresponding polling agency's websites and were, when necessary, checked for accuracy with published newspaper articles written on the polls' outcomes. Figure 1 depicts the number of seats that the polls assigned identical to the subsequent election result, including a benchmark for the previous election result.

Figure 1 provides three sets of insights. First, the figure shows that the interim polls of all polling agencies are able to forecast the upcoming election result quite well: approximately 70% (105 seats) to 90% (135 seats) of the 150 seats in the Dutch Lower House are predicted correctly over the months and even years leading up to election day. On the one hand, this is high when we take the high volatility of Dutch elections into account. On the other hand, similar percentages are attained by the benchmark included in the graph. In other words, vote intention polls generally offer

[2] Moreover, the 1998–2002 government period ended transforming the Dutch party system (Pellikaan et al. 2007).

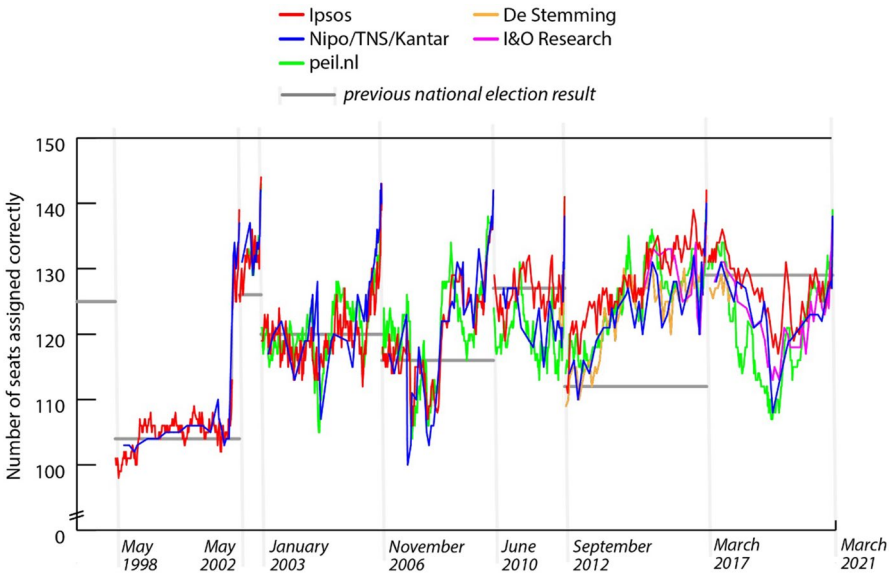

Fig. 1 The macro-level predictive value of polls 1998–2021 (5 pollsters, 7 full electoral cycles)

no more predictive value than the previous election outcome during long stretches before the elections.

Second, the timing of the poll clearly plays a role. We see a steep increase in the predictive value of all polls in the final weeks or months of the election cycle. Sometimes this steep increase takes place in the final weeks (such as in 2012, with a volatile campaign); in other years this is a longer, gradual process (such as in 2002, 2006, 2010, and 2017). This finding is very much in line with prior research by Jennings and Wlezien (2016), who conclude that polls become increasingly revealing about the election result as Election Day comes closer. In fact, only in the last phase of the election cycle do vote intention polls become structurally better at forecasting the election result compared to the previous election outcome.

However, several election cycles show trends that differ from this structural pattern. The 2012–2017 election cycle is remarkable, as Fig. 1 shows a rather structural upward trend from the onset. This is largely due to the 2012 elections, when large shares of voters opted strategically for either the conservative party (VVD) or particularly the social-democratic party (PvdA) to keep the other out of government. When the two parties formed a coalition after the 2012 elections, these strategic voters were quick to renounce their support. The 2006–2010 cycle also differs from the pattern, as polls became substantially more informative halfway through the cycle. This was due to a specific external event, i.e., the nationalization of the ABN Amro Bank in October 2008, which boosted support for the junior-coalition partner PvdA (see also Van der Meer et al. 2015).

Third, when we compare the different polling agencies to one another, the similarities stand out. The structural differences between the polling agencies are relatively small, and they all capture similar trends. None of the polling agencies' polls

seem to perform structurally better or worse by a substantive measure in forecasting seat shares over the whole timeframe under study (1998–2021). If any of the pollsters stands out, Ipsos tends to have the most stable trends and a somewhat higher predictive value (i.e., allocates the highest share of seats in line with the election outcome) over the last three election cycles.

Meso level analysis: differences between parties and pollsters

Even though there are relatively few differences between pollsters when it comes to the macro-level differences between their polls and the election result, clear differences between pollsters exist with regard to the vote intentions for each party. Particular parties structurally score higher or lower in the polls of some pollsters than those of others; this is the so-called pollster's 'house effect' (Converse and Traugott 1986, p. 1095; Jackman 2005). While house effect primarily stem from diverging methodologies of polling companies (Traugott 2014), the size of the house effect can be related to party characteristics (De Stefano et al. 2022, p. 21). In particular, 'challenger' parties,[3] which campaign on the basis of an at least somewhat anti-establishment platform, often receive varying scores in vote intention polls of various pollsters. Voters for these types of parties may be harder to reach for pollsters and it is more difficult to use poststratification weights based on previous voting behaviour, particularly for new challenger parties. Overall, we therefore expect lower levels of predictive accuracy for challenger parties, as well as larger differences between pollsters when it comes to the predictive accuracy of their estimates of vote intentions for challenger parties.

We tested this expectation using a generalized additive model (GAM) that explains the absolute difference between a party's standing in a particular poll and the number of seats obtained by that party in the subsequent election.[4] For example, if a party was on 30 seats in a poll and received 34 seats in the election, the absolute difference is 4. Using the GAM, we can model this absolute difference as a smooth function of the number of days until the election, which is more flexible than assuming a linear trend. Apart from the time until the election, we include whether a party was classified as a challenger (and its interaction with time), the pollster (and its interaction with time), the party's score at the next election and the parliamentary term in the model. We are mostly interested in the differences between pollsters and between challenger and non-challenger parties.

Figure 2 shows that our model estimates the differences between polls and election results to be generally larger for challenger parties than for non-challengers. At the same time, we observe substantial differences between pollsters in this respect. For most of the election cycle, the 'predictive accuracy gap' (between challenger and

[3] For the purposes of our analysis we treated the following parties as challengers: BIJ1, 50PLUS, Code Oranje, Denk, DPK, FvD, GeenPeil, Nieuwe Wegen, GroepVerdonk, PVV, JA21, BBB, LHK, Splinter.

[4] For this analysis, we focus the election cycles since 2006, because all pollsters have used internet based polling methods since this election.

