Anomalies and Frauds in the Korea 2020 Parliamentary Election*

Walter R. Mebane, Jr.†

May 9, 2020

*Thanks to Hun Chung for highlighting the concerns with the election (as did several others) and for pointing to the dataset originally used in the analysis.

†Professor, Department of Political Science and Department of Statistics, Research Professor, Center for Political Studies, University of Michigan, Haven Hall, Ann Arbor, MI 48109-1045 (E-mail: wmebane@umich.edu).
The 2020 parliamentary election in Korea is controversial, with fraud allegations. I examine data from the election using eforensics, tests from the Election Forensics Toolkit and the spikes test. This paper improves on a previous version (“Frauds in the Korea 2020 Parliamentary Election,” April 29, 2020) by using updated complete data and by adding Election Forensics Toolkit and spikes test results. The estimates and tests all exhibit anomalies that suggest the election data were fraudulently manipulated.

1 eforensics

The statistical model implemented in eforensics offers evidence that fraudulent votes occurred in the election that may have changed some election outcomes. The statistical model operationalizes the idea that “frauds” occur when one party gains votes by a combination of manufacturing votes from abstentions and stealing votes from opposing parties. The Bayesian specification allows posterior means and credible intervals for counts of “fraudulent” votes to be determined both for the entire election and for observed individual aggregation units.

It is important to keep in mind that “frauds” according to the eforensics model may or may not be results of malfeasance and bad actions. How much estimated “frauds” may be produced by normal political activity, and in particular by strategic behavior, is an open question that is the focus of current research. Statistical findings such as are reported here should be followed up with additional information and further investigation into what happened. The statistical findings alone cannot stand as definitive evidence about what happened in an election.

Figure 1 shows the distribution of turnout and vote proportions across aggregation units. Each turnout proportion is (Number Valid)/(Number Eligible), and each vote

1 https://github.com/UMeforensics/eforensics_public
2 Ferrari, McAlister and Mebane [2018] and http://www.umich.edu/~wmebane/efslides.pdf
The proportion is \((\text{Number Voting for Party})/(\text{Number Eligible})\). The data include counts for \(n = 19131\) units. 329 “abroad_office” observations have zero eligible voters but often a small number of votes—the largest number is 23—and are omitted from the plots. One “prevote_in” unit with zero voters and zero votes is also omitted. Figure 1(a) uses vote proportions defined based on Democratic Party votes, and Figure 1(b) uses vote proportions defined based on the votes received by the party with the most votes in each constituency. Fraud allegations have focused on the Democratic Party, but a principled way to analyze the single-member district (SMD) election data is to consider that frauds potentially benefited the leading candidate in each constituency. In the figure differences between the two distributions are apparent, but both share a distinctive multimodal pattern. There appear to be clusters of observations that share distinctive levels of turnout and votes, some with low, medium, high and very high turnout. The diagonal edge feature in the plots results from using Number Eligible as the denominator for both proportions: when the party receives nearly all the valid votes, then the observation is near that diagonal.

Figures 2 and 3 show that the different clusters in Figure 1 correspond with observations that are administratively distinctive. Figure 2 displays data for Democratic Party votes, and Figure 3 shows data for constituency leader votes. The four sets of units that have distinctive distributions are district-level, election-day units that are not abroad (Figures 2(a) and 3(a)), voting post, election-day units (Figures 2(b) and 3(b)), abroad.


