**Predicting the 2018 Hungarian General Elections**

**Tactical Voting Summary and Election Predictions using UNS and GLM**

**1 Election system and tactical voting overview**

In 2012, a new election system has been introduced in Hungary. In the new system 106 seats are distributed through single seat constituencies (OEVK) and 93 seats are proportionally shared based on party list votes. In addition, a new “winner bonus” system has been added to this system where the winning candidates in OEVKs receive carry some votes to the party list where the seats are distributed between parties. This new bonus means that those parties that win the majority of seats will have a stronger majority than in the former, more proportional election system.

Because of this change in the system, opposition parties had to coordinate their efforts in seats to win as many seats as possible. This coordination was quite limited with only 43 candidates withdrawing from the race in 29 seats. Withdrawals mainly happened in key districts in Budapest, but even there the coordination was quite limited between parties. To increase the number of opposition seats, voters had to opt for tactical voting in many districts where at least two (but often more) opposition candidates stood against the governing party, Fidesz.

Since voters had to choose from many potential opposition candidates, taktikaiszavazas.hu (meaning “tactical voting”) was launched with the list of constituencies and candidates in all 106 districts with suggestions on who the strongest candidate is in each district. Similar sites have also been created such as kireszavazzunk.hu (“who should we vote for?”) where small sample (500 respondents) telephone polling was carried out in all districts. This site also provided recommendation on the strongest candidates.

The election predictions on taktikaiszavazas.hu used a mixture of constituency-level polling in 20 OEVKs carried out by Medián and Závecz Research, two polling agencies, and the 2014 election data on which a uniform national swing (UNS) model was used. The polling was ordered by Közös Ország Mozgalom (Common Country Movement or KOM), an NGO group that facilitated political discussion between parties, and they also tried to increase voter turnout among opposition voters. Using the constituency polls and UNS, it was possible to identify 40-45 districts (about 40% of all seats) where opposition parties had some chance to win the seat by either standing head to head (with the strongest candidate) against Fidesz, or by encouraging voters to vote tactically for a certain candidate.

In the following sections, I will explain the following aspects of this election prediction project:

* Different data sources and solutions used to run the website and create the predicitions
* Prediction models
* Election results and the deviations from predictions
* Conclusions and further research potential

**2 Technical solutions and data sources**

**2.1 Running the taktikaiszavas.hu website**

The main website, taktikaiszavazas.hu, was set up as a static website running as GitHub page. The reason to use this approach was to reduce the impact of DDos (denial of service) attacks. Each visitor was also redirected to Cloudflare to further mitigate the overload attacks potential.

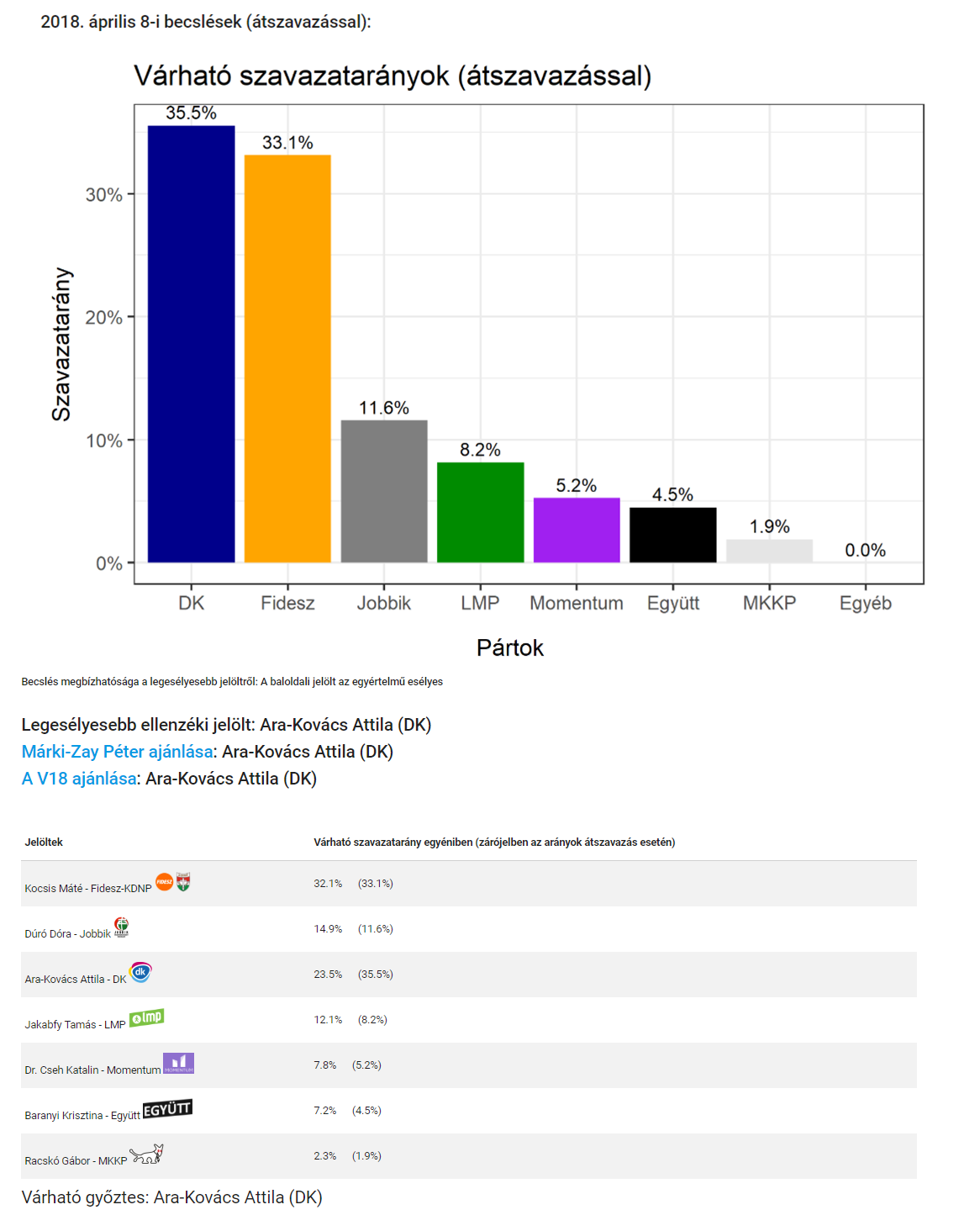
The website is divided into different sections most of them focusing on single seat races.

