Brief Election Forensics Report on the 2015 Legislative Elections in Turkey

Walter R. Mebane, Jr.*

January 20, 2016

*Professor, Department of Political Science and Department of Statistics, University of Michigan, Haven Hall, Ann Arbor, MI 48109-1045 (E-mail: wmebane@umich.edu).

1 Introduction

I briefly summarize the most telling diagnostics regarding anomalies and potential frauds in the legislative elections in Turkey during 2015. To analyze the June and November elections I use polling station data.[1] Methods used with the polling station data to analyze each election separately are mostly methods implemented at the prototype Election Forensics Toolkit website (Hicken and Mebane 2015; Mebane 2015*b*).[2] These methods include tests based on the digits in turnout and vote counts and in turnout and vote percentages, tests of unimodality and estimates of positive empirical models of election frauds. I examine patterns of geographic clustering in the probability of "fraud" estimated for each polling station. To check how voting patterns in the June elections relate to patterns in the November elections, I aggregate the polling station data to produce neighborhood observations: to see whether the distribution of votes among parties in the November election is like the distribution in the June election I use a robust overdispersed multinomial regression model (Mebane and Sekhon 2004*a*).

The election rules in Turkey mean it is meaningful to compute various statistics separately using the polling stations in each of the 85 districts. The rules specify a national minimum threshold of ten percent for gaining seats but allocate seats by district using a closed-list proportional representation (D'Hondt) system (Álvarez-Rivera 2015; Turkish Press 2010; Yüksek Seçim Kurulu 2015).

1.1 Digit and Modality Tests

We consider results from using the Election Forensics Toolkit website to estimate four kinds of statistics separately in each of district. Second-digit mean ("2BL") and last-digit mean ("LastC") statistics are applied to the counts themselves, while the indicator variable

[1] Data from the Computer Supported Central Voter Registry System SEÇSiS Project at `https://sonuc.ysk.gov.tr/module/GirisEkrani.jsf` were made available by Rob Barry and Tom McGinty of the Wall Street Journal on November 19, 2015.

[2] Election Forensics Toolkit website URL `http://electionforensics.ddns.net:3838/EFT_USAID/`.

Table 1: Leading Party in Each District, Turkey 2015, June

Party	Districts
AKP	Adana, Adiyaman, Afyonkarahisar, Aksaray, Amasya, Ankara_I, Ankara_II, Antalya, Artvin, Balikesir, Bartin, Bayburt, Bilecik, Bingöl, Bolu, Burdur, Bursa, Çankiri, Çorum, Denizli, Düzce, Elazig, Erzincan, Erzurum, Gaziantep, Giresun, Gümüshane, Hatay, Isparta, Istanbul_I, Istanbul_II, Istanbul_III, Kahramanmaras, Karabük, Karaman, Kastamonu, Kayseri, Kilis, Kirikkale, Kirsehir, Kocaeli, Konya, Kütahya, Malatya, Manisa, Nevsehir, Nigde, Ordu, Rize, Sakarya, Samsun, Sanliurfa, Sinop, Sivas, Tokat, Trabzon, Usak, Yalova, Yozgat
HDP	Agri, Ardahan, Batman, Bitlis, Diyarbakir, Hakkari, Igdir, Kars, Mardin, Mus, Siirt, Sirnak, Tunceli, Van
CHP	Aydin, Çanakkale, Edirne, Eskisehir, Izmir_I, Izmir_II, Kirklareli, Mersin, Mugla, Tekirdag, Zonguldak
MHP	Osmaniye

Note: party with the most votes in each district
AKP, Justice and Development Party; CHP, Republican People's Party; HDP, Peoples' Democratic Party; MHP, Nationalist Movement Party.

mean ("P05s")[3] statistic and unimodality hypothesis test ("DipT") are based on percentages: the percentage of eligible voters voting for "Turnout," and the percentage of valid votes cast for each party for the party variables. Statistics are computed for turnout and for the leading party (the party with the most votes) in each district. Tables 1 and 2 identify the leading party in each district in June and in November. Points shown in red differ significantly from the values some have asserted should occur in the absence of frauds or anomalies, while points in blue do not.

As shown in Figures 1 and 2, both the June and November elections have many suspicious features. Several statistics for turnout suggest that frauds occurred. In both November and June, P05s is significantly greater than .2 in many districts. Values of P05s that are significantly elevated above P05s = .2 suggest that turnout was being manipulated by "agents" who wished for their manipulations to be detectable (Beber and Scacco 2012; Kalinin and Mebane 2011; Rundlett and Svolik 2015). Who these "agents" are and how

[3] P05s is the mean of a binary variable that is one if the last digit of the rounded percentage of votes for the referent party or candidate is zero or five.

Table 2: Leading Party in Each District, Turkey 2015, November

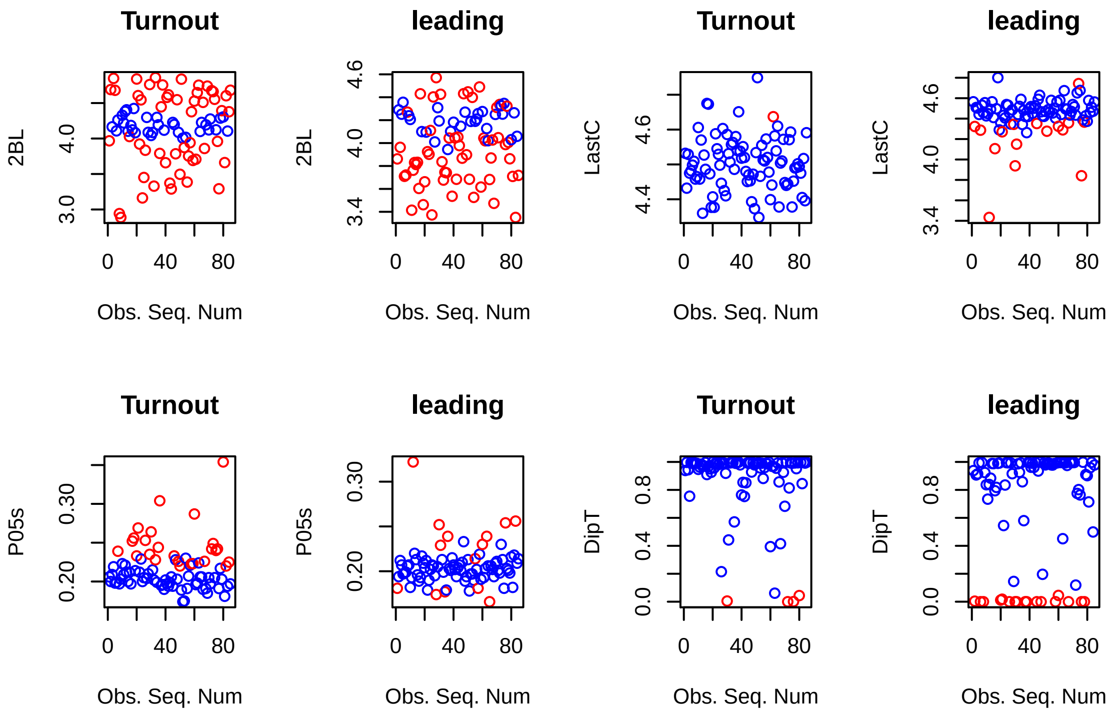
Party	Districts
AKP	Adana, Adiyaman, Afyonkarahisar, Aksaray, Amasya, Çanakkale, Ankara_I, Ankara_II, Antalya, Ardahan, Artvin, Balikesir, Bartin, Bayburt, Bilecik, Bingöl, Bolu, Burdur, Bursa, Çankiri, Çorum, Denizli, Düzce, Elazig, Erzincan, Erzurum, Eskisehir, Gaziantep, Giresun, Gümüshane, Hatay, Isparta, Istanbul_I, Istanbul_II, Istanbul_III, Kahramanmaras, Karabük, Karaman, Kars, Kastamonu, Kayseri, Kilis, Kirikkale, Kirsehir, Kocaeli, Konya, Kütahya, Malatya, Manisa, Mersin, Nevsehir, Nigde, Ordu, Osmaniye, Rize, Sakarya, Samsun, Sanliurfa, Sinop, Sivas, Tokat, Trabzon, Usak, Yalova, Yozgat, Zonguldak
HDP	Agri, Batman, Bitlis, Diyarbakir, Hakkari, Igdir, Mardin, Mus, Siirt, Sirnak, Tunceli, Van
CHP	Aydin, Edirne, Izmir_I, Izmir_II, Kirklareli, Mugla, Tekirdag

Note: party with the most votes in each district
AKP, Justice and Development Party; CHP, Republican People's Party; HDP, Peoples' Democratic Party.

exactly they manipulated turnout is not clear, nor is it clear which party may have benefited from or been harmed by their actions. According to the DipT p-value (Hartigan and Hartigan 1985), turnout percentages have a multimodal distribution within a few districts in both elections. Turnout percentage distributions that are not unimodal are generally considered to be suspicious (Myagkov, Ordeshook and Shaikin 2009). The P05s statistics for the leading party suggest frauds are likely in several districts. P05s for the leading party is significantly greater than .2 in several districts in both elections.

The patterns observed for the leading parties' second significant digit means (2BL) also suggest that frauds occur (Pericchi and Torres 2011; Deckert, Myagkov and Ordeshook 2011; Mebane 2011). With multiple parties the leading party should have 2BL < 4 if that party gains from some voters favoring that party for strategic reasons (Mebane 2013, n.d.). If strategic voting is not helping the leading party, then in a multiparty contest the leading party should have 2BL no greater than about 4.25. But in several districts 2BL > 4.4. Seeing 2BL values both greater than 4.187 and less than 4.187 is suspicious.

Figure 1: Distribution and Digit Tests by District, Turkey 2015, June



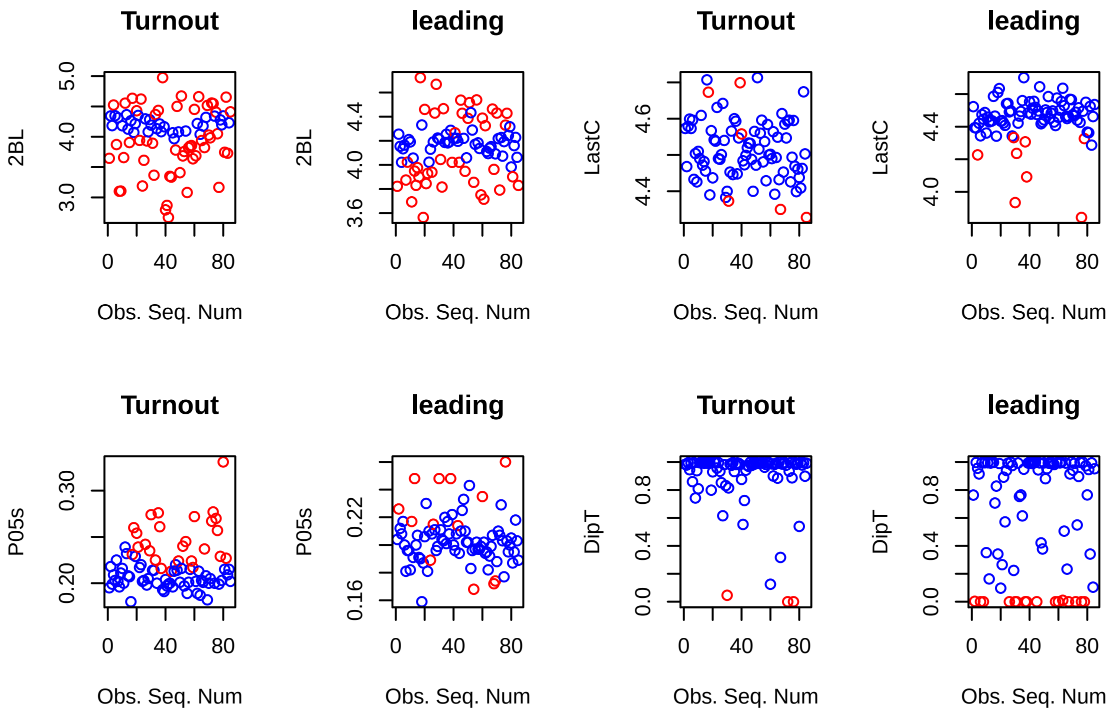
Note: statistics and tests based on polling station observations. "2BL," second-digit mean; "LastC," last-digit mean; "P05s," mean of variable indicating whether the last digit of the rounded percentage of votes for the referent party or candidate is zero or five.