4“Number Valid” is the number voting for any candidate, and “Number Eligible” is sum_people in korea_election_regional_21_eng.sqlite. Candidates are mapped from the candidate table of korea_election_regional_21_eng.sqlite to parties, which in the party table in korea_election_regional_21_eng.sqlite are numbered as 3, 5, 7, 9, 11, 532, 616, 695, 1195, 3006, 4133, 10237, 11069, 11955, 13243, 13316, 13318, 13321, 13323, 13676, 17524, 18267, then votes for any of those parties are summed for each aggregation unit observation to produce “Number Valid” for that unit.
Figure 1: Korea 2020 Parliamentary Election Data Plots

(a) Democratic Party
(b) Constituency leaders

Note: plots show turnout (number voting/number eligible) and vote proportions (number voting for party/number eligible) for (a) the Democratic Party or (b) the party the most votes in each constituency in aggregation units in the Korea 2020 parliamentary election. Plots show scatterplots with estimated bivariate densities overlaid, with histograms along the axes. 329 “abroad_office” observations reported with zero eligible voters but often with a positive number of votes are omitted. One “prevote” unit with zero voters and zero votes is also omitted.

units (Figures 2(c) and 3(c)) and pre-vote units (Figures 2(d) and 3(d)). Each subset of units (a), (b) and (d) has a mostly unimodal distribution: the marginal histograms are mostly near symmetric. But exceptional points are evident in each of these subsets. Abroad units are more distinctively bimodal when constituency leaders are considered than when the Democratic Party is in focus.
Figure 2: Korea 2020 Parliamentary Election Data Plots, Democratic Party

(a) district, election-day, not abroad

(b) voting post, election-day

(c) abroad

(d) pre-vote

Note: plots show turnout (number voting/number eligible) and vote proportions (number voting for Democratic party/number eligible) for four subsets of observations: (a) district-level, election-day, not abroad; (b) voting post election-day; (c) abroad; (d) pre-vote. Plots show scatterplots with estimated bivariate densities overlaid, with histograms along the axes. 329 “abroad_office” observations reported with zero eligible voters but often with a positive number of votes are omitted. One “prevote” unit with zero voters and zero votes is also omitted.
Figure 3: Korea 2020 Parliamentary Election Data Plots, Constituency Leaders

(a) district, election-day, not abroad
(b) voting post, election-day
(c) abroad
(d) pre-vote

Note: plots show turnout (number voting/number eligible) and vote proportions (number voting for constituency-leading party/number eligible) for four subsets of observations: (a) district-level, election-day, not abroad; (b) voting post election-day; (c) abroad; (d) pre-vote. Plots show scatterplots with estimated bivariate densities overlaid, with histograms along the axes. 329 “abroad_office” observations reported with zero eligible voters but often with a positive number of votes are omitted. One “prevote” unit with zero voters and zero votes is also omitted.
I estimate the eforensics model separately for the two definitions of leading party votes. Covariates for turnout and vote choice include indicators for pre-vote, voting post, abroad and disabled-ship status and fixed effects for the 253 constituencies included in the data. The two specifications agree that 506 aggregation units are fraudulent, but 664 additional units are fraudulent in the Democratic party specification and 988 additional units are fraudulent in the constituency-leading party specification. As Table 1 shows, key parameter estimates are similar in the models. Parameters for the probabilities of incremental and extreme frauds (\( \pi_2, \pi_3 \)) are very slightly greater in the constituency leader specification, and coefficients for the turnout equation (\( \tau_0 - \tau_4 \)) are similar. Coefficients for vote choice (\( \beta_0 - \beta_4 \)) differ, reflecting the differences in vote proportions being modeled.

Figure 4 uses plots by subset of Democratic party focused observations to illustrate which observations are fraudulent according to the eforensics model with the Democratic party focused specification. Nonfraudulent observations are plotted in blue and fraudulent observations appear in red. The frequencies of fraudulent and not fraudulent units appear in the note at the bottom of the figure. Visually and by the numbers, frauds occur most frequently for pre-vote units (27.9% are fraudulent), next most frequently for for district-level, election-day, not abroad units (3.04% fraudulent) then next most frequently voting post election day units (.67% are fraudulent) then abroad units (.30% are fraudulent).