The different tables and graphics have been predefined in the page template that was fed from two different sources.

The table are pulled from a csv file that with a Perl script. The script automatically amends values and generates new single static pages (one for each single seat race). It was relatively easy to update the website every week when new, adjusted predictions were published.

The graphics showing the graphical representation of the expected vote shares in the table were generated in ggplot. They also corresponded to the auto-generated webpage names so all of them could be inserted to the website without any manual editing.

The other parts of the website were also partially auto-generated through the Perl script such as the list of all expected results by constituency in a summary page. Overall, managing the website required some effort when images such as the constituency maps had to be updated, but the calculated values and related visuals were mostly automated.



**Typical view of a webpage showing expected vote shares in one constituency**

**2.2 Data sources used for predictions**

To create predictions for the 2018 election, data from the previous 2014 election was used. The data was shared by Gábor Tóka, who was an advisor and analyst for KOM during the constituency-level polling. Also, I combined this data with different demographics data acquired from the Hungarian Central Statistical Office (KSH). The 2018 election data was also downloaded and shared online in the coming days after the election. Fortunately, all data sets were in .csv format, so the data cleansing required less effort. The main issue was with the encoding of constituency and town names as some of them were not read correctly in utf-8 format. Both in R and Excel, I created a set of golden data sources, where all the fields are formatted correctly and can be used for analysis. These files are the following:

**Initial sources (csv)**

vote\_counts\_precincts\_2b (2014 data)

2018\_egyeni\_listas\_29May (2018 data)

6 additional files from KSH data (tstar data)

**Generated and cleaned output files (csv)**

2018\_UNS\_baseline (used as the prediction base)

2018\_by\_oevk\_clean (2018 results for each OEVK)

oevk2014\_clean\_new (2014 results for each OEVK)

oevk2014\_telepules (2014 results by town)

2014\_2018\_tstar\_all\_by\_telep (2014-2018 results combined with KSH data by town)

2018\_by\_telep (2018 results by town)

2018\_by\_szavazokor (2018 results by precincts)

2018\_egyeni\_becsles\_elteres (Prediction error between pre-election predictions vs. 2018 results)

Some of the manual changes in these files included adding some additional data about OEVKs such as the typical profile of the district (based on the types and numbers of town making up that voting area). These extra metrics were used during the pre-election prediction (for example, by calculating with higher than expected number of votes cast on smaller parties in major cities).

In addition to the 2014 data and other population data from KSH, the summary of 20 constituency-level polls were also used for predictions together with national polls conducted by different polling agencies between January-March 2018. The constituency level polls included the summary of potential vote shares of each party, the vote shares of each candidate and the expected tactical voting behavior of people. The last of these aspects was very important to create a simple but plausible approach to tactical voting, which was hard to predict with any prior pattern.

**Prediction model – UNS with a twist**

UNS with candidate strength added

Small parties – taken from LMP voter base

**Tactical voting predictions**

Different polls on tactical voting behavior

Final prediction’s model used

**Election results and prediction errors**

Summary of results

Highlight key districts with errors

Explain errors based on UNS results and GLM and T-star data

**Conclusion**

Tactical voting results

Further research (voter fraud)