1.2 Robust Overdispersed Multinomial Regression Models

I use robust overdispersed multinomial regression models (Mebane and Sekhon 2004*a*,*b*) to see whether the distribution of votes among parties in one election is like the distribution in a previous election and to identify "outlier" observations for which the relationship between elections is not like the majority of observations (Wand, Shotts, Sekhon, Mebane, Herron and Brady 2001; Mebane and Herron 2005; Mebane 2010). This method requires that particular vote tabulation units be matched across the different elections. In Turkey it is feasible to aggregate the polling station data into neighborhoods and then to match "neighborhoods" across the June and November elections.[4] The robust overdispersed

[4]In all there are 50,038 neighborhoods and 49,997 neighborhoods with complete data.

Figure 2: Distribution and Digit Tests by District, Turkey 2015, November



Note: statistics and tests based on polling station observations. "2BL," second-digit mean; "LastC," last-digit mean; "P05s," mean of variable indicating whether the last digit of the rounded percentage of votes for the referent party or candidate is zero or five.

multinomial regression model for the November neighborhood vote counts in each district uses linear predictors that are functions of regressor variables p_{ij}, $j \in \{\text{AKP}, \text{CHP}, \text{MHP}, \text{HDP}, \text{Other}\}$, which are the proportions of the vote received in each neighborhood i by each party in the June election:

$$\mu_{i\mathrm{AKP}} = \beta_{10} + \beta_{11} p_{i\mathrm{AKP}} + \beta_{12} p_{i\mathrm{CHP}} + \beta_{13} p_{i\mathrm{MHP}} + \beta_{14} p_{i\mathrm{HDP}} + \beta_{15} p_{i\mathrm{Other}} \tag{1a}$$

$$\mu_{i\mathrm{MHP}} = \beta_{20} + \beta_{21} p_{i\mathrm{AKP}} + \beta_{22} p_{i\mathrm{CHP}} + \beta_{23} p_{i\mathrm{MHP}} + \beta_{24} p_{i\mathrm{HDP}} + \beta_{25} p_{i\mathrm{Other}} \tag{1b}$$

$$\mu_{i\mathrm{HDP}} = \beta_{30} + \beta_{31} p_{i\mathrm{AKP}} + \beta_{32} p_{i\mathrm{CHP}} + \beta_{33} p_{i\mathrm{MHP}} + \beta_{34} p_{i\mathrm{HDP}} + \beta_{35} p_{i\mathrm{Other}} \tag{1c}$$

$$\mu_{i\mathrm{Other}} = \beta_{40} + \beta_{41} p_{i\mathrm{AKP}} + \beta_{42} p_{i\mathrm{CHP}} + \beta_{43} p_{i\mathrm{MHP}} + \beta_{44} p_{i\mathrm{HDP}} + \beta_{45} p_{i\mathrm{Other}} \tag{1d}$$

$$\mu_{i\mathrm{Nonvote}} = \beta_{50} + \beta_{51} p_{i\mathrm{AKP}} + \beta_{52} p_{i\mathrm{CHP}} + \beta_{53} p_{i\mathrm{MHP}} + \beta_{54} p_{i\mathrm{HDP}} + \beta_{55} p_{i\mathrm{Other}} \tag{1e}$$

$$\mu_{i\mathrm{CHP}} = 0 \tag{1f}$$

The β_{kj}, $k = 1, \ldots, 5$, $j = 0, \ldots, 5$, are constant coefficients estimated independently for each district.

Two criteria determine whether the model flags votes in a neighborhood as potentially problematic: (1) whether coefficients that measure the relationship between the support a party received in June and the support the party received in November have the wrong sign; and (2) whether one or more of the vote counts recorded in the neighborhood in November is an outlier. On coefficient signs, I expect the coefficients that relate the June vote proportion of the votes for a party to the votes for the same party in November to be positive. In terms of the notation in the linear predictor equations (1a)–(1f), I expect $\beta_{11} > 0$, $\beta_{23} > 0$ and $\beta_{34} > 0$. In addition, because CHP is the principal party opposing AKP as well as the reference party for the linear predictor, I expect $\beta_{12} < 0$. On outliers, for each neighborhood there are five orthogonalized residuals that may be outliers.[5] I estimate a separate model in each district.[6]

Table 3 reports some details for the districts in which either one of the four key coefficients has the wrong sign, or more than five percent of the neighborhoods are outliers, or both. The coefficients for HDP (β_{34}) always have the correct sign, but incorrect signs

[5] In terms of the notation in Mebane and Sekhon (2004*a*, 408), the orthogonalized residuals for neighborhood observation i are denoted $\hat{r}^*_{ij}$.

[6] Efforts to estimate the model failed in seven districts: Afyonkarahisar, Bayburt, Bolu, Düzce, Kilis, Rize and Sakarya.

are sometimes observed for the other three key coefficients. Table 3 indicates which of these coefficients have the wrong sign in which districts. In every district where at least five percent of the orthogonalized residuals are outliers, at least five percent of the first orthogonalized residuals in the district are outliers. Table 3 therefore reports the proportion of the first orthogonalized residuals that are outliers. The table also reports the number of neighborhood observations in each of the districts that have a suspicious profile, along with the identity of the party receiving the most votes in the district.

Districts in which the June and November voting patterns do not much resemble one another according to the criteria used to make Table 3 are almost all districts in eastern Turkey. The map of regions in Turkey in Figure 3 shows that almost all regions east of a line traced through Samsun, Tokat, Sivas, Kahramanmaras and Sanliurfa (or Kilis, if districts for which estimation failed are included) exhibit substantial tendencies for the November vote counts not to resemble the patterns that the June election results might have led one to expect. Almost all the districts that exhibit anomalous November voting patterns according to Table 3 are east of that rough line. As reported in Table 3, AKP is the leading party in 22 of the districts listed in Table 3, while HDP is the leading party in the other nine districts.[7]

Differences between June and November voting patterns do not of themselves imply that electoral frauds occurred in November. It is unclear whether the June election patterns represent typical patterns one should treat as baselines for the November results. But we will see that detailed comparisons of positive empirical "frauds" models and of turnout and vote proportion distributions between June and November (Figures 8–21) will strongly suggest that more fraudulent activity occurred in November.

[7] For context, across the entire country in November AKP is the leading party in 66 districts, HDP leads in 12 districts and CHP leads in seven districts. See Table 2.

Figure 3: Map of Regions in Turkey

Image source: `http://www.invest.gov.tr/en-US/investmentguide/investorsguide/Pages/Incentives.aspx` (obtained December 2, 2015).

Table 3: Incorrect Signs and Outliers in Robust Regression Models for November Votes

	wrong signs[a]					
District	AKP	MHP	CHP	outliers[b]	N[c]	leader[d]
Adiyaman			*	.90	615	AKP
Agri				.08	660	HDP
Aksaray	*			.96	329	AKP
Bitlis				.08	462	HDP
Diyarbakir				.06	1039	HDP
Elazig	*			.92	699	AKP
Erzincan	*			.59	674	AKP
Erzurum				.83	1184	AKP
Giresun	*			.99	744	AKP
Gümüshane	*	*		.96	392	AKP
Hakkari				.15	171	HDP
Kahramanmaras	*			.84	682	AKP
Karaman	*			.97	289	AKP
Kars	*			.02	436	AKP
Kirikkale	*			.96	270	AKP
Konya	*			.99	1188	AKP
Kütahya	*			.99	767	AKP
Malatya	*			.81	716	AKP
Mardin				.11	693	HDP
Mus	*			.43	473	HDP
Ordu	*			.99	725	AKP
Samsun	*			.98	1244	AKP
Sanliurfa			*	.93	1365	AKP
Siirt				.05	339	HDP
Sivas	*			.69	1485	AKP
Tokat	*			.79	930	AKP
Trabzon	*		*	.99	688	AKP
Tunceli				.17	399	HDP
Van	*			.12	683	HDP
Çankiri	*		*	.98	466	AKP
Çorum	*			.80	883	AKP

Note: Districts shown are ones in which either one of four key coefficients has the wrong sign, or more than five percent of the neighborhoods include outliers, or both. [a] * the November vote count variable for AKP or for MHP has a negative coefficient for the June vote proportion variable for the same party, or the November vote count variable for CHP has a positive coefficient for the June vote proportion variable for AKP. [b] proportion of the first orthogonalized residuals that are outliers. [c] number of neighborhood observations. [d] party with the most votes in each district.

1.3 Positive Empirical Models of Election Frauds

The positive empirical model of election frauds introduced by Klimek, Yegorov, Hanel and Thurner (2012) produces substantial evidence of fraudulent activity. According to Klimek et al.'s idea, frauds always benefit only one party, which they designate to be the party that has the most votes—the leading party. In fact any party can be declared the "leading party," although it's always true with their concept that only one party can benefit from frauds. For the November election I estimate frauds probabilities produced using a finite mixture model implementation of Klimek et al.'s idea (Mebane 2015*a*), estimating the model separately in each district. The statistical methodology includes a likelihood ratio test of the hypothesis that there are no "frauds" (Mebane and Wall 2015).

1.3.1 Party with the most votes in each district is the leading party

Results from using the party with the most votes in each district as the leading party in that district are displayed in Figure 4.[8] In Figure 4, parameter values for districts in the November election in which the likelihood ratio test rejects the "no-frauds" hypothesis appear in red. Otherwise points are blue. Significant "frauds" occur in most districts.[9] In most districts the "frauds" that materially affect votes are "incremental frauds." The "incremental fraud" probabilities ($f_{\rm i}$) are mostly small, but some are greater than .01.[10] If $f_{\rm i} = .01$, then one expects "incremental fraud" to affect one percent of the polling stations in the referent district.