Figure 5 uses plots by subset of constituency-leader focused observations to illustrate which observations are fraudulent according to the eforensics model with the constituency-leader focused specification. Nonfraudulent observations are plotted in blue and fraudulent observations appear in red. The frequencies of fraudulent and not fraudulent units appear in the note at the bottom of the figure. Visually and by the numbers, frauds occur most frequently for pre-vote units (30.9% are fraudulent), next most frequently for district-level, election-day, not abroad units (2.73% fraudulent) then next most frequently voting post election day units (2.14% are fraudulent). None of the
abroad units are fraudulent.
Table 1: Korea 2020 Parliamentary eforensics Estimates

(a) Democratic Party specification

<table>
<thead>
<tr>
<th>Parm.</th>
<th>Covariate</th>
<th>Mean</th>
<th>HPD.lo&lt;sup&gt;a&lt;/sup&gt;</th>
<th>HPD.up&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>π₁</td>
<td>No Fraud</td>
<td>.934</td>
<td>.930</td>
<td>.938</td>
</tr>
<tr>
<td>π₂</td>
<td>Incremental Fraud</td>
<td>.0611</td>
<td>.0573</td>
<td>.649</td>
</tr>
<tr>
<td>π₃</td>
<td>Extreme Fraud</td>
<td>.00492</td>
<td>.00383</td>
<td>.00582</td>
</tr>
<tr>
<td>γ₀</td>
<td>(Intercept)</td>
<td>.676</td>
<td>.655</td>
<td>.696</td>
</tr>
<tr>
<td>γ₁</td>
<td>pre-vote</td>
<td>1.16</td>
<td>1.09</td>
<td>1.24</td>
</tr>
<tr>
<td>γ₂</td>
<td>voting post</td>
<td>-.0461</td>
<td>-.0545</td>
<td>-.0359</td>
</tr>
<tr>
<td>γ₃</td>
<td>abroad</td>
<td>-.0488</td>
<td>-.0516</td>
<td>-.0463</td>
</tr>
<tr>
<td>γ₄</td>
<td>disabled-ship</td>
<td>.0539</td>
<td>.0459</td>
<td>.0623</td>
</tr>
<tr>
<td>β₀</td>
<td>(Intercept)</td>
<td>-.124</td>
<td>-.134</td>
<td>-.113</td>
</tr>
<tr>
<td>β₁</td>
<td>pre-vote</td>
<td>.0721</td>
<td>.0634</td>
<td>.0806</td>
</tr>
<tr>
<td>β₂</td>
<td>voting post</td>
<td>-.111</td>
<td>-.115</td>
<td>-.105</td>
</tr>
<tr>
<td>β₃</td>
<td>abroad</td>
<td>.170</td>
<td>.159</td>
<td>.178</td>
</tr>
<tr>
<td>β₄</td>
<td>disabled-ship</td>
<td>-.0264</td>
<td>-.0325</td>
<td>-.0192</td>
</tr>
</tbody>
</table>