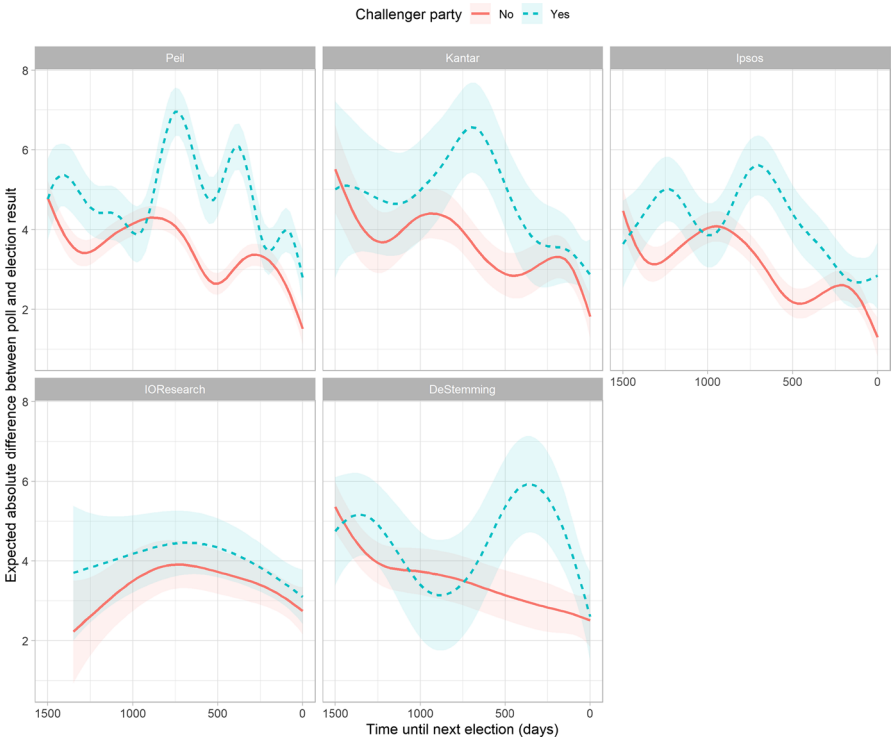

Fig. 2 Expected values for the absolute difference between a poll estimate and party election score (Based on data for 2006–2021. Expected values are simulated keeping all other variables at their mean or mode)

non-challenger parties) is smaller in the polls of I&O Research compared to Peil, De Stemming and Kantar. Only at the start and at the very end of an electoral cycle this predictive accuracy gap diminishes. Our data show that the predictive accuracy gap is mostly due to challenger party vote intention estimates that are higher than their eventual election result, particularly for De Stemming and Peil polls.[5] Even though this may reflect actual trends in vote intention—challenger parties may indeed 'peak' in support mid-term and then fall back to lower levels of support—this shows that any mid-term gains for these types of parties need to be interpreted with some caution and not be treated as very predictive for the election result.

While some pollsters are, therefore, more optimistic about the vote intentions for challengers mid-way through an electoral cycle, we also observe that the gap between challengers and non-challengers as well as the differences between pollsters seem to become smaller as the election draws closer. This raises the question if some pollsters might 'follow' other pollsters in picking up trends in party support. It may be the case

[5] We ran the same analysis with the difference (positive or negative) between a poll and the election result as the dependent variable. The results are available in "Appendix 2".

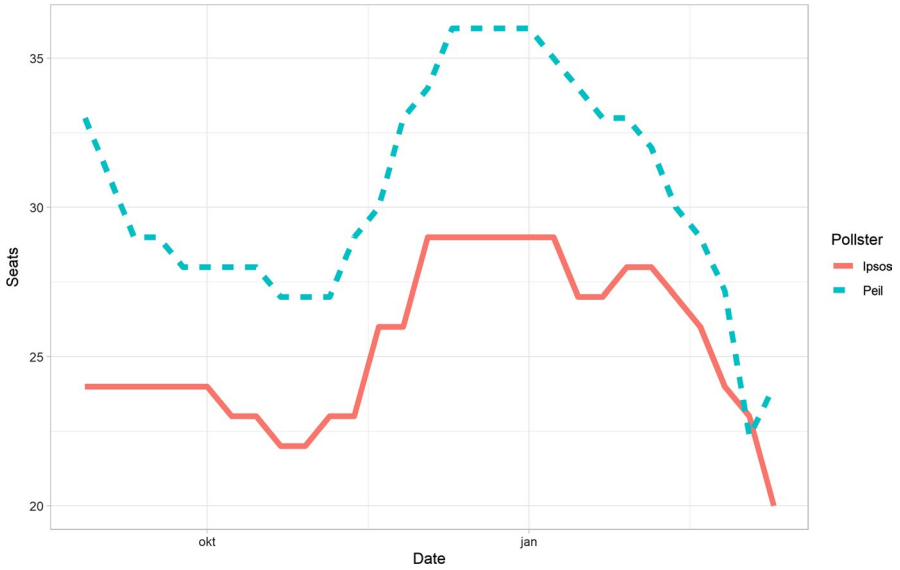

Fig. 3 Weekly seat estimates for PVV 2016–2017, Ipsos and Peil polls

that some pollsters are able to pick up changes in vote intentions more quickly than others. To study these patterns, we estimated a Vector Autoregressive Model (VAR), in which the weekly vote intention estimates for one party for one pollster are estimated as a linear function of all pollster's estimates for the party from the previous two weeks (we used two lags). If one pollster was 'leading' the others, we would estimate significant and positive coefficients for that pollster's lagged terms. We have studied these patterns for the five largest parties in the final nine months of the 2012–2017 election cycle, which we selected due to the relatively high number of pollsters (5) who polled during this time. If a pollster did not poll during a week, we took its estimate for the previous week; when it polled more than once, we took the mean estimate.

Overall, we do not find evidence of one pollster 'leading the pack'. For the purpose of illustration, consider the PVV, for which the scores of pollsters Ipsos and Peil are plotted in Fig. 3. We see that the rise in PVV scores in late 2016 was picked up by both pollsters, and perhaps slightly earlier by Ipsos, but the decline was visible earlier in the Peil polls. We should note that the party won 20 seats at the elections, so one could argue that the Peil estimates were mostly trending towards the Ipsos estimates over this period of time. The VAR results show that both Ipsos polls are somewhat predictive of Peil polls and the other way around (see "Appendix 3"). This suggests that both polls measure the same trends and that at some point in time the one pollster picks a trend up slightly earlier than the other, and vice versa. For the other parties, we see even less evidence for particular pollsters picking up trends before others.

Micro level analysis: predictive accuracy and consistency of vote intentions of voters

Finally, we turn to the micro-level. Macro-level predictive accuracy must be underpinned by micro-level electoral preferences. The predictive accuracy of vote intention polls is likely to differ across subgroups of voters. The literature on electoral volatility (e.g., Lachat 2007; Van der Meer et al. 2015) and the literature on predictability (e.g., Box-Steffensmeier et al. 2015) are informative in this respect. Partisans have higher levels of stability and predictability than non-partisans (Campbell et al. 1960). With regards to political sophistication, the conclusions tend to be a bit more mixed: whereas high levels of sophistication tend to provide some higher level of stability of preferences, low levels of sophistication may undermine voters' responsiveness to external events and thereby stabilize preferences. Hence, various studies concluded that particularly middle groups are most likely to be volatile and less predictable (cf. Lachat 2007; Van der Meer et al. 2015). Macro-level analyses of predictive accuracy cannot pull these differences apart, but micro-level analyses on preferential stability in these polls can.