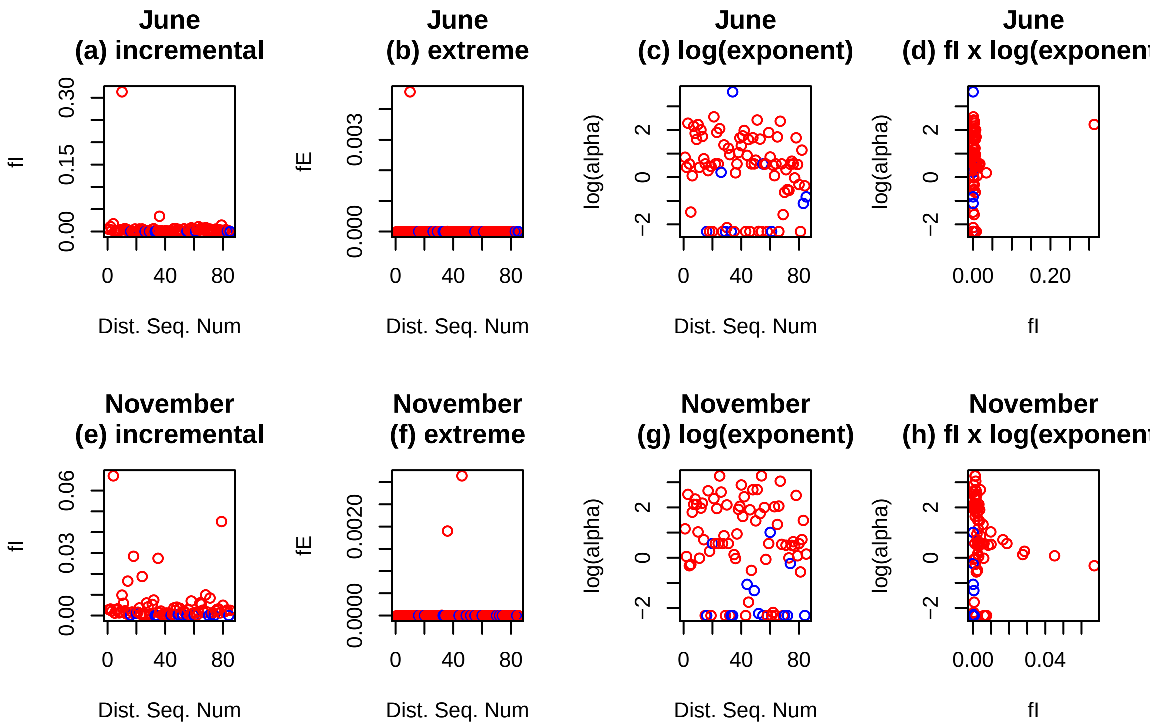
With "incremental fraud," some votes that should have gone to other parties instead go to the leading party, and turnout is inflated by some nonvotes being counted as votes for the leading party. Let N_i be the observed number of electors, let τ_i be the proportion of

[8] Table 4 in the Appendix lists the values of $f_{\rm i}$ and $f_{\rm e}$ estimated for each district.

[9] The test does not discriminate which of the two types of frauds in the model is producing the significant test result, so even when the "extreme fraud" probability ($f_{\rm e}$) estimate is zero so that no "extreme fraud" occurs, this parameter may be shown in red. Only one district in June and two districts in November have $f_{\rm e} > 0$.

[10] The "incremental fraud" probability in June that has value $f_{\rm i} = .31$ in Figures 4(a–d) occurs in Ardahan district. The largest value $f_{\rm i} = .067$ that occurs in November (Figures 4(e–h)) is for Agri district.

Figure 4: Finite Mixture Likelihood Model by District, Turkey 2015



Note: statistics and tests based on polling station observations, using the party with the most votes in each district as the leading party in that district. Parameters of the finite mixture model of Mebane (2015*a*): `fI`, incremental fraud probability (f_{i}); `fE`, extreme fraud probability (f_{e}); `log(exponent)`, log exponent ($\log\alpha$). Parameter values for districts in which the "no-frauds" hypothesis can be rejected based on a likelihood ratio test (Mebane and Wall 2015) appear in red. Otherwise points are blue.

electors who turn out to vote in the absence of frauds, and let ν_i be the proportion of votes the leading party receives in the absence of frauds. With no frauds the leading party is expected to receive $N_i\tau_i\nu_i$ votes in polling station i, but with "incremental fraud" the leading party is expected to receive the following number of votes:

$$N_i\left(\tau_i\nu_i + x_i\left(1-\tau_i\right) + x_i^{\alpha}\left(1-\nu_i\right)\tau_i\right) \qquad (2)$$

where x_i is the proportion of genuine nonvotes that are counted as votes for the leading party and x_i^α is the proportion of votes that were genuinely cast for other parties but instead are counted as votes for the leading party. τ_i, ν_i and x_i are all unobserved random variables and $\alpha > 0$ is a constant to be estimated. τ_i and ν_i are truncated Normal variables and x_i is a truncated half-Normal variable. See Mebane (2015*a*) for further details.

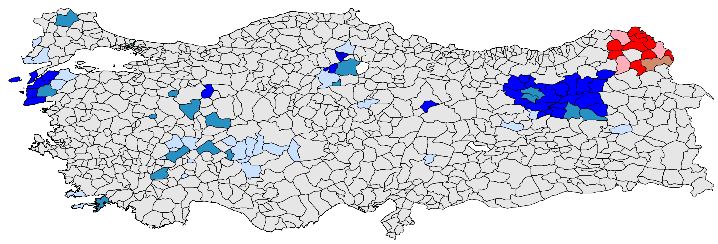
Whether the "fraud" involves more vote stealing or more vote manufacturing is measured by the exponent α. If $\log(\alpha) = 0$ then both processes are equally affecting votes. If $\log(\alpha) < 0$ then $x_i < x_i^\alpha$ and vote stealing is more important, and if $\log(\alpha) > 0$ then $x_i > x_i^\alpha$ and manufacturing votes from nonvoters is more important. Figures 4(c,g) show that $\log(\alpha) > 0$ more often than $\log(\alpha) < 0$, so manufacturing votes from nonvoters is more often the kind of "fraud" the model suggests occurs. In both June and November the vote-manufacturing variety of "incremental fraud" is more important in the districts where $f_{\mathrm{i}} > .01$ (Figures 4(d,h)), but in November there is more vote-stealing "incremental fraud" than in June.

Figure 5 displays estimates of geographic "hotspots" at which clusters of incremental frauds probabilities estimated for each polling station are either significantly higher or significantly lower than the probabilities that occur generally. Because we lack information about the geographic location of each polling station but do have the location of each town (Hijmans 2015), Figure 5 shows hotspot analysis of the mean fraud probability in each town.[11] Red polygons in Figure 5 indicate towns that have high incremental fraud probabilities and blue polygons indicate towns that have low incremental fraud probabilities.

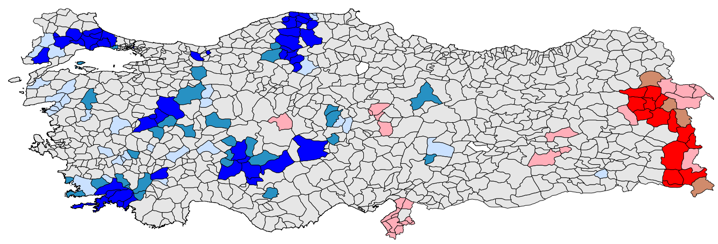
In Figure 5 more hotspots with incremental frauds probabilities greater than the prevailing average occur in November than in June. In June the clusters of high frauds probabilities occur in northeast Turkey, while in November the clusters occur mostly in

[11] Hotspot analysis uses the Getis-Ord G_i statistic (Getis and Ord 1992; Ord and Getis 1995). This measures whether the mean of values geographically close to observation i differs from the overall mean. To test whether each G_i value is significantly larger or smaller than would be expected by chance, we use permutation test methods to estimate p-values. Details are specified in Mebane (2015*b*). The p-values are corrected for multiple testing using false discovery rate procedures (Benjamini and Hochberg 1995) for the test levels α shown in Figure 22 in the Appendix ($\alpha = .01$, $\alpha = .05$ and $\alpha = .1$).

Figure 5: Hotspot Analysis, Fraud Probabilities by District, Turkey 2015



(a) June



(b) November

Note: Using the party with the most votes in each district as the leading party in all districts. Polling station incremental fraud probability, town average hotspot analysis using Getis-Ord G_i. Each polygon corresponds to a town. Overall averages: June, $\hat{f}_{\mathrm{i}} = 0.00594$; November, $\hat{f}_{\mathrm{i}} = 0.00572$.

eastern Turkey.

1.3.2 AKP is always treated as the leading party

More insight into where fraudulent activites likely occurred in the elections is produced by treating AKP as the leading party in all districts, regardless of whether AKP has the most votes in each particular district. We'll show this first by considering estimates of the finite mixture model and then by directly inspecting some scatterplots that display the joint distribution of turnout and vote proportions.

Figure 6 shows results from estimating the finite mixture model when AKP is always treated as the leading party. Significant "frauds" occur in most districts, and "incremental fraud" is the predominant type of "fraud."[12]

In Figure 6 it is clear that "frauds" that benefit AKP are about equally prevalent in both the June and November elections. Figures 6(c,g) show that in November more of the frauds involve more vote stealing than vote manufacturing: $\log(\alpha) < 0$ somewhat more often in November than in June. In the districts where $f_{\rm i}$ is largest—where the probabilities of "incremental fraud" are greatest—vote manufacturing tends to be more prevalent than vote stealing (see Figures 6(d,h)).

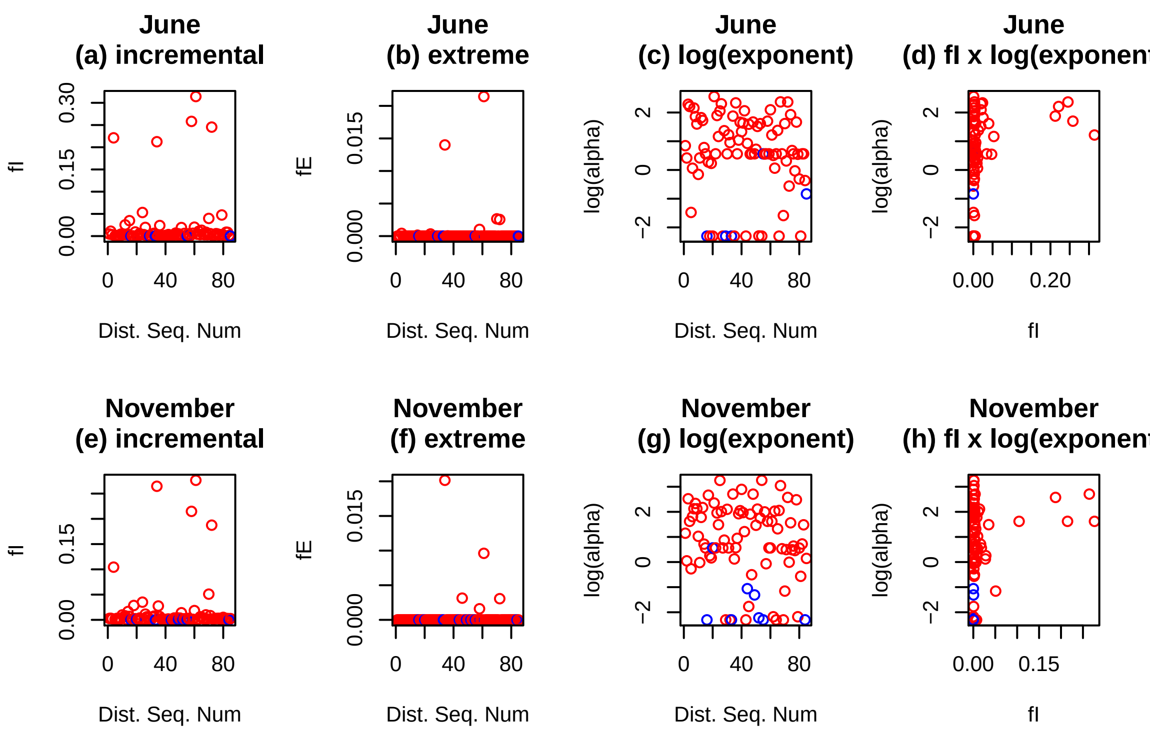
The greater prevalence of patterns that have the appearance of vote-stealing in November can be seen by examining scatterplots of the turnout proportions by the AKP-vote proportions for each polling station. Consider Figure 8, for example, which displays such scatterplots for the two elections in the Mus district, along with the unidimensional empirical densities for each kind of proportion. In the areas in Figures 8(c,f) where both turnout and AKP-vote proportions are greater than 0.8, more polling station observations are present in November than in June. Figure 8 displays distributional results for Mus, but similar patterns can be seen also in Hakkari, Mardin, Sirnak, Agri, Siirt, Diyarbakir, Bingöl, Van, Batman, Igdir, Adiyaman, Hatay and Tunceli (see Figures 9–21). All those districts are in eastern Turkey, all except Tunceli have $f_{\rm i} > .01$ in November, and all except Bingöl, Hatay and Tunceli have $f_{\rm i} > .01$ in June.[13]

When AKP is always treated as the leading party in the finite mixture model, frauds appear to be clustered in different places than when the party with the most votes in each district is treated as the leading party. Like Figure 5, Figure 7 shows that hotspots with incremental frauds probabilities greater than the prevailing average occur in eastern Turkey, but in Figure 7 both in June and November all the clusters of high frauds

[12] Only five districts have $f_{\rm e} > 0$, and three of those values are small.