(b) constituency leader specification

<table>
<thead>
<tr>
<th>Parm.</th>
<th>Covariate</th>
<th>Mean</th>
<th>HPD.lo&lt;sup&gt;a&lt;/sup&gt;</th>
<th>HPD.up&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>π₁</td>
<td>No Fraud</td>
<td>.910</td>
<td>.906</td>
<td>.915</td>
</tr>
<tr>
<td>π₂</td>
<td>Incremental Fraud</td>
<td>.0813</td>
<td>.0767</td>
<td>.0858</td>
</tr>
<tr>
<td>π₃</td>
<td>Extreme Fraud</td>
<td>.00841</td>
<td>.00712</td>
<td>.00987</td>
</tr>
<tr>
<td>γ₀</td>
<td>(Intercept)</td>
<td>.674</td>
<td>.649</td>
<td>.701</td>
</tr>
<tr>
<td>γ₁</td>
<td>pre-vote</td>
<td>1.14</td>
<td>1.06</td>
<td>1.21</td>
</tr>
<tr>
<td>γ₂</td>
<td>voting post</td>
<td>-.0424</td>
<td>-.0494</td>
<td>-.0373</td>
</tr>
<tr>
<td>γ₃</td>
<td>abroad</td>
<td>-.0817</td>
<td>-.0917</td>
<td>-.0693</td>
</tr>
<tr>
<td>γ₄</td>
<td>disabled-ship</td>
<td>.0553</td>
<td>.0490</td>
<td>.0626</td>
</tr>
<tr>
<td>β₀</td>
<td>(Intercept)</td>
<td>.209</td>
<td>.201</td>
<td>.219</td>
</tr>
<tr>
<td>β₁</td>
<td>pre-vote</td>
<td>-.0845</td>
<td>-.0891</td>
<td>-.0800</td>
</tr>
<tr>
<td>β₂</td>
<td>voting post</td>
<td>.0498</td>
<td>.0464</td>
<td>.0537</td>
</tr>
<tr>
<td>β₃</td>
<td>abroad</td>
<td>.00227</td>
<td>-.00611</td>
<td>.00606</td>
</tr>
<tr>
<td>β₄</td>
<td>disabled-ship</td>
<td>-.0376</td>
<td>-.0532</td>
<td>-.0266</td>
</tr>
</tbody>
</table>

Note: selected eforensics model parameter estimates. Constituency fixed effects are not shown. For parameter notation see [http://www.umich.edu/~wmebane/efslides.pdf](http://www.umich.edu/~wmebane/efslides.pdf).<sup>a</sup>

<sup>a</sup> 95% highest posterior density credible interval lower bound. <sup>b</sup> 95% highest posterior density credible interval upper bound.
Figure 4: Korea 2020 Fraud Plots, Democratic Party

(a) district, election-day, not abroad  

(b) voting post, election-day

(c) abroad

(d) pre-vote

Note: plots show turnout (number voting/number eligible) and vote proportions (number voting for Democratic Party/number eligible) for four subsets of observations: (a) district-level, election-day, not abroad (10 fraudulent, 319 not); (b) voting post election-day (96 fraudulent, 14234 not); (c) abroad (1 fraudulent, 328 not); (d) pre-vote (1063 fraudulent, 2750 not). Plots show scatterplots with nonfraudulent observations in blue and fraudulent observations in red. 329 “abroad_office” observations reported with zero eligible voters but often with a positive number of votes are omitted. One “prevote” unit with zero voters and zero votes is also omitted.
Figure 5: Korea 2020 Fraud Plots, Constituency Leaders

(a) district, election-day, not abroad

(c) abroad

(b) voting post, election-day

(d) pre-vote

Note: plots show turnout (number voting/number eligible) and vote proportions (number voting for constituency-leading party/number eligible) for four subsets of observations: (a) district-level, election-day, not abroad (9 fraudulent, 320 not); (b) voting post election-day (307 fraudulent, 14023 not); (c) abroad (0 fraudulent, 329 not); (d) pre-vote (1178 fraudulent, 2635 not). Plots show scatterplots with nonfraudulent observations in blue and fraudulent observations in red. 329 “abroad_office” observations reported with zero eligible voters but often with a positive number of votes are omitted. One “prevote” unit with zero voters and zero votes is also omitted.
I use a counterfactual method to calculate how many votes are fraudulent. Table 2 reports the observed counts of eligible voters, valid votes and votes for the (a) Democratic party and (b) constituency-leading party totaled over all units in the analysis, along with fraudulent vote count totals. The total of “manufactured” votes is reported separately from the total number of fraudulent votes: manufactured votes are votes that the model estimates should have been abstentions but instead were observed as votes for the leading party. Both posterior means and 95% and 99.5% credible intervals are reported. The results show that for the Democratic-Party-focused specification over all about 1,382,524 votes are fraudulent, and of the fraudulent votes about 1,030,562 are manufactured (the remaining 351,962 are stolen—counted for the leading party when they should have been counted for a different party). Overall, according to the eforensics model, about 9.6% of the votes for Democratic Party candidates are fraudulent. The results show that for the constituency-leading-focused specification over all about 1,404,919 votes are fraudulent, and of the fraudulent votes about 1,105,458 are manufactured (the remaining 299,461 are stolen—counted for the leading party when they should have been counted for a different party). Overall, according to the eforensics model, about 8.7% of the votes for constituency-leading candidates are fraudulent.