For that purpose, we analyse detailed panel data on two election cycles. Dutch daily news show EenVandaag has long hosted its own public opinion panel. The data on the 2006–2010 and 2010–2012 election cycles have been made available for scholarly research under the label 1VOP (cf. Aaldering et al. 2018). While not a random sample of the general Dutch population (a.o. overestimating the share of older and higher educated voters), the 1VOP data are particularly effective in assessing within-person changes over time.

Evidently, it is crucial for the analyses to link respondents' vote intention to the vote they would cast at the subsequent Lower House elections. For the analysis in this contribution, we therefore rely solely on respondents who reported the political party they voted for during the election in the post-election poll.[6] Additionally, to assess the *predictive accuracy of the polls* in both election cycles, we selected respondents who participated in at least 1 poll before these elections. For the assessment of the *consistency rate of the vote intention*, we relied on respondents who participated in at least 13 (for the 2010 election cycle) or 9 (for the 2012 election cycle) polls before these elections.

All in all, our analyses on the 2010 election cycle cover 27,572 respondents who reported their voting behavior in 2010, of whom 98.6% participated in at least one of maximally 46 polls[7] between 2006 and 2010, and 71.8% participated in at least 13 polls. Our analyses on the 2012 election cycle cover 35,574 respondents who reported their voting behavior in 2010, of whom 96.9% participated in at least one of 31 polls between 2010 and 2012, and 72.3% participated in at least 9 polls. We weighted the analyses by the vote cast in the subsequent election, and checked the robustness of our findings against other participation rates.

[6] We thereby also exclude non- and blanc-voters from our analyses.

[7] We left out polls that only approached a subset of the 1VOP panel, such as party members or adherents of specific parties.

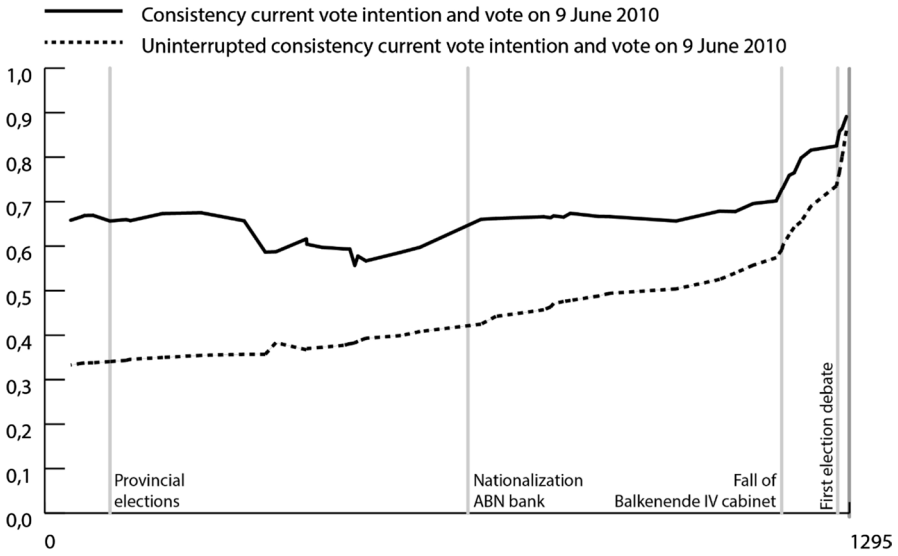

Fig. 4 The micro-level predictive value of vote intentions (1VOP, 2006–2010)

First, we cover the 2006–2010 election cycle. The straight line in Fig. 4 shows the share of respondents in each poll between 2006 and 2010 whose vote intention matches the vote that they would eventually cast on election day in 2010. Up to the fall of the incumbent government in February 2010, this share fluctuates between 55 and 70%. These figures are substantially lower than the predictive accuracy of the macro-polls in Fig. 1 (circling around 80% between 2006 and 2010). There are two reasons for this difference. First, at the micro-level, we account for respondents who do not yet know their vote intention or do not (yet) want to share it. Second, the micro-level reports the gross levels of similarity, whereas the macro-level reports the net similarity of polls: at the macro-level the predictive inaccuracies of multiple respondents can be canceled out against each other, particularly if these inaccuracies are random.[8]

The broken line in Fig. 4 represents the share of respondents who from that poll onwards do not report an intention to vote for any other party than the one ultimately voted for at the 2010 elections. We only cover respondents who participated in at least 12 waves. In any of the intervening polls, these respondents may not have participated, may have reported not knowing which party to vote for or may have refused to answer the question. Yet, as long as they do not mention a different party

[8] Consider the easiest example of two respondents. Respondent A who intends to vote for the PvdA but votes for the VVD, and Respondent B who intends to vote for the VVD but votes for the PvdA. At the micro-level both respondents are classified as inaccurate. At the macro-level the errors cancel each other out.

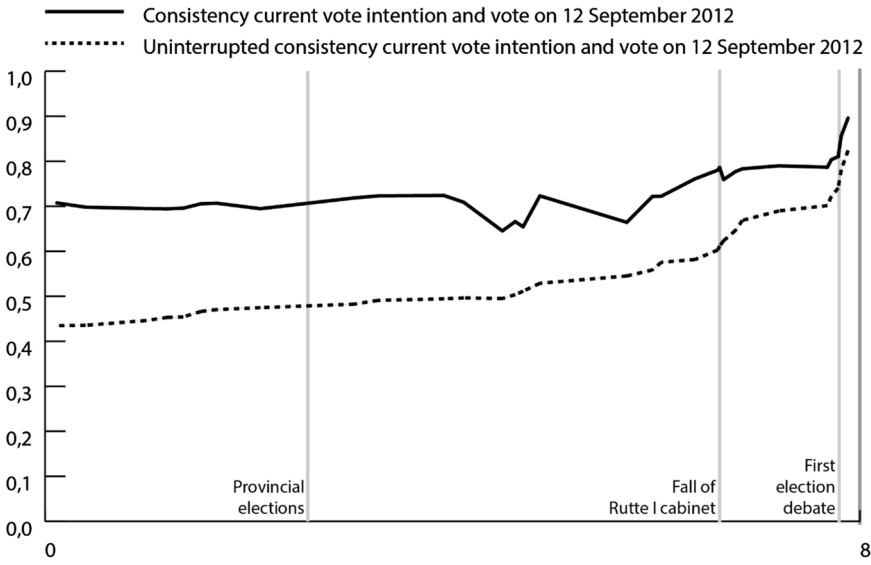

Fig. 5 The micro-level predictive value of vote intentions (1VOP, 2010–2012)

than before, we analyze these answers as not showing evidence for a changed substantive vote intention.[9]

Evidently, the consistency rate of the vote intention in Fig. 4 is lower than the predictive accuracy. The consistency rises steadily from 33% weeks after the previous election to 59% before the fall of the incumbent government, less than four months before the upcoming elections. We performed event history analyses to assess whether the timing when vote intentions get cemented differs with gender, education, and partisanship (see "Appendix 1"). Although levels of consistency evidently differ across groups of voters—most notably, partisans are more likely to have firmly cemented vote intentions—the *timing* when vote intentions become fixed follows similar trends across these groups.