[13] The finite mixture model cannot be estimated with AKP treated as the leading party in Tunceli because the estimation routine has been unable to find valid starting values because the median vote proportion is too low: the parameter ν must be less than that median (Mebane 2015*a*, 7). In June $f_{\rm i} = .0001$ in Bingöl and $f_{\rm i} = .002$ in Hatay.

Figure 6: Finite Mixture Likelihood Model by District, Turkey 2015



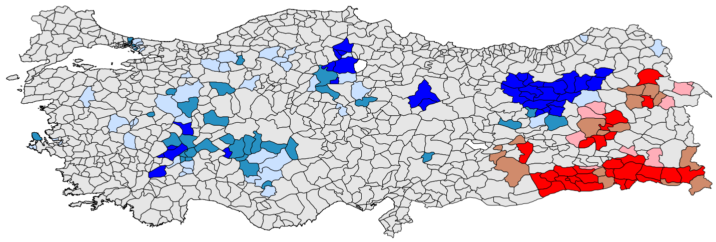
Note: statistics and tests based on polling station observations, using AKP as the leading party in all districts. Parameters of the finite mixture model of Mebane (2015*a*): `fI`, incremental fraud probability ($f_{\rm i}$); `fE`, extreme fraud probability ($f_{\rm e}$); `log(exponent)`, log exponent ($\log\alpha$). Parameter values for districts in which the "no-frauds" hypothesis can be rejected based on a likelihood ratio test (Mebane and Wall 2015) appear in red. Otherwise points are blue.

probabilities occur in southeast Turkey. For the most part the same towns are implicated in those hotspots in both elections.

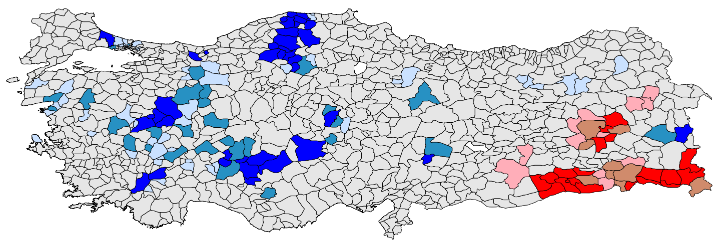
Bimodality that is apparent in turnout (Figures 8(b)–21(b)) goes with apparent multimodality in scatterplots of vote proportion by turnout (Figures 8(c)–21(c)), where there is a mode in which high turnout goes with a high vote proportion for AKP. A few other districts that are not in eastern Turkey also have $f_{\rm i} > .01$, but only a couple of those have patterns that roughly match those observed in the eastern Turkey districts.[14]

[14]When the finite mixture model is estimated with AKP treated as the leading party, in June Aydin, Osmaniye and Nigde have $f_{\rm i} > .01$ and in November Mugla, Bartin, Kirklareli and Edirne have $f_{\rm i} > .01$. Only

Figure 7: Hotspot Analysis, AKP-favoring Fraud Probabilities by District, Turkey 2015



(a) June



(b) November

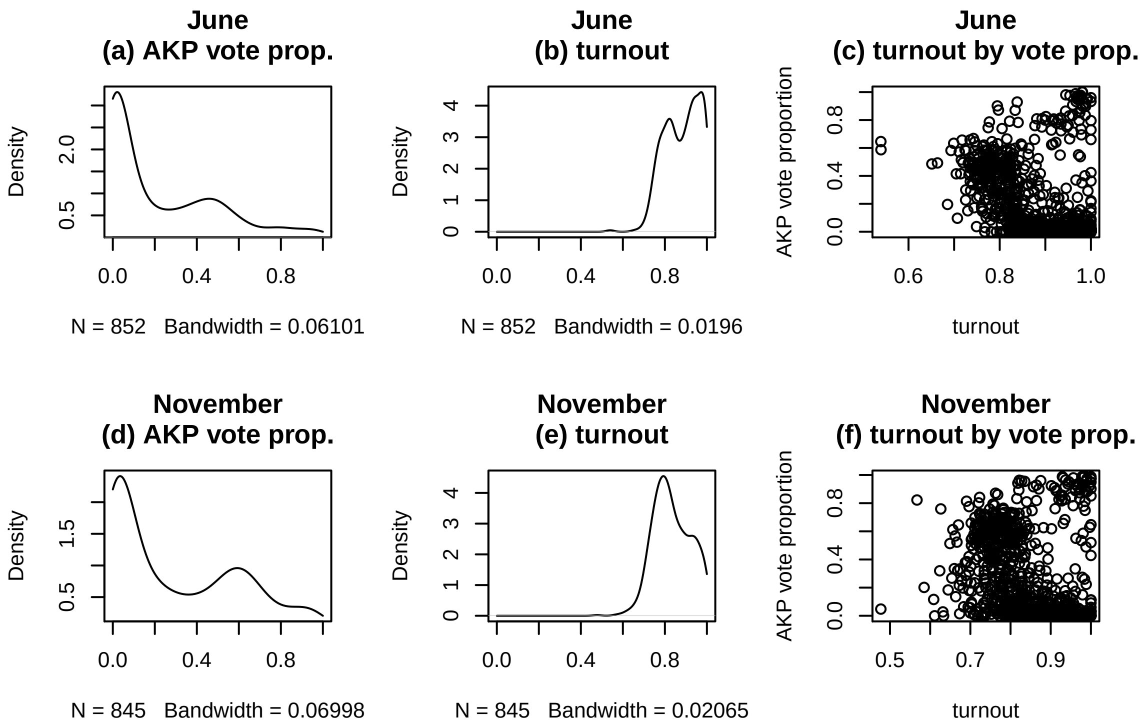
Note: Using AKP as the leading party in all districts. Polling station incremental fraud probability, town average hotspot analysis using Getis-Ord G_i. Each polygon corresponds to a town. Overall averages: June, $\hat{f}_{\mathrm{i}} = 0.0188$; November, $\hat{f}_{\mathrm{i}} = 0.0146$.

In all but six of the districts with $f_{\mathrm{i}} > .01$ in either election, the party with the most votes in the district in both months is either HDP (Mus, Hakkari, Mardin, Sirnak, Agri, Siirt, Diyarbakir, Van, Batman, Igdir, Tunceli) or CHP (Mugla, Kirklareli, Edirne, Aydin). AKP is the actual leading party in five of the districts (Bingöl, Hatay, Bartin, Adiyaman, Nigde). MHP has the most votes in Osmaniye in June but AKP has the most votes in that district in November. For some of the districts $f_{\mathrm{e}} > 0$ as well. The f_{i} and f_{e} values estimated for each district are shown at the bottom of each of Figures 8–20.

"Incremental frauds" that benefit AKP are most prevalent in eastern Turkey.

Aydin and Nigde have multimodal patterns for turnout and AKP-vote proportions that strongly resemble the patterns observed in eastern Turkey districts.

Figure 8: Vote and Turnout Distribution,Turkey 2015, November, Mus

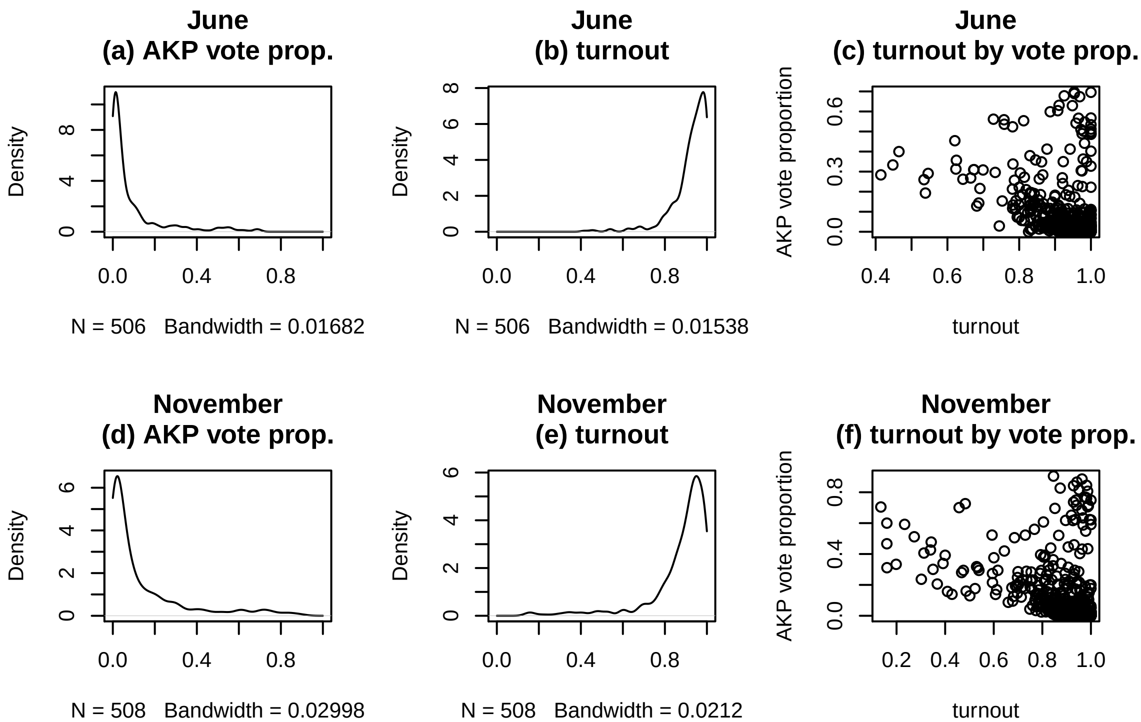


Note: distributions and scatterplot based on polling station observations. Finite mixture model estimates: June, $f_{\mathrm{i}} = .314$, $f_{\mathrm{e}} = .0214$; November, $f_{\mathrm{i}} = .2763$, $f_{\mathrm{e}} = .0096$. Number of polling stations: 845. Party with the most votes in the district: HDP.

Multimodality in the joint distribution of turnout and AKP vote proportions is apparent in all the plots shown for districts in eastern Turkey (Figures 8–21), with the multimodality being perhaps the least obvious visually in the displays for Sirnak (Figure 11). Multimodalities are much more apparent in the distribution for the districts in eastern Turkey than they are in the other districts that have $f_{\mathrm{i}} > .01$. Not only does "incremental fraud" appear to be more prevalent in the districts in eastern Turkey, but the patterns of vote stealing and vote manufacturing there differ from patterns in most of the rest of Turkey.