Fraudulent vote occurrence varies over constituencies. Counts of frauds by aggregation unit appear in a supplemental file but I use the unit-specific fraudulent vote counts from the constituency-leader focused specification to assess whether the number of fraudulent votes is ever large enough apparently to change the winner of a constituency contest. For 226 constituencies it is not, but for 27 constituencies the number of fraudulent votes is large enough apparently to change the winner of the constituency contest. In 14 instances the apparently fraudulently winning party is the “Democratic Party,” in 11 instances it is

---

5For a description of the method see “approach two” described at http://www.umich.edu/~wmebane/efslides.pdf

6See the original R output files wrkef2a_Korea2020dAC_1d.Rout and wrkef2a_Korea2020daAC_1d.Rout in Korea2020ef2.zip for the numbers of fraudulent votes at each aggregation unit.
Table 2: Korea 2020 eforesnics Estimated Fraudulent Vote Counts

(a) Democratic Party specification fraudulent counts

<table>
<thead>
<tr>
<th></th>
<th>Observed Counts</th>
<th>95% interval</th>
<th>99.5% interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Voters Valid Votes</td>
<td>Manufactured</td>
<td>Total</td>
</tr>
<tr>
<td></td>
<td>43961157 28588899 14343693</td>
<td>1030561.8 996221.3 1067862.3 505242.4 1081627.3</td>
<td>1382523.7 1348340.1 1415145.3 1023134.9 1430344.3</td>
</tr>
</tbody>
</table>

(b) constituency leader specification fraudulent counts

<table>
<thead>
<tr>
<th></th>
<th>Observed Counts</th>
<th>95% interval</th>
<th>99.5% interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Voters Valid Votes</td>
<td>Manufactured</td>
<td>Total</td>
</tr>
<tr>
<td></td>
<td>43961157 28588899 16118788</td>
<td>1105458.2 1064714.5 1146934.4 553215.4 1159670.8</td>
<td>1404918.7 1363474.7 1445172.4 1028382.0 1457197.6</td>
</tr>
</tbody>
</table>

Note: observed counts and total fraud posterior means and credible intervals based on eforesnics model estimates. n = 18801.