Figure 4 shows that the predictive accuracy rises at three moments: (1) the nationalization of ABN Amro Bank in 2008 (which led to a small rally 'round the flag effect and could also be witnessed in Fig. 1), (2) the fall of the incumbent government in February 2010, and (3) the start of the election campaign in May 2010, a few weeks before the elections. Just like at the macro-level, we find that the predictive value of reported vote intentions in polls increases particularly in the final months before the election. The consistency rate, by contrast, rises in response to

[9] We checked the robustness of our findings to a different operationalization of these non-substantive answers between two different substantive preferences. In this robustness check we coded these non-substantive answers as evidence for the upcoming change (rather than as not yet evidence for the upcoming change). Our findings are substantially robust to this difference. Although the curve is somewhat steeper in the robustness check, particularly in the months after the fall of the government coalitions in 2010 and 2012, the differences in the patterns central to this paper are small (see "Appendix 4").

the latter two events (fall of government, start of the campaign), but not to the earlier event (nationalization).

Figure 5 paints a rather similar picture of the micro-level predictive accuracy between 2010 and 2012. The straight line shows that—up to the fall of the incumbent Rutte I government—65% to 72% of the respondents has a vote intention similar to the vote they cast at the 2012 elections. Predictive accuracy remains on a somewhat higher plateau (75–80%) until the final weeks before the 2012 elections. The broken line shows the share of respondents who do not change their vote intentions in any intermediate poll (and who participated in at least 9 survey waves). Although we generally see a rise in cemented vote intentions over time (from 43 to 59%), the consistent predictive accuracy of these vote intentions rises more rapidly after the fall of the incumbent government, and with the short election campaign. Additional event history analyses confirm that the timing of this trend shows similarities across different subgroups, albeit starting out from different levels of consistency (see "Appendix 1").

Conclusion

Even though vote intention polls are not strictly equipped to forecast the future election results, they are nevertheless often used for that purpose by voters, media, and even politicians. This research note set out to assess the predictive value of vote intention polls not only at the macro-level (complete poll results, set against a substantive benchmark), but also at the meso-level (parties, pollsters), and at the micro-level (voters). To that purpose, we focused on the highly proportional, fragmented multiparty system of the Netherlands between 1998 and 2021.

First, we find that vote intention polls tend to predict 70% to 90% of the seats in the Dutch Lower House correctly in the months and years before the election. This suggests that the grand outlines of the election outcomes are indeed visible in the polls far before the elections take place, similar to findings in the international-comparative analyses (Jennings and Wlezien 2016, 2018). However, this predictive accuracy is not very impressive once we take a crucial benchmark into account, i.e., the current composition of parliament as the outcome of the *previous* election. Vote intention polls only provide structurally more predictive power than this benchmark in the last few weeks (sometimes: months) before the elections. The added predictive value of vote intention polls rises particularly during the election campaign.

There are exceptions to this general pattern. During one election cycle vote intention polls were more informative than the benchmark from the start. However, that had more to do with the previous elections (e.g. high levels of short-term strategic voting) than with the upcoming elections.

Second, the predictive accuracy of vote intention polls is even lower for challenger parties than for non-challenger parties. Challenger parties tend to do better in opinion polls midway through the election cycle than in the election. Some pollsters suffer more from this issue than others. Nevertheless, we find no evidence that one pollster leads the other(s) in picking up trends, or that one pollster follows the other(s).

Third, vote intention polls are not very informative predictors of actual vote choice at the micro-level. Predictive accuracy at the micro-level is much lower than at the macro-level as voters' preferences are cemented relatively late. In the aggregate (i.e., in polls as a whole, at the macro-level) some of these micro-level predictive inaccuracies cancel each other out. While some groups of voters are more consistent than others, the *timing* of the events is quite similar across groups, in line with the temporal patterns found at the macro-level. All in all, the predictive accuracy of polls diminishes at lower levels of aggregation, but differences between voters hardly account for changes in that accuracy.

Data availability in the Netherlands, a highly proportional system with a diverse set of political parties, allowed us to identify the predictive value of vote intention polls at three levels of analysis. Yet, one can imagine that the patterns identified in this research note play out differently in systems that are less proportional (leading to more uncertainty in the translation of votes into seats, to less fragmentation, and an incentive of exit over voice), systems that are less fragmented (leading to less volatility within blocks of parties), and to systems that are more polarized (leading to less volatility between blocks). This calls for further analyses at the meso- and micro-level in different types of countries.

All in all, we conclude that vote intention polls are not very informative forecasts of the future election outcome at the micro-, meso-, and even macro-level, beyond the established benchmark of the extant composition of parliament determined by the previous election. This is mainly due to changes in voter preferences throughout the electoral cycle as well as undecided voters making up their minds. Voters, media, and politicians should therefore be modest in their use of vote intention polls, using them as descriptive (but imperfect) *nowcasts* rather than predictive election *forecasts*. Forecasting elections via vote intention polls runs the risk of procyclical swings: the more trends in vote intention polls are considered meaningful, they can become influential by strategic and bandwagon voting (by voters), selective media attention (by journalists), and political strategizing (by politicians).

We may be better off by not using vote intention polls as election forecasts, unless when they are embedded in more elaborate forecasting models that also take into account structural factors and historical trends (e.g. Lewis-Beck and Dassonneville 2015). Yet, to date, such elaborate forecast models are employed much less commonly in Western Europe than regular vote intention polls.

Appendix 1

Event history analysis allows us to explain the timing of a specific event. The event that we are interested in here, is the moment when the vote intention becomes fixed (as an indicator of micro-level predictive accuracy). EHA only studies respondents that are at risk (and, hence, not stable in their vote intention throughout the period under study). It is evident that party members, in particular, are more stable in their vote intention. EHA shows the differences between groups, *given that they are at risk*. It thus analyses the timing rather than the level.