An important limitation of the Klimek et al. (2012) concept is that it imagines that all "frauds" benefit only one party. It is not possible to allow several different parties diversely

Figure 9: Vote and Turnout Distribution,Turkey 2015, November, Hakkari



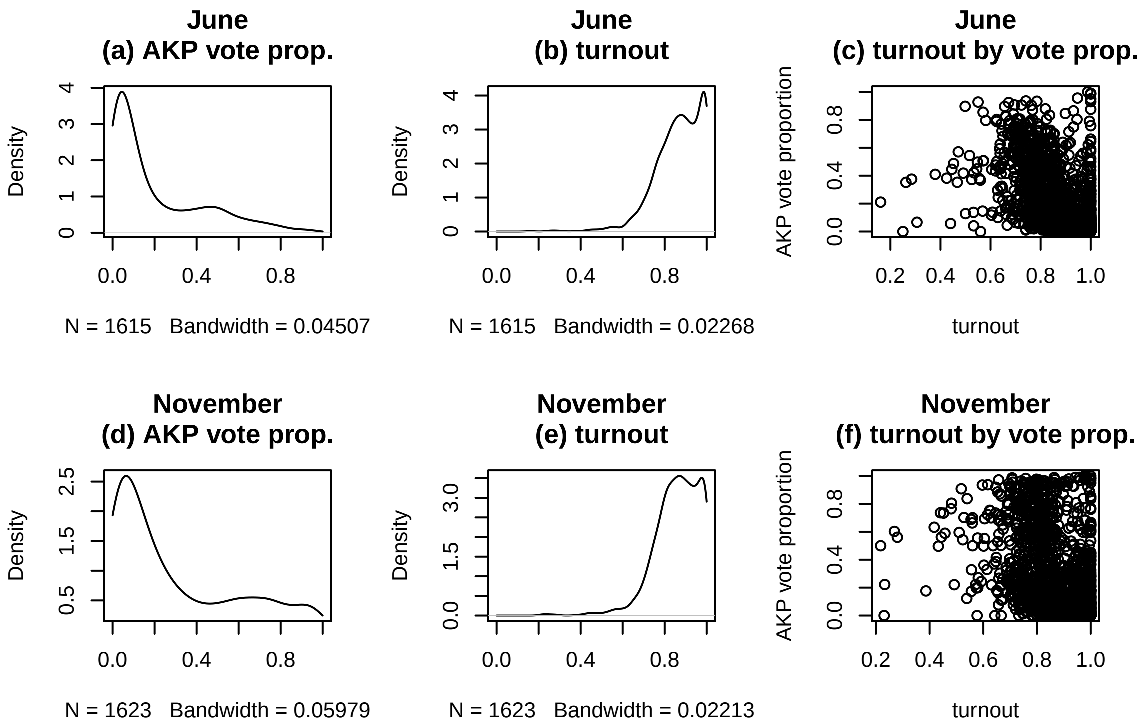
Note: distributions and scatterplot based on polling station observations. Finite mixture model estimates: June, $f_{\mathrm{i}} = .212$, $f_{\mathrm{e}} = .014$; November, $f_{\mathrm{i}} = .2645$, $f_{\mathrm{e}} = .0201$. Number of polling stations: 508. Party with the most votes in the district: HDP.

to benefit from "frauds." The analysis in this subsection is produced by using AKP as the leading party in all districts. The analysis is the appropriate one to consider if the idea is that AKP benefits from "frauds" regardless of whether it has the most votes in a district. The pattern of an increased number of polling stations with turnout and AKP vote proportions greater than 0.8 in November is visible evidence that supports this idea.

2 Summary

All in all these results suggest that new fraudulent activites affected votes and turnout in November. The discrepancies between the joint distributions of turnout and AKP-vote

Figure 10: Vote and Turnout Distribution,Turkey 2015, November, Mardin

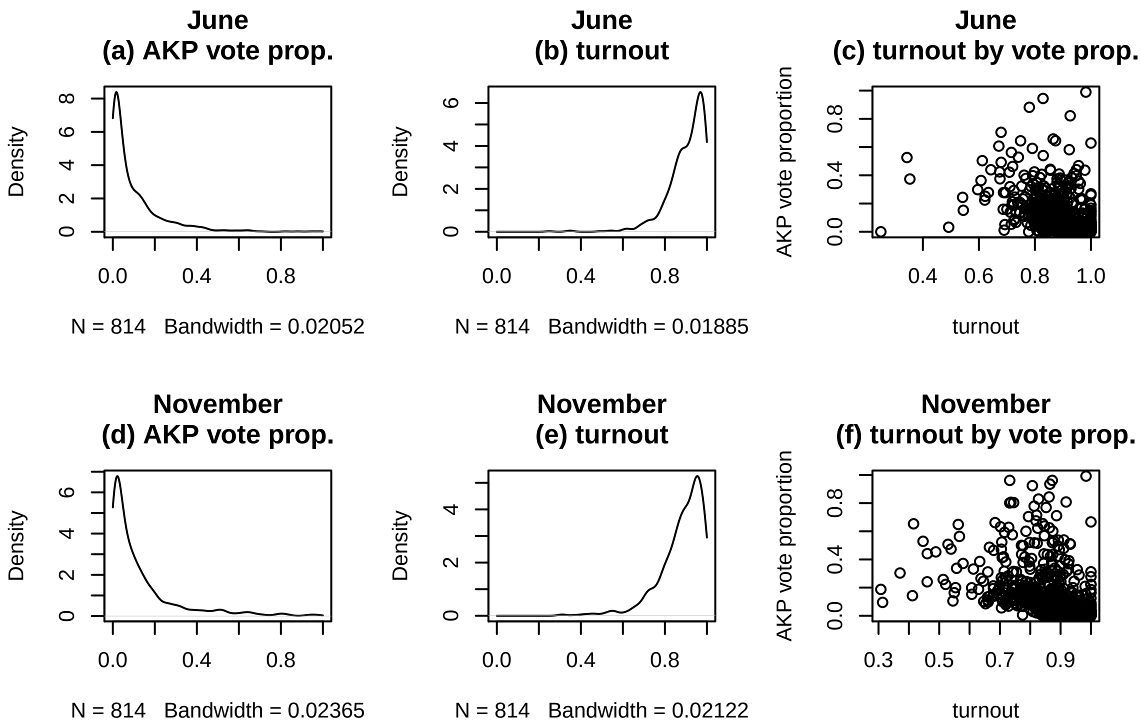


Note: distributions and scatterplot based on polling station observations. Finite mixture model estimates: June, $f_{\mathrm{i}} = .258$, $f_{\mathrm{e}} = .00103$; November, $f_{\mathrm{i}} = .2148$, $f_{\mathrm{e}} = .0016$. Number of polling stations: 1623. Party with the most votes in the district: HDP.

proportions in the June and November elections go a long way to explain why the pattern of election outcomes in November often does not much resemble the pattern in June (recall the wrong signs and outlier frequencies reported in Table 3). Using detailed results from the finite mixture models, we can list the polling stations at which "frauds" of the kind that model identifies are highly likely to have happened. We omit such a listing here.

Many districts have features that can be considered anomalous. Most of districts in eastern Turkey exhibit distinctive patterns for enough different kinds of statistics in both June and November that the results in those districts are very likely affected by substantial frauds. Some kinds of frauds seem to occur more extensively in November than in June, but all things considered it is not clear which election exhibited the most consequential

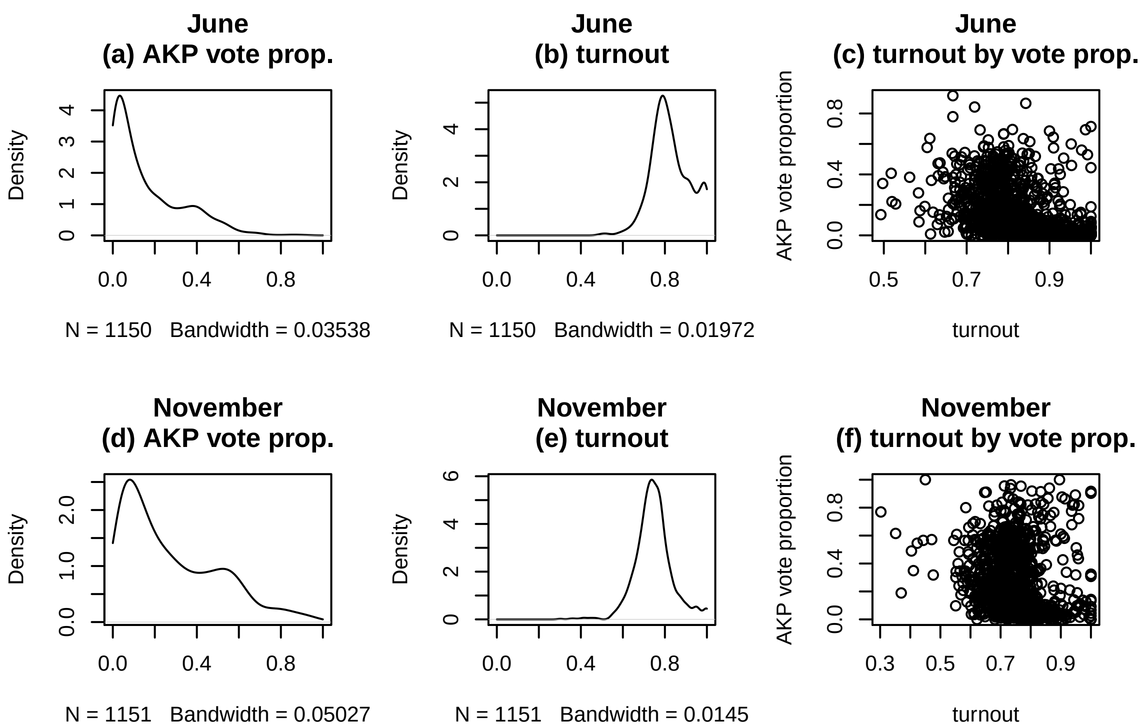
Figure 11: Vote and Turnout Distribution,Turkey 2015, November, Sirnak



Note: distributions and scatterplot based on polling station observations. Finite mixture model estimates: June, $f_\mathrm{i} = .245$, $f_\mathrm{e} = .00255$; November, $f_\mathrm{i} = .1875$, $f_\mathrm{e} = .0031$. Number of polling stations: 814. Party with the most votes in the district: HDP.

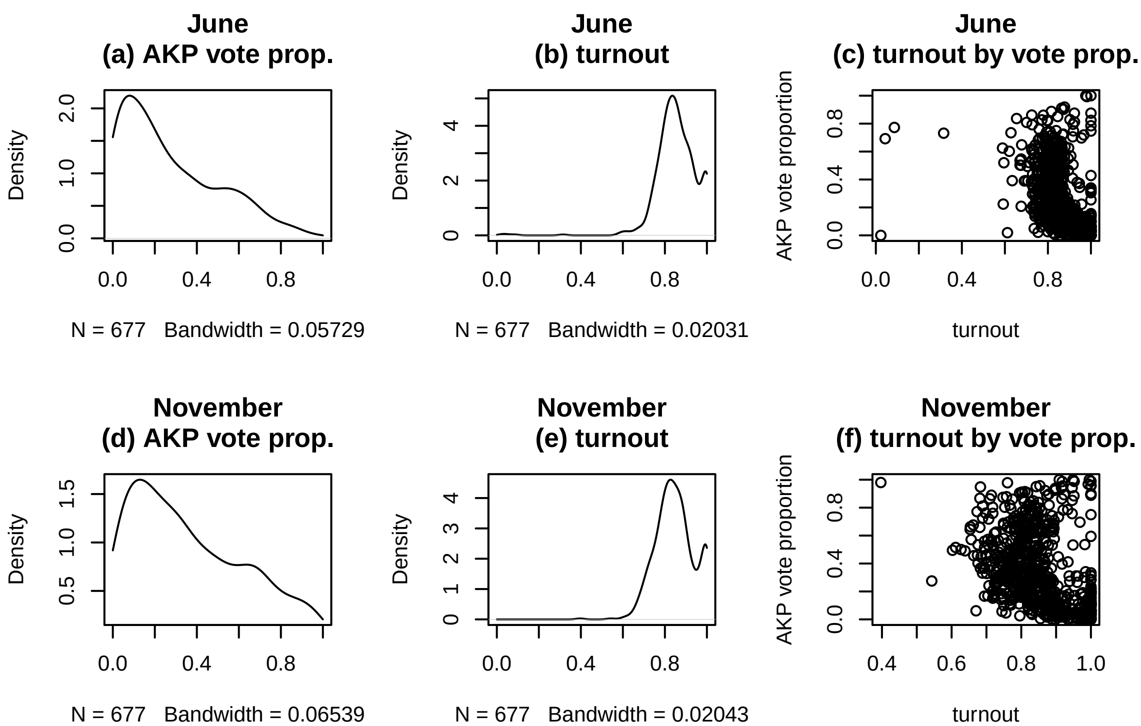
anomalies. AKP is the party that by some measures benefited from the most extensive "frauds," but other parties appear occasionally to have benefited from "frauds" as well. The diversity of the anomalies and frauds is a bit beyond the scope of existing election forensics technologies to diagnose perfectly.

Figure 12: Vote and Turnout Distribution,Turkey 2015, November, Agri



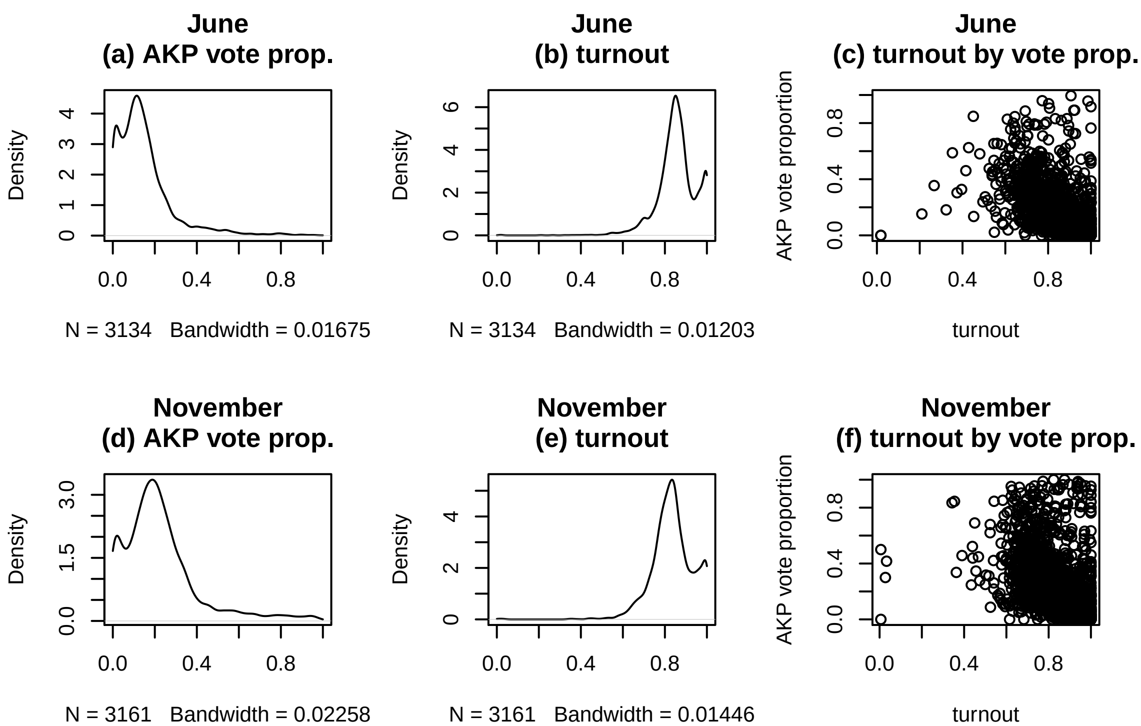
Note: distributions and scatterplot based on polling station observations. Finite mixture model estimates: June, $f_{\mathrm{i}} = .221$, $f_{\mathrm{e}} = .000421$; November, $f_{\mathrm{i}} = .1044$, $f_{\mathrm{e}} = 0$. Number of polling stations: 1151. Party with the most votes in the district: HDP.

Figure 13: Vote and Turnout Distribution,Turkey 2015, November, Siirt



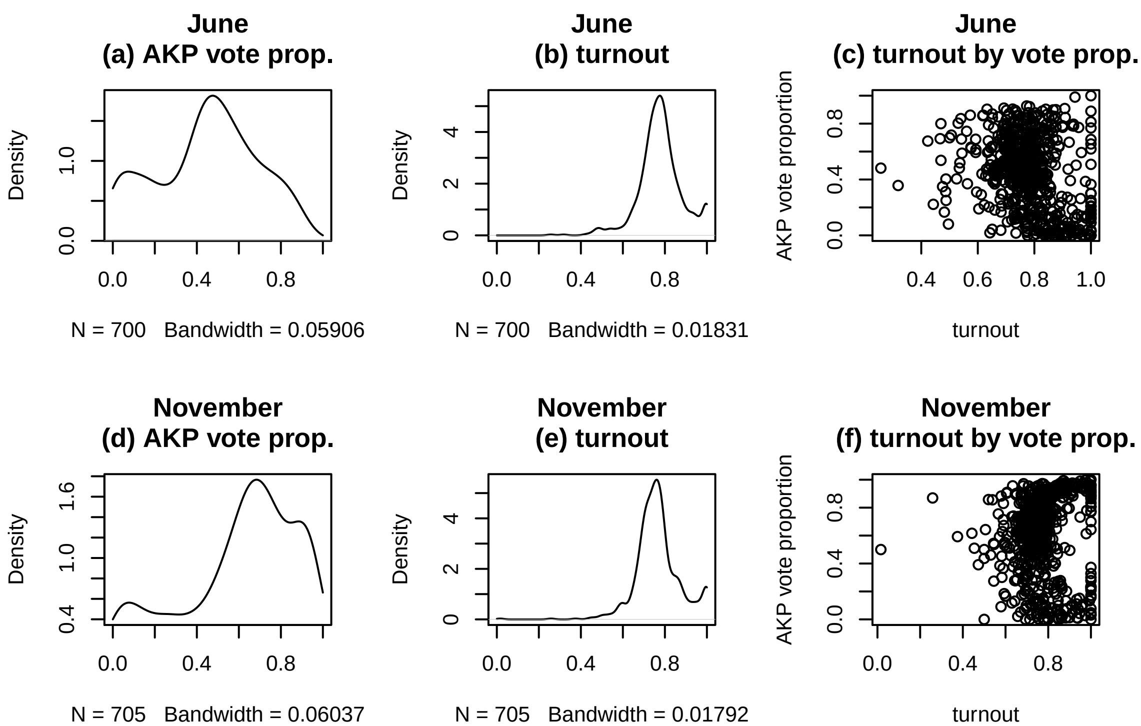
Note: distributions and scatterplot based on polling station observations. Finite mixture model estimates: June, $f_{\mathrm{i}} = .0398$, $f_{\mathrm{e}} = .00265$; November, $f_{\mathrm{i}} = .05098$, $f_{\mathrm{e}} = 0$. Number of polling stations: 677. Party with the most votes in the district: HDP.

Figure 14: Vote and Turnout Distribution,Turkey 2015, November, Diyarbakir



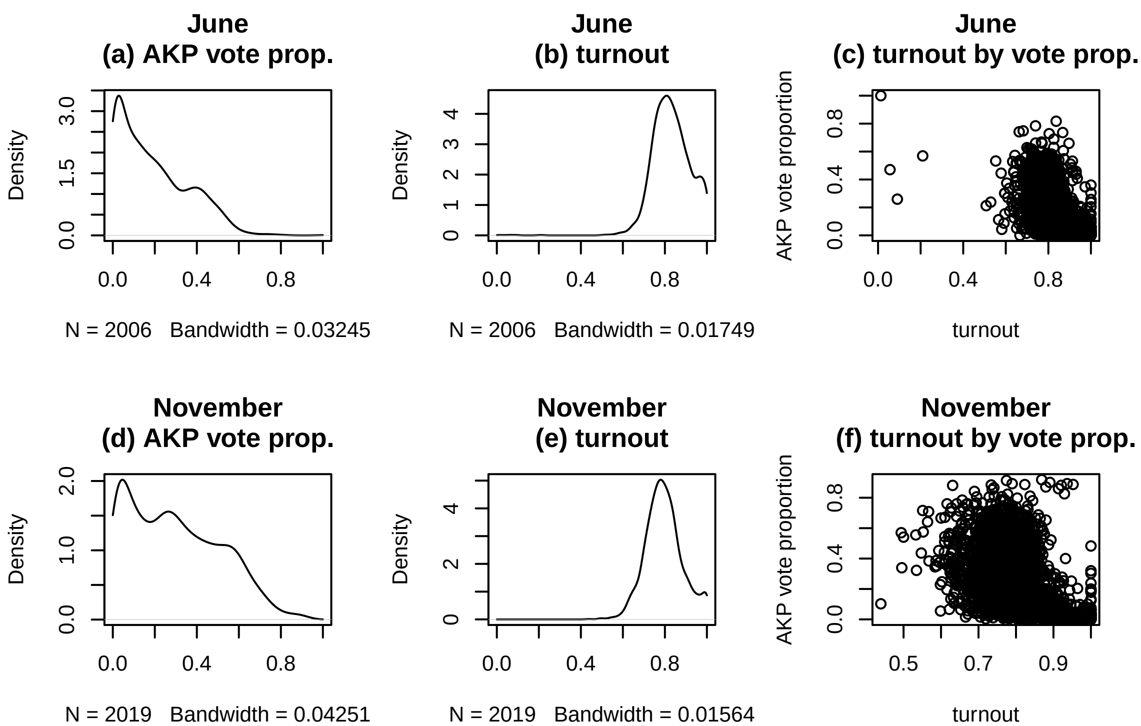
Note: distributions and scatterplot based on polling station observations. Finite mixture model estimates: June, $f_{\mathrm{i}} = .0532$, $f_{\mathrm{e}} = .000332$; November, $f_{\mathrm{i}} = .0351$, $f_{\mathrm{e}} = 3.539\mathrm{e}{-06}$. Number of polling stations: 3161. Party with the most votes in the district: HDP.

Figure 15: Vote and Turnout Distribution,Turkey 2015, November, Bingöl



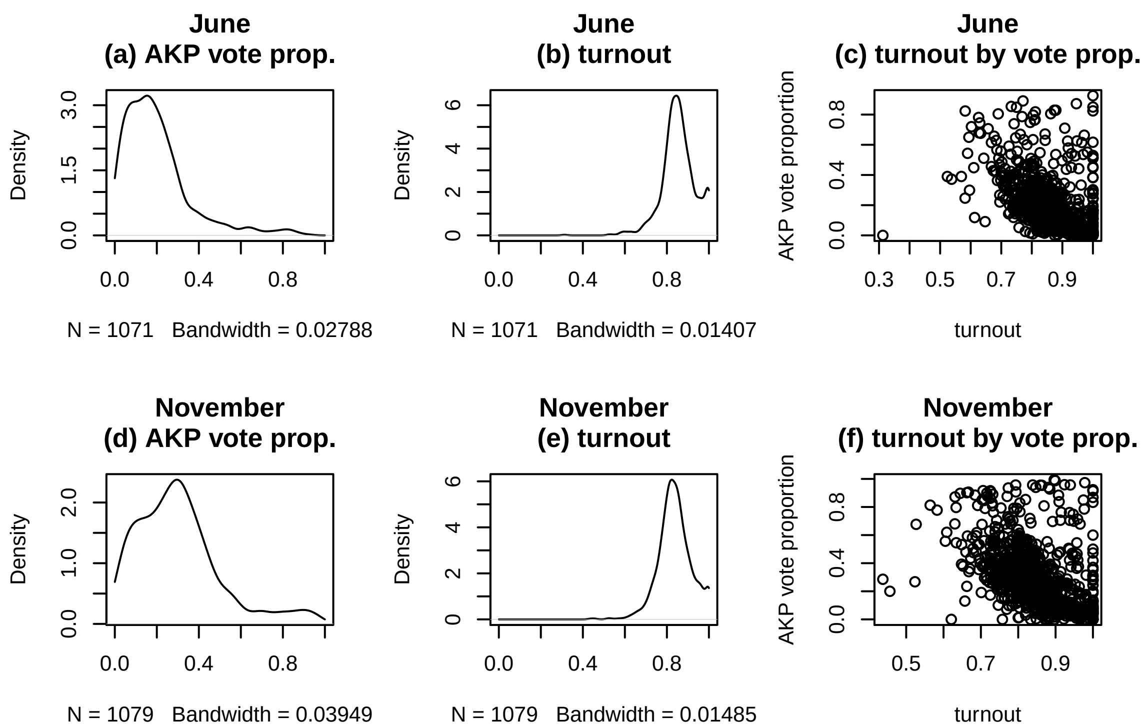
Note: distributions and scatterplot based on polling station observations. Finite mixture model estimates: June, $f_{\mathrm{i}} = .000127$, $f_{\mathrm{e}} = 0$; November, $f_{\mathrm{i}} = .01839$, $f_{\mathrm{e}} = 0$. Number of polling stations: 705. Party with the most votes in the district: AKP.