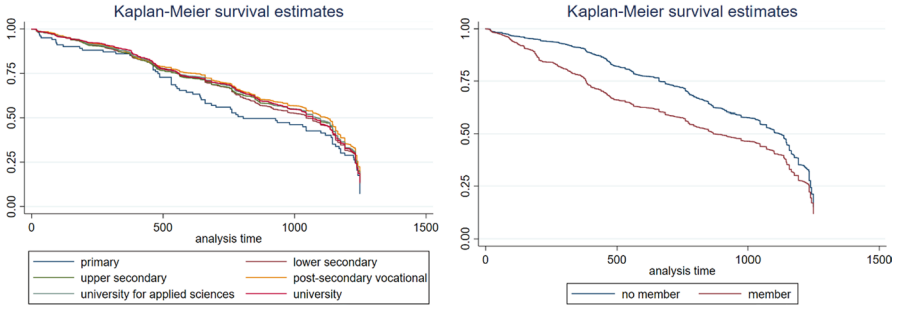

Fig. 6 Survival estimates 2006–2010, by level of education and party membership

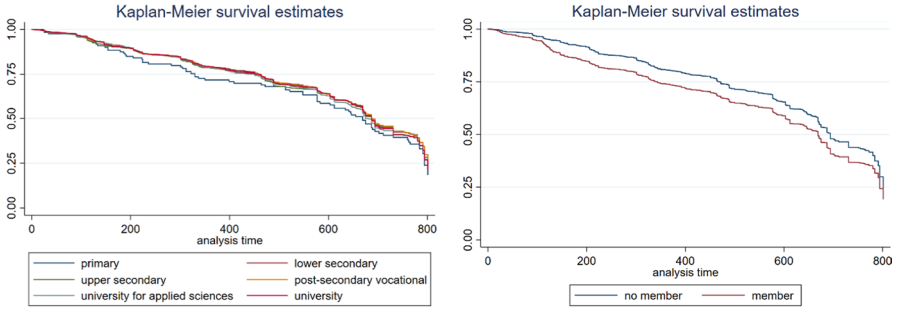

Fig. 7 Survival estimates 2010–2012, by level of education and party membership

We estimated EHA models in Stata 15 using the sts commands.

Figure 6 shows the outcomes of EHA for the 2006–2010 election cycle; Fig. 7 the outcomes for the 2010–2012 cycle. We find that the differences between levels of education tend to be rather small, except for the (small and deviant group of) respondents who only had primary education. That relatively old group tends to experience the event (fixed, accurate preference) earlier.

For party membership, we find more defined differences. Party members are not only less likely to be at risk in general (having more stable preferences); if they are at risk, they are substantially more likely to cement their new preference. This difference between members and non-members is stronger between 2006 and 2010 than between 2010 and 2012.

Yet, given these differences, we see rather similar trends in the timing of these events across groups. The trends accelerate at similar moments, particularly in the last few days and weeks before the election.

Appendix 2: Full results of the macro-level models

See Table 1 and Fig. 8.

Table 1 Output of GAM regression model with linear link

	Model 1 DV: absolute difference	Model 2 DV: difference
(Intercept)	1.94 (0.17)***	3.18 (0.25)***
Pollster (Ref: De Stemming)		
IandO Research	− 0.10 (0.24)	0.66 (0.35)
Ipsos	− 0.46 (0.17)**	0.60 (0.24)*
Kantar	0.12 (0.19)	0.52 (0.27)
Peil	− 0.04 (0.16)	0.15 (0.23)
Challenger	0.98 (0.31)**	2.31 (0.46)***
Seats at next election	0.25 (0.00)***	− 0.26 (0.00)***
Term (Ref: 2006–2012)		
Term 2010–2012	− 0.22 (0.10)*	0.02 (0.15)
Term 2012–2017	− 1.66 (0.09)***	− 0.47 (0.13)***
Term 2017–2021	− 1.49 (0.09)***	− 0.60 (0.14)***
Pollster × Challenger interaction		
I&O Research × Challenger	− 0.30 (0.46)	− 2.34 (0.68)***
Ipsos × Challenger	0.13 (0.35)	− 2.32 (0.51)***
Kantar × Challenger	0.18 (0.41)	− 2.20 (0.60)***
Peil × Challenger	0.41 (0.33)	− 0.29 (0.48)
Effective degrees of freedom of smooth terms		
EDF: time till next election × De Stemming × Non-challenger	4.09 (5.07)***	1.00 (1.00)
EDF: time till next election × IandO Research × Non-challenger	2.67 (3.30)*	1.00 (1.00)
EDF: time till next election × Ipsos × Non-challenger	7.94 (9.65)***	1.00 (1.00)***
EDF: time till next election × Kantar × Non-challenger	6.74 (8.26)***	1.00 (1.00)
EDF: time till next election × Peil × Non-challenger	9.16 (10.94)***	4.79 (5.96)***
EDF: time till next election × De Stemming × Challenger	4.79 (5.93)**	2.72 (3.39)**
EDF: time till next election × IandO Research × Challenger	1.95 (2.40)	1.89 (2.33)**
EDF: time till next election × Ipsos × Challenger	7.46 (9.10)***	9.16 (10.90)***
EDF: time till next election × Kantar × Challenger	5.72 (7.06)***	4.09 (5.09)**
EDF: time till next election × Peil × Challenger	12.01 (13.33)***	12.29 (13.48)***
AIC	85,971.32	98,638.52
BIC	86,565.18	99,051.63
Log likelihood	− 42,908.14	− 49,265.33
Deviance	218,403.66	491,275.24
Deviance explained	0.37	0.21
Dispersion	13.99	31.43
R^2	0.37	0.21
GCV score	14.06	31.54
Num. obs.	15,684	15,684
Num. smooth terms	10	10

Coefficients with standard errors in parentheses. For the smooth terms the effective degrees of freedom is displayed with reference degrees of freedom

***$p < 0.001$; **$p < 0.01$; *$p < 0.05$

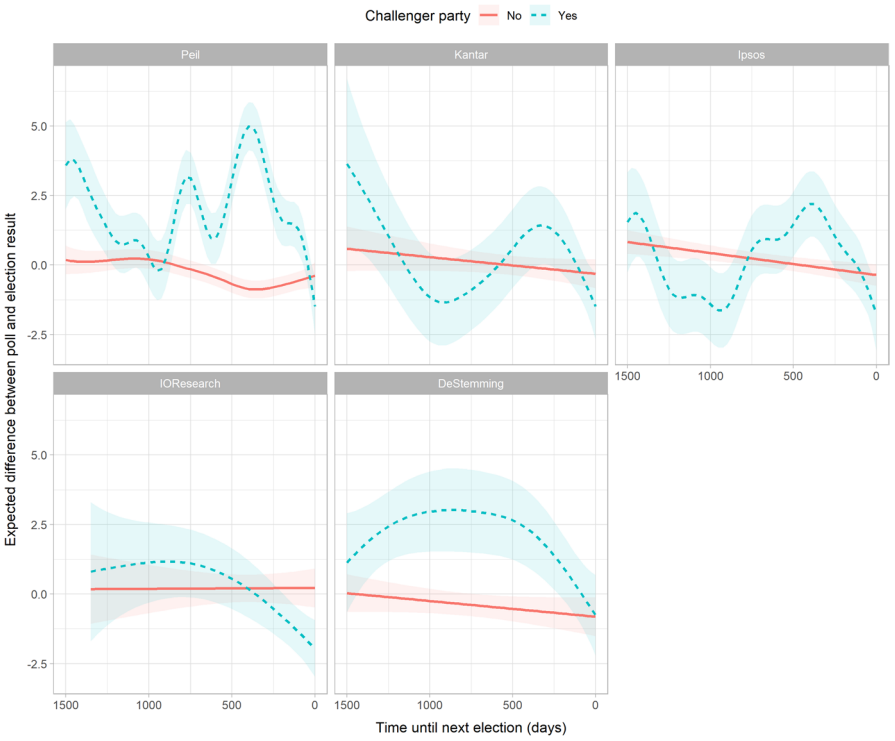

Fig. 8 Visualisation of output GAM model for difference (Model 2 in Table 1)

Appendix 3: Results of VAR models

See Tables 2, 3, 4, 5 and 6.

Table 2 VAR models for VVD seats in polls 2016–2017

	Ipsos	Peil	Kantar	IOResearch	DeStemming
Ipsos.l1	0.34 (0.27)	− 0.06 (0.19)	− 0.26 (0.36)	− 0.11 (0.24)	0.19 (0.15)
Peil.l1	0.59 (0.35)	0.71* (0.25)	0.33 (0.47)	0.20 (0.31)	− 0.07 (0.20)
Kantar.l1	0.01 (0.20)	0.19 (0.14)	0.28 (0.27)	0.27 (0.18)	− 0.08 (0.11)
IOResearch.l1	− 0.13 (0.32)	0.04 (0.24)	0.24 (0.44)	0.42 (0.29)	0.17 (0.19)
DeStemming.l1	0.46 (0.47)	0.30 (0.34)	− 0.13 (0.64)	0.92* (0.42)	0.33 (0.27)
Ipsos.l2	0.09 (0.31)	− 0.10 (0.22)	0.50 (0.42)	− 0.21 (0.27)	− 0.01 (0.18)
Peil.l2	− 0.16 (0.38)	0.34 (0.27)	0.58 (0.51)	0.30 (0.33)	0.17 (0.22)
Kantar.l2	− 0.39 (0.21)	− 0.30 (0.15)	− 0.55 (0.28)	− 0.02 (0.18)	0.04 (0.12)
IOResearch.l2	0.13 (0.29)	0.09 (0.21)	− 0.58 (0.39)	− 0.22 (0.25)	0.13 (0.16)
DeStemming.l2	− 0.08 (0.47)	− 0.64 (0.34)	− 0.20 (0.63)	− 0.21 (0.42)	− 0.27 (0.27)
Const.	5.71 (8.54)	11.31 (6.19)	19.85 (11.57)	− 6.40 (7.60)	8.65 (4.90)
R^2	0.75	0.85	0.63	0.82	0.77
Adj. R^2	0.59	0.75	0.40	0.71	0.63
Num. obs.	27	27	27	27	27

Table 3 VAR models for PVV seats in polls 2016–2017

	Ipsos	Peil	Kantar	IOResearch	DeStemming
Ipsos.l1	0.31 (0.23)	1.27*** (0.29)	0.05 (0.85)	0.30 (0.81)	0.14 (0.41)
Peil.l1	0.71*** (0.17)	0.85** (0.22)	0.77 (0.64)	1.54* (0.62)	0.37 (0.31)
Kantar.l1	− 0.02 (0.07)	− 0.29** (0.09)	0.20 (0.27)	− 0.11 (0.26)	− 0.10 (0.13)
IOResearch.l1	− 0.08 (0.07)	0.01 (0.09)	0.26 (0.26)	0.49 (0.25)	0.05 (0.13)
DeStemming.l1	0.09 (0.14)	0.33 (0.18)	0.26 (0.52)	− 0.20 (0.50)	0.40 (0.25)
Ipsos.l2	0.14 (0.23)	− 0.58 (0.28)	0.46 (0.83)	− 1.54 (0.80)	− 0.56 (0.40)
Peil.l2	− 0.19 (0.18)	− 0.18 (0.22)	0.18 (0.66)	− 0.28 (0.63)	0.04 (0.32)
Kantar.l2	− 0.11 (0.08)	− 0.02 (0.10)	− 0.16 (0.29)	0.27 (0.28)	0.04 (0.14)
IOResearch.l2	− 0.02 (0.07)	0.24* (0.09)	− 0.25 (0.28)	− 0.15 (0.26)	0.10 (0.13)
DeStemming.l2	0.05 (0.15)	− 0.45* (0.19)	− 0.66 (0.55)	− 0.62 (0.52)	− 0.03 (0.27)
Const.	0.22 (4.07)	− 2.02 (5.11)	− 3.06 (15.01)	28.91 (14.39)	13.80 (7.29)
R^2	0.95	0.96	0.81	0.75	0.72
Adj. R^2	0.92	0.94	0.69	0.59	0.54
Num. obs.	27	27	27	27	27

Table 4 VAR models for CDA seats in polls 2016–2017

	Ipsos	Peil	Kantar	IOResearch	DeStemming
Ipsos.l1	0.57* (0.26)	− 0.11 (0.19)	0.12 (0.27)	− 0.27 (0.24)	− 0.31 (0.20)
Peil.l1	0.23 (0.43)	1.23** (0.31)	0.96* (0.43)	1.45** (0.39)	1.03** (0.32)
Kantar.l1	− 0.07 (0.25)	− 0.06 (0.18)	0.42 (0.26)	− 0.24 (0.23)	− 0.27 (0.19)
IOResearch.l1	− 0.36 (0.34)	− 0.24 (0.24)	0.58 (0.34)	1.11** (0.31)	0.19 (0.25)
DeStemming.l1	0.02 (0.45)	0.38 (0.33)	− 0.96 (0.46)	− 1.00* (0.41)	0.04 (0.34)
Ipsos.l2	− 0.40 (0.26)	− 0.04 (0.19)	0.12 (0.26)	0.45 (0.23)	0.10 (0.19)
Peil.l2	0.72 (0.80)	− 0.32 (0.58)	− 0.48 (0.81)	− 0.73 (0.73)	− 0.16 (0.59)
Kantar.l2	− 0.13 (0.28)	− 0.00 (0.20)	− 0.19 (0.28)	0.19 (0.25)	0.18 (0.21)
IOResearch.l2	0.52 (0.40)	0.51 (0.29)	− 0.12 (0.41)	− 0.75 (0.37)	− 0.65* (0.30)
DeStemming.l2	− 0.55 (0.48)	− 0.41 (0.35)	0.60 (0.49)	0.56 (0.44)	0.33 (0.36)
Const.	10.28 (5.81)	2.10 (4.21)	0.13 (5.91)	2.69 (5.28)	8.49 (4.32)
R^2	0.68	0.91	0.78	0.82	0.89
Adj. R^2	0.48	0.85	0.65	0.70	0.83
Num. obs.	27	27	27	27	27