Figure 16: Vote and Turnout Distribution,Turkey 2015, November, Van



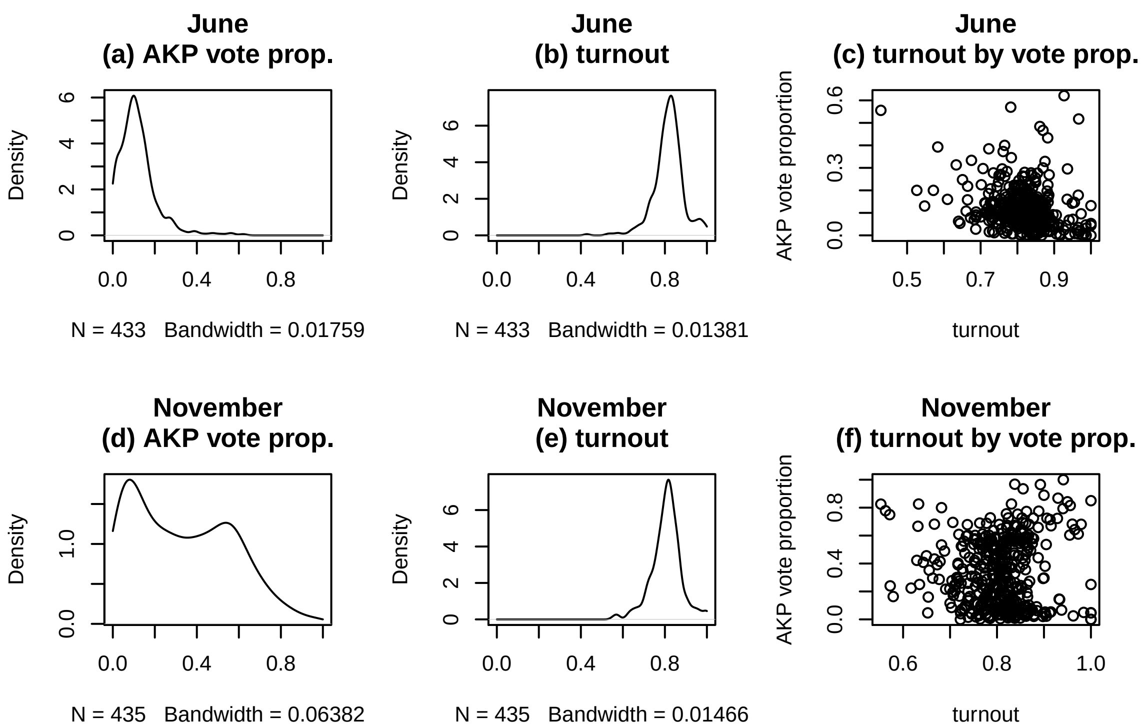
Note: distributions and scatterplot based on polling station observations. Finite mixture model estimates: June, $f_{\mathrm{i}} = .0475$, $f_{\mathrm{e}} = 0$; November, $f_{\mathrm{i}} = .2763$, $f_{\mathrm{e}} = .0096$. Number of polling stations: 845. Party with the most votes in the district: HDP.

Figure 17: Vote and Turnout Distribution,Turkey 2015, November, Batman



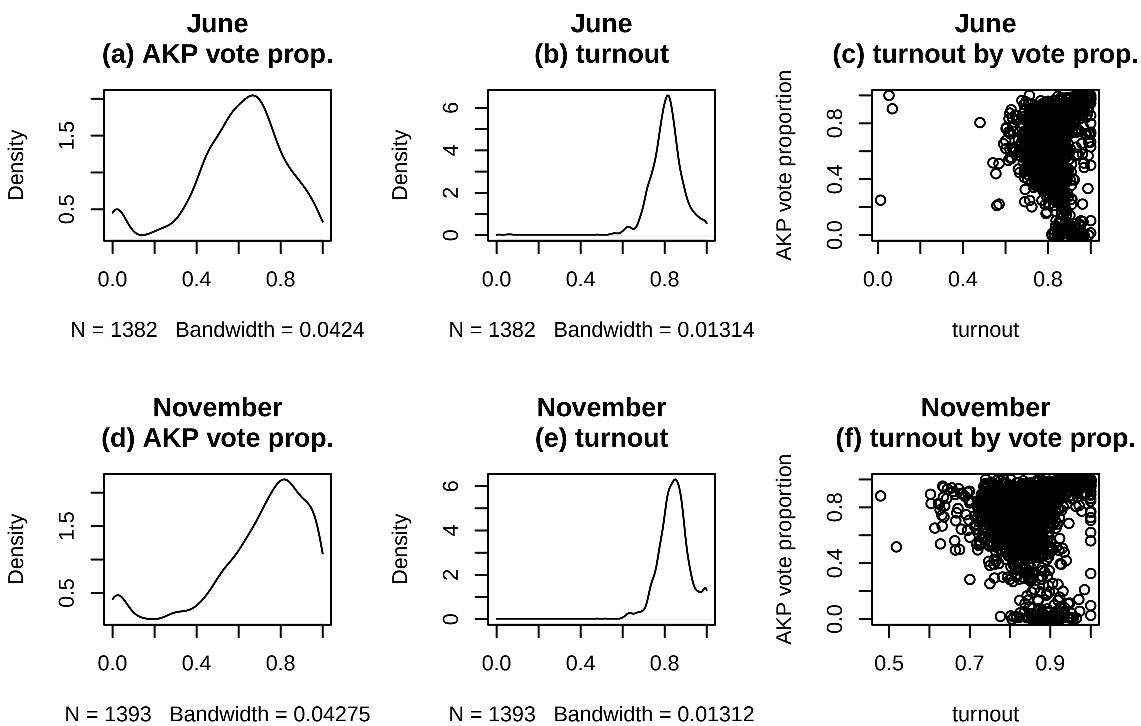
Note: distributions and scatterplot based on polling station observations. Finite mixture model estimates: June, $f_{\mathrm{i}} = .0348$, $f_{\mathrm{e}} = .000144$; November, $f_{\mathrm{i}} = .2763$, $f_{\mathrm{e}} = .0096$. Number of polling stations: 845. Party with the most votes in the district: HDP.

Figure 18: Vote and Turnout Distribution,Turkey 2015, November, Igdir



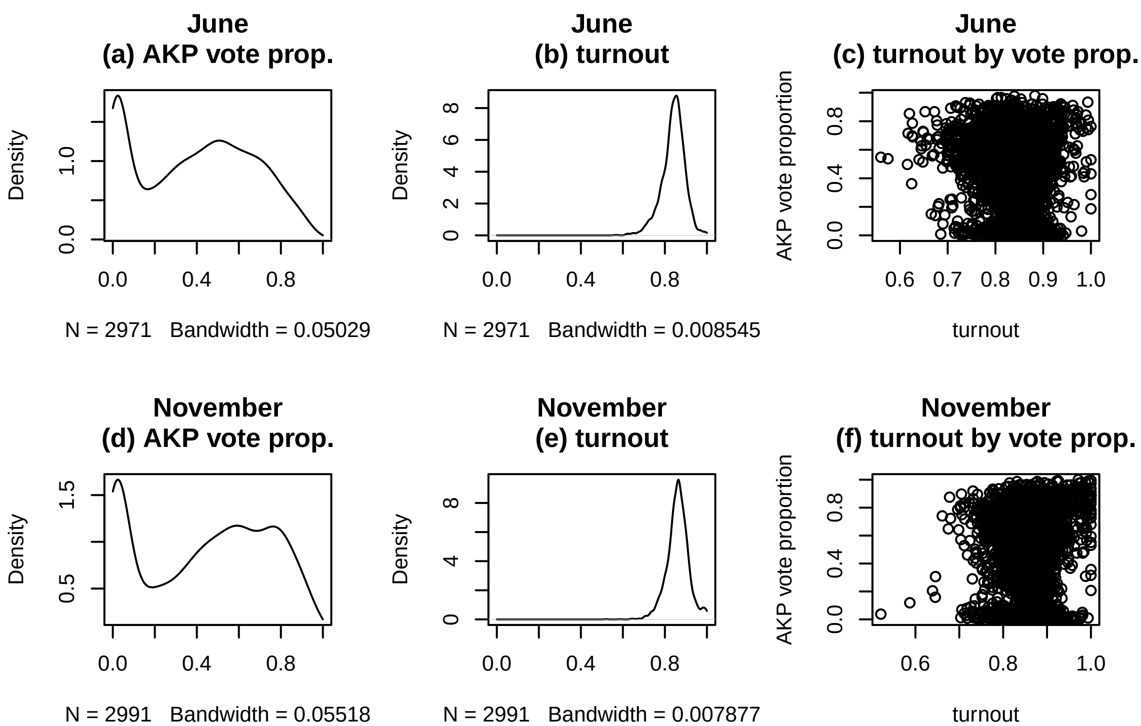
Note: distributions and scatterplot based on polling station observations. Finite mixture model estimates: June, $f_{\mathrm{i}} = .0239$, $f_{\mathrm{e}} = .0000356$; November, $f_{\mathrm{i}} = .2763$, $f_{\mathrm{e}} = .0096$. Number of polling stations: 845. Party with the most votes in the district: HDP.

Figure 19: Vote and Turnout Distribution,Turkey 2015, November, Adiyaman



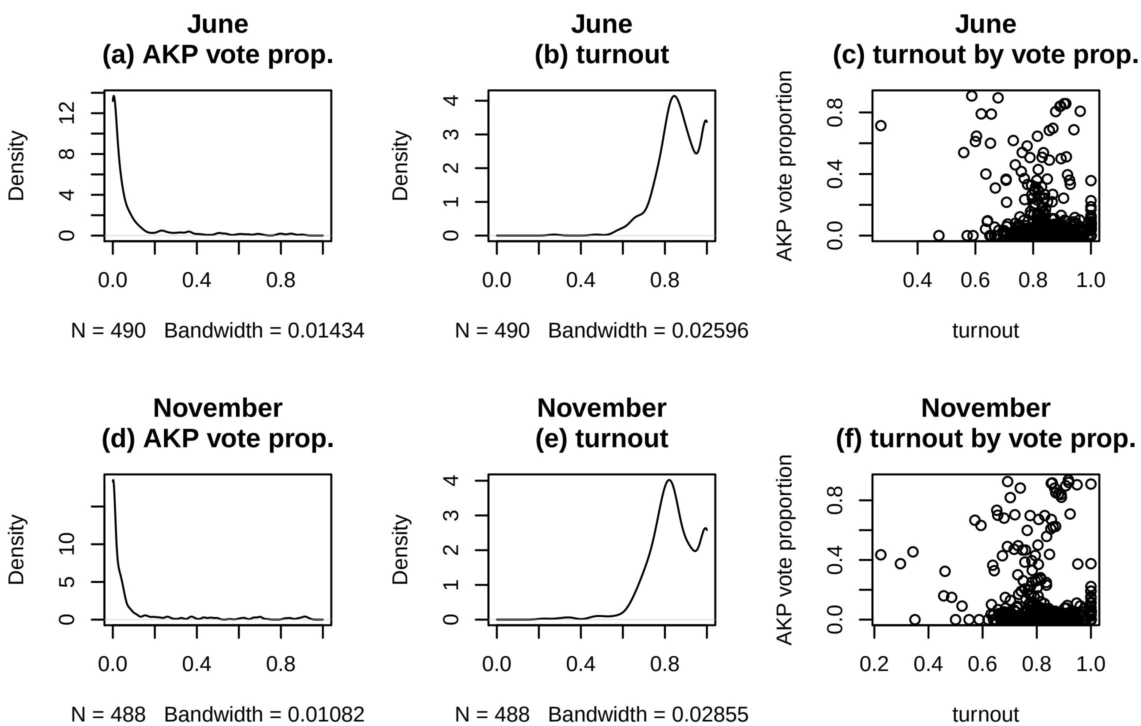
Note: distributions and scatterplot based on polling station observations. Finite mixture model estimates: June, $f_{\mathrm{i}} = .0111$, $f_{\mathrm{e}} = 0$; November, $f_{\mathrm{i}} = .2763$, $f_{\mathrm{e}} = .0096$. Number of polling stations: 845. Party with the most votes in the district: AKP.