Table 5 VAR models for D66 seats in polls 2016–2017

	Ipsos	Peil	Kantar	IOResearch	DeStemming
Ipsos.l1	0.47 (0.27)	0.04 (0.16)	0.19 (0.42)	0.46 (0.54)	0.05 (0.32)
Peil.l1	0.43 (0.42)	0.51 (0.25)	− 0.69 (0.65)	1.01 (0.83)	− 0.94 (0.50)
Kantar.l1	0.04 (0.18)	− 0.04 (0.11)	0.54 (0.29)	− 0.24 (0.37)	− 0.28 (0.22)
IOResearch.l1	− 0.09 (0.11)	− 0.03 (0.07)	0.02 (0.18)	0.83** (0.23)	0.12 (0.14)
DeStemming.l1	− 0.06 (0.17)	− 0.16 (0.10)	− 0.25 (0.27)	0.01 (0.35)	0.65** (0.21)
Ipsos.l2	− 0.11 (0.24)	− 0.04 (0.14)	− 0.21 (0.38)	− 1.29* (0.48)	0.36 (0.29)
Peil.l2	− 0.94 (0.56)	− 0.25 (0.33)	0.60 (0.88)	− 1.45 (1.12)	1.39 (0.67)
Kantar.l2	0.03 (0.18)	0.18 (0.11)	0.19 (0.28)	0.07 (0.36)	0.32 (0.21)
IOResearch.l2	− 0.01 (0.13)	0.17* (0.08)	0.30 (0.20)	− 0.17 (0.26)	0.23 (0.16)
DeStemming.l2	0.11 (0.18)	0.04 (0.10)	− 0.40 (0.28)	− 0.19 (0.36)	− 0.16 (0.21)
Const.	18.03 (8.89)	8.51 (5.24)	12.41 (13.88)	31.91 (17.80)	− 10.78 (10.69)
R^2	0.47	0.75	0.79	0.74	0.64
Adj. R^2	0.14	0.60	0.67	0.58	0.42
Num. obs.	27	27	27	27	27

Table 6 VAR models for GroenLinks seats in polls 2016–2017

	Ipsos	Peil	Kantar	IOResearch	DeStemming
Ipsos.l1	0.70** (0.22)	− 0.07 (0.13)	0.02 (0.19)	− 0.02 (0.23)	0.01 (0.14)
Peil.l1	− 0.06 (0.41)	0.29 (0.25)	− 0.31 (0.36)	0.07 (0.43)	− 0.13 (0.27)
Kantar.l1	0.07 (0.32)	− 0.12 (0.19)	0.65* (0.28)	1.11** (0.33)	− 0.03 (0.21)
IOResearch.l1	− 0.10 (0.21)	0.09 (0.13)	0.22 (0.18)	0.63* (0.22)	0.05 (0.14)
DeStemming.l1	− 0.53 (0.34)	0.11 (0.21)	− 0.14 (0.29)	0.15 (0.36)	0.65* (0.22)
Ipsos.l2	− 0.51* (0.22)	0.09 (0.13)	− 0.17 (0.19)	0.10 (0.23)	− 0.41* (0.15)
Peil.l2	0.10 (0.51)	0.54 (0.31)	0.40 (0.44)	− 0.21 (0.53)	0.07 (0.34)
Kantar.l2	− 0.11 (0.46)	− 0.01 (0.28)	− 0.48 (0.40)	− 1.21* (0.48)	0.00 (0.30)
IOResearch.l2	0.01 (0.18)	0.09 (0.11)	0.01 (0.15)	− 0.03 (0.19)	0.08 (0.12)
DeStemming.l2	1.03* (0.41)	0.04 (0.25)	0.38 (0.35)	0.44 (0.43)	0.43 (0.27)
Const.	6.12 (3.60)	− 0.24 (2.17)	4.98 (3.10)	2.11 (3.75)	3.57 (2.35)
R^2	0.68	0.90	0.74	0.91	0.95
Adj. R^2	0.48	0.84	0.58	0.86	0.92
Num. obs.	27	27	27	27	27

Appendix 4

See Figs. 9 and 10.

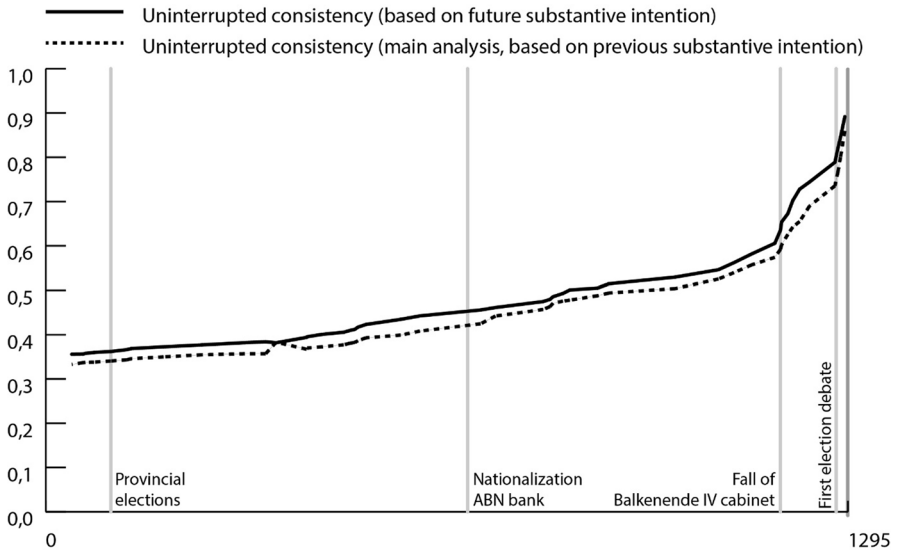

Fig. 9 Uninterrupted consistency by method of dealing with non-substantive preferences (2007–2010)

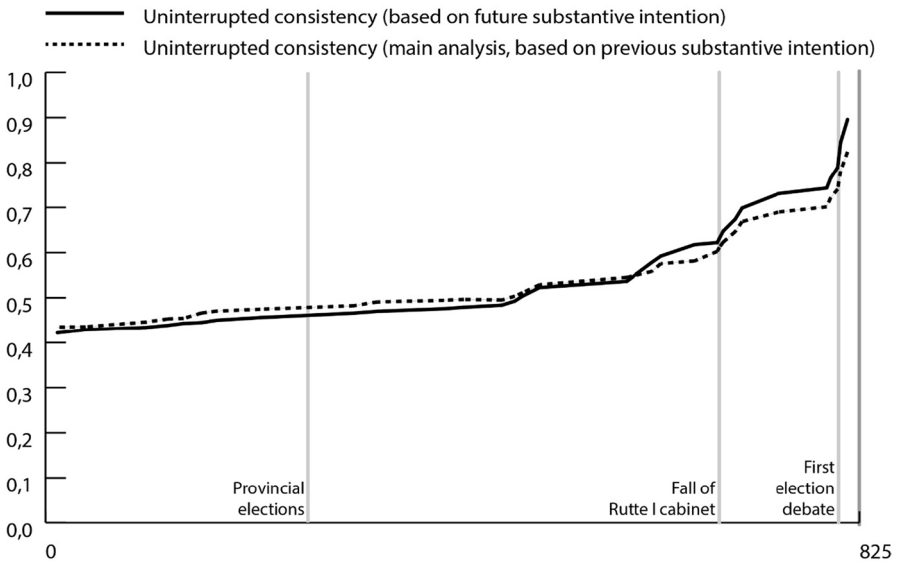

Fig. 10 Uninterrupted consistency by method of dealing with non-substantive preferences (2010–2012)

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1057/s41269-022-00250-x.

Declarations

Conflict of interest The authors hereby state that there is no conflict of interest—social, financial, or otherwise—with regards to this manuscript.

References

Aaldering, L., T.W.G. van der Meer, and W. Van der Brug. 2018. Mediated leader effects: The impact of newspapers' portrayal of party leadership on electoral support. *The International Journal of Press/Politics* 23 (1): 70–94.

Andeweg, R.B., ed. 1986. *Tussen steekproef en stembus. Beschouwingen over verkiezingspeilingen naar aanleiding van de discrepantie tussen de peilingen en de verkiezingsuitslag van 21 mei 1986*. Leiden: DSWO Press.

Box-Steffensmeier, J., M. Dillard, D. Kimball, and W. Massengill. 2015. The long and short of it: The unpredictability of late deciding voters. *Electoral Studies* 39: 181–194.

Brettschneider, F. 1997. The press and the polls in Germany, 1980–1994. Poll coverage as an essential part of election campaign reporting. *International Journal of Public Opinion Research* 9 (3): 248–265.

Caballe, A., P. Grima, and L. Marco-Almagro. 2013. Are Election Polls Right? Analysis of the Accuracy of Election Polls Predictions Published in the Press. *Revista Española De Investigaciones Sociológicas* 142: 24–46.

Campbell, A., P. Converse, W. Miller, and D. Stokes. 1960. *The American Voter*. New York: Wiley.

Converse, P., and M. Traugott. 1986. Assessing the accuracy of polls and surveys. *Science* 234 (4780): 1094–1098.

De Lange, S.L. 2013. Vinger aan de polls: Peilers, pers en politicologen over opiniepeilingen. *Beleid En Maatschappij* 40 (2): 168–170.

De Stefano, D., F. Pauli, and N. Torelli. 2022. Pre-electoral polls variability: a hierarchical bayesian model to assess the role of house effects with application to Italian elections. *The Annals of Applied Statistics* 16 (1): 460–476. https://doi.org/10.1214/21-AOAS1507.

Donsbach, W., and U. Hartung. 2008. The legal status of public opinion research in the world. In *The Sage handbook of public opinion research*, ed. W. Donsbach and M.W. Traugott, 431–440. London: Sage.

Evans, G., and J. Tilley. 2012. The depoliticization of inequality and redistribution: Explaining the decline of class voting. *The Journal of Politics* 74: 963–976.

Goot, M. 2021. How good are the polls? Australian election predictions, 1993–2019. *Australian Journal of Political Science* 56 (1): 35–55.

Irwin, G.A., and J.J.M. Van Holsteyn. 2002. According to the polls: The influence of opinion polls on expectations. *The Public Opinion Quarterly* 66 (1): 92–104.

Jackman, S. 2005. Pooling the polls over an election campaign. *Australian Journal of Political Science* 40 (4): 499–517.

Jaworski, A., and R. Fitzgerald. 2008. 'This poll has not happened yet': Temporal play in election predictions. *Discourse & Communication* 2 (1): 5–27.

Jennings, W., and Chr Wlezien. 2016. The timeline of elections: A comparative perspective. *American Journal of Political Science* 60 (1): 219–233.

Jennings, W., and Chr Wlezien. 2018. Election polling errors across time and space. *Nature Human Behaviour* 2 (4): 276–283.

Keeter, S., N. Hatley, C. Kennedy, and A. Lau. 2017. *What low response rates mean for telephone surveys*. Pew Research Center. http://www.pewresearch.org/2017/05/15/what-low-response-rates-mean-for-telephone-surveys/.

Kriesi, H., E. Grande, R. Lachat, M. Dolezal, S. Bornschier, and T. Frey. 2008. *West European politics in the age of globalisation*. Cambridge: Cambridge University Press.

Lachat, R. 2007. *A heterogeneous electorate: Political sophistication, predisposition strength, and the voting decision process*. Baden-Baden: Nomos Verlag.

Lewis-Beck, M.S., and R. Dassonneville. 2015. Forecasting elections in Europe: Synthetic models. *Research & Politics* 2 (1): 1–11.

Meffert, M.E., and Th. Gschwend. 2011. Polls, coalition signals and strategic voting: An experimental investigation of perceptions and effects. *European Journal of Political Research* 50: 636–667.

Nadeau, R., R. Dassonneville, M.S. Lewis-Beck, and Ph. Mongrain. 2020. Are election results more unpredictable? A forecasting test. *Political Science Research and Methods* 8: 764–771.

Oscarsson, H., and M. Rosema. 2019. Consideration set models of electoral choice: Theory, method, and application. *Electoral Studies* 57: 256–262.

Patterson, Th.E. 2005. Of polls, mountains: U.S. journalists and their use of election surveys. *Public Opinion Quarterly* 69 (5): 716–724.

Pellikaan, H., S.L. De Lange, and T.W.G. Van der Meer. 2007. Fortuyn's legacy: party system change in the Netherlands. *Comparative European Politics* 5 (3): 282–302.

Seymour, G., A. Crossley, P.F. Lazarsfeld, and G. Gallup. 1949. Should public opinion polls make election forecasts? *A Symposium. Journalism Quarterly* 26 (2): 131–144.

Toff, B. 2019. The 'Nate Silver effect' on political journalism: Gatecrashers, gatekeepers, and changing newsroom practices around coverage of public opinion polls. *Journalism* 20 (7): 873–889.

Traugott, M. 2014. Public opinion polls and election forecasting. *PS: Political Science & Politics* 47 (2): 342–344.

Van der Meer, T.W.G., A. Hakhverdian, and L. Aaldering. 2016. Off the fence, onto the bandwagon? A large-scale survey experiment on effect of real-life poll outcomes on subsequent vote intentions. *International Journal of Public Opinion Research* 28 (1): 46–72.

Van der Meer, T., R. Lubbe, E. Van Elsas, M. Elff, and W. Van der Brug. 2012. Bounded volatility in the Dutch electoral battlefield: A panel study on the structure of changing vote intentions in the Netherlands during 2006–2010. *Acta Politica* 47 (4): 333–355.

Van der Meer, T.W.G., E. Van Elsas, R. Lubbe, and W. Van der Brug. 2015. Are volatile voters erratic, whimsical or seriously picky? A panel study of 58 waves into the nature of electoral volatility (The Netherlands 2006–2010). *Party Politics* 21 (1): 100–114.

Weimann, G. 1990. The obsession to forecast: pre-election polls in the Israeli Press. *The Public Opinion Quarterly* 54 (3): 396–408.

Zerback, Th., Th. Koch, and B. Krämer. 2015. Thinking of others: Effects of implicit and explicit media cues on climate of opinion perceptions. *Journalism & Mass Communication Quarterly* 92 (2): 421–443.