Figure 20: Vote and Turnout Distribution,Turkey 2015, November, Hatay



Note: distributions and scatterplot based on polling station observations. Finite mixture model estimates: June, $f_{\mathrm{i}} = .00211$, $f_{\mathrm{e}} = 3.09\text{e-}09$; November, $f_{\mathrm{i}} = .0275$, $f_{\mathrm{e}} = 0$. Number of polling stations: 2991. Party with the most votes in the district: AKP.

Figure 21: Vote and Turnout Distribution,Turkey 2015, November, Tunceli



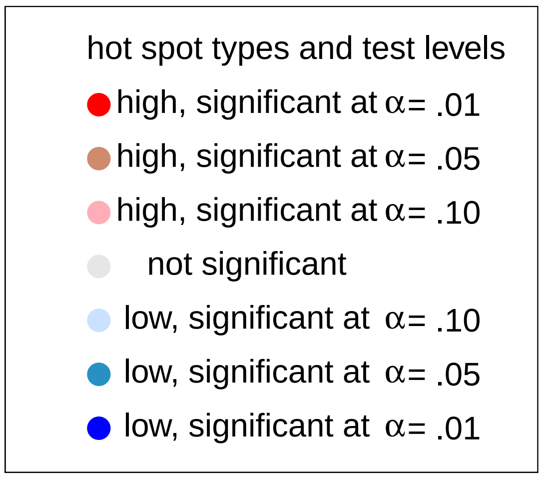
Note: distributions and scatterplot based on polling station observations. Finite mixture model could not be estimated. Party with the most votes in the district: HDP.

3 Appendix

Table 4: Finite Mixture Model Frauds Probabilities, Turkey 2015, November

District	f_i	f_e	District	f_i	f_e
Adana	0.0028	0	Karabük	4e-05	0
Adiyaman	0.0031	0	Karaman	0.0006	0
Afyonkarahisar	0.0021	0	Kars	0.0037	0.0031
Agri	0.0671	0	Kastamonu	0.0025	0
Aksaray	0.0012	0	Kayseri	0.0040	0
Amasya	0.0026	0	Kilis	0.0004	0
Ankara_I	0.0032	0	Kirikkale	0.0035	0
Ankara_II	0.0013	0	Kirklareli	0.0012	0
Antalya	0.0018	0	Kirsehir	0	0
Ardahan	0.0098	0	Kocaeli	0.0021	0
Artvin	0.0059	3e-07	Konya	0.0012	0
Aydin	0.0022	0	Kütahya	1.e-07	0
Balikesir	0.0008	0	Malatya	0.0045	0
Bartin	0.0165	0	Manisa	0.0045	0
Batman	0.0002	0	Mardin	0.0051	2e-08
Bayburt	0.0002	0	Mersin	0.0011	0
Bilecik	0.0017	0	Mugla	9e-06	0
Bingöl	0.0285	0	Mus	0.0015	0
Bitlis	0.0015	0	Nevsehir	0.0003	0
Bolu	0.0010	0	Nigde	0.0019	0
Burdur	0.0016	0	Ordu	0.0061	5e-07
Bursa	0.0037	0	Osmaniye	0.0055	0
Denizli	0.0006	0	Rize	0.0026	0
Düzce	0.0012	0	Samsun	0.0099	0
Diyarbakir	0.0187	1e-08	Sanliurfa	0.0017	6e-07
Edirne	0.0004	0	Sakarya	0.0015	0
Elazig	0.0060	0	Siirt	4e-06	0
Erzincan	0.0041	0	Sinop	0.0084	0
Erzurum	0.0003	0	Sirnak	0.0006	2e-08
Eskisehir	0.0006	0	Sivas	0.0014	0
Gaziantep	0.0053	1e-09	Tekirdag	5e-06	0
Giresun	0.0075	6e-06	Tokat	0.0023	0
Gümüshane	1e-05	0	Trabzon	0.0030	0
Hakkari	8e-05	0	Tunceli	0.0032	0
Hatay	0.0275	0	Usak	0.0013	0
Igdir	0.0013	0.0019	Van	0.0451	0
Isparta	0.0021	0	Yalova	0.0050	0
Istanbul_I	0.0004	0	Yozgat	0.0018	0
Istanbul_II	0.0005	0	Zonguldak	0.0022	0
Istanbul_III	0.0004	0	Çanakkale	0.0027	0
Izmir_I	0.0012	0	Çankiri	1e-07	0
Izmir_II	0.0010	0	Çorum	0.0025	0
Kahramanmaras	0.0002	0			

Figure 22: Hotspot Analysis Legend



Note: Significance levels refer to tests adjusted for the false discovery rate (Benjamini and Hochberg 1995). This figure displays the legend for hotspot maps used in this paper. Red colors show areas where local average scores are significantly above the overall average. Blue colors show areas where local average scores are significantly below the overall average.

References

Álvarez-Rivera, Manuel. 2015. "Election Resources on the Internet: Elections to the Turkish Grand National Assembly." URL `http://electionresources.org/tr/`. November 14, 2015 update.

Beber, Bernd and Alexandra Scacco. 2012. "What the Numbers Say: A Digit-Based Test for Election Fraud." *Political Analysis* 20(2):211–234.

Benjamini, Yoav and Yosef Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society, Series B* 57(1):289–300.

Deckert, Joseph, Mikhail Myagkov and Peter C. Ordeshook. 2011. "Benford's Law and the Detection of Election Fraud." *Political Analysis* 19(3):245–268.

Getis, Arthur and J. K. Ord. 1992. "The Analysis of Spatial Association by Use of Distance Statistics." *Geographical Analysis* 24(3):189–206.

Hartigan, J. A. and P. M. Hartigan. 1985. "The Dip Test of Unimodality." *Annals of Statistics* 13:70–84.

Hicken, Allen and Walter R. Mebane, Jr. 2015. "A Guide to Election Forensics." Working paper for IIE/USAID subaward #DFG-10-APS-UM, "Development of an Election Forensics Toolkit: Using Subnational Data to Detect Anomalies".

Hijmans, Robert. 2015. "GADM Database: Global Administrative Areas." Version 2.8 (November 2015). URL `http://www.gadm.org`.

Kalinin, Kirill and Walter R. Mebane, Jr. 2011. "Understanding Electoral Frauds through Evolution of Russian Federalism: from "Bargaining Loyalty" to "Signaling Loyalty"." Paper presented at the 2011 Annual Meeting of the Midwest Political Science Association, Chicago, IL, March 31–April 2.

Klimek, Peter, Yuri Yegorov, Rudolf Hanel and Stefan Thurner. 2012. "Statistical Detection of Systematic Election Irregularities." *Proceedings of the National Academy of Sciences* 109:16469–16473.

Mebane, Jr., Walter R. 2010. "Fraud in the 2009 Presidential Election in Iran?" *Chance* 23:6–15.

Mebane, Jr., Walter R. 2011. "Comment on 'Benford's Law and the Detection of Election Fraud'." *Political Analysis* 19(3):269–272.

Mebane, Jr., Walter R. 2013. "Election Forensics: The Meanings of Precinct Vote Counts' Second Digits." Paper presented at the 2013 Summer Meeting of the Political Methodology Society, University of Virginia, July 18–20, 2013.

Mebane, Jr., Walter R. 2015*a*. "Election Forensics: Latent Dimensions of Election Frauds and Strategic Voting." Paper presented at the 2015 Summer Meeting of the Political Methodology Society, Rochester, July 23–25.

Mebane, Jr., Walter R. 2015*b*. "Election Forensics Toolkit DRG Center Working Paper." Working paper for IIE/USAID subaward #DFG-10-APS-UM, "Development of an Election Forensics Toolkit: Using Subnational Data to Detect Anomalies".

Mebane, Jr., Walter R. n.d. "Election Forensics." book MS.

Mebane, Jr., Walter R. and Jasjeet S. Sekhon. 2004*a*. "Robust Estimation and Outlier Detection for Overdispersed Multinomial Models of Count Data." *American Journal of Political Science* 48(2):391–410.

Mebane, Jr., Walter R. and Jasjeet Singh Sekhon. 2004*b*. "MultinomRob: Robust Estimation of Overdispersed Multinomial Regression Models." Computer software: Comprehensive R Archive Network.

Mebane, Jr., Walter R. and Jonathan Wall. 2015. "Election Frauds, Postelection Legal Challenges and Geography in Mexico." Paper presented at the 2015 Annual Meeting of the American Political Science Association, San Francisco, CA, September 3–6.

Mebane, Jr., Walter R. and Michael C. Herron. 2005. Ohio 2004 Election: Turnout, Residual Votes and Votes in Precincts and Wards. In *Democracy at Risk: The 2004 Election in Ohio*, ed. Democratic National Committee Voting Rights Institute. Washington, D.C.: Democratic National Committee. June 9, 2005.

Myagkov, Mikhail, Peter C. Ordeshook and Dimitry Shaikin. 2009. *The Forensics of Election Fraud: With Applications to Russia and Ukraine*. New York: Cambridge University Press.

Ord, J. K. and Arthur Getis. 1995. "Local Spatial Autocorrelation Statistics: Distributional Issues and an Application." *Geographical Analysis* 27(4):286–306.

Pericchi, Luis Raúl and David Torres. 2011. "Quick Anomaly Detection by the Newcomb-Benford Law, with Applications to Electoral Processes Data from the USA, Puerto Rico and Venezuela." *Statistical Science* 26(4):502–516.

Rundlett, Ashlea and Milan W. Svolik. 2015. "Deliver the Vote! Micromotives and Macrobehavior in Electoral Fraud." Working paper.

Turkish Press. 2010. "Elections in Turkey." URL `http://www.turkishelections.com/`.

Wand, Jonathan, Kenneth Shotts, Jasjeet S. Sekhon, Walter R. Mebane, Jr., Michael Herron and Henry E. Brady. 2001. "The Butterfly Did It: The Aberrant Vote for Buchanan in Palm Beach County, Florida." *American Political Science Review* 95:793–810.

Yüksek Seçim Kurulu. 2015. "136 Sayili Genelge: İl Seçim Kurullarının Seçim Sonuçlarına İlişkin Görevleri ile Yurt Düzeyi Seçim Sonuçlarının Belirlenmesinde Uygulanacak Esas ve İlkeleri (Circular No. 136: Dormitory Level Task Relating to the Provincial Election Committee Election, Election Results Principles and Guidelines Will be Applied in Determining the Outcome)." URL `http://www.ysk.gov.tr/ysk/content/conn/YSKUCM/path/Contribution%20Folders/Genelgeler/2015MVES-Genelge136.pdf` (September 9, 2015).