7The particular constituencies that have these conditions can be identified by matching constituencies sequentially in “list of winners” tables available from [http://info.nec.go.kr/main/showDocument.xhtml?electionId=0020200415&topMenuId=EP&secondMenuId=EPEIO1](http://info.nec.go.kr/main/showDocument.xhtml?electionId=0020200415&topMenuId=EP&secondMenuId=EPEIO1) (as of May 9, 2020 17:12 EST). Province constituency-sequence-number (party posterior mean fraudulent): Chung-cheong bukdo 2 (DP 5321.9 fraudulent), Chungcheongnam-do 4 (FIP 3115.4 fraudulent), Chungcheongnam-do 5 (FIP 2574.6 fraudulent), Chungcheongnam-do 6 (FIP 2554.8 fraudulent), Chungcheongnam-do 1 (DP 5381.3 fraudulent), Gyeongsangnam-do 15 (DP 7573.1 fraudulent), Gyeongsangnam-do 5 (FIP 3697.5 fraudulent), Seoul 43 (FIP 6627.6 fraudulent), Seoul 48 (DP 11947.6 fraudulent), Seoul 45 (FIP 6240.1 fraudulent), Seoul 46 (FIP 7214.7 fraudulent), Seoul 4 (FIP 5166.0 fraudulent), Busan 7 (DP 2385.7 fraudulent), Busan 3 (FIP 5138.3 fraudulent), Busan 8 (DP 4055.7 fraudulent), Busan 12 (DP 2854.2 fraudulent), Busan 15 (FIP 6176.0 fraudulent), Daegu Metropolitan City 8 (I 4372.3 fraudulent), Incheon Metropolitan City 3 (I 1176.3 fraudulent), Incheon Metropolitan City 5 (DP 4056.5 fraudulent), Daejeon 7 (DP 4505.1 fraudulent), Daejeon 2 (DP 4458.2 fraudulent), Gyeonggi-do 8 (FIP 6255.3 fraudulent), Gyeonggi-do 9 (DP 8463.6 fraudulent), Gyeonggi-do 27 (DP 4187.1 fraudulent), Gyeonggi-do 49 (DP 4538.8 fraudulent), Gyeonggi-do 44 (DP 9679.8 fraudulent). In the “list of winners” tables, as translated by Google Translate, the constituency winner is
Given two specifications, which one is better? Probably neither model is correct, strictly speaking, even beyond the generality that no model is ever correct, but some are useful. If frauds only ever benefit the Democratic Party, then those frauds may have induced apparent frauds when we constrain frauds to benefit only constituency-leading candidates, because many of these do not affiliate with the Democratic Party. Similarly if only constituency-leading candidates benefit from frauds, then eforensics may be producing misleading results when we constrain frauds to benefit only the Democratic Party. Or perhaps other candidates—or several in each constituency—benefit from frauds and both specifications are producing misleading results. Possibly, of course, there are no frauds and something else is going on.

Caveats are many. The most basic caution is to keep in mind that “frauds” according to the eforensics model may or may not be results of malfeasance and bad actions. If some normal political situation makes the apparently fraudulent aggregation units appear fraudulent to the eforensics model and estimation procedure, then the frauds estimates may be signaling that “frauds” occur where in fact something else is happening. In particular there maybe something benign that leads many of the pre-vote units to have a turnout and vote choice distribution that differs so much especially from the distribution for election-day voting post units, the latter comprising the bulk of the data. Likewise something benign may distinguish the election-day voting post units that the eforensics model identifies as fraudulent. Beyond that general caution, there may something about the particular data used for the analysis that triggers the “fraud” findings—for instance, the vote totals in the data for constituency-leading candidates do not always match totals reported in “lists of winners.” And there may be something about the model specification that should be improved that would produce different results.

designated as associated with (“Party Name”) “Democratic Party” (DP), “Future Integration Party” (FIP) or “Independent” (I).
2 Election Forensics Toolkit and Spikes

I use the Election Forensics Toolkit (EFT, a website developed as part of a USAID-funded project) (Hicken and Mebane 2015; Mebane 2015) to look at features of the data. The EFT results add to the impression that the election results are manipulated.

Results for five tests (see Hicken and Mebane 2015 for explanations of the tests) computed using the entire set of aggregation units all together appear in Table 3. The DipT statistics for Turnout shows there is significant multimodality, a result that matches what can be seen visually in Figure 1. The P05s statistic for Turnout is significantly below the expected value of .2: this does not match the excessively high value of P05s that occurs in the case of signalling in Russia (Kalinin and Mebane 2011; Kalinin 2017), but it is difficult to think of natural processes that would produce frequencies of percentages that end in 0 or 5 that are too low. The 2BL statistics differ significantly from the expected value of 4.187, but the values that occur for the candidates’ votes match what we observe given strategic turnout and voting with multiparty competition (Mebane 2013a), so these 2BL statistics do not support a diagnosis that there are frauds.

Table 3: Distribution and Digit Tests, Korea 2020

<table>
<thead>
<tr>
<th>Name</th>
<th>2BL</th>
<th>LastC</th>
<th>P05s</th>
<th>C05s</th>
<th>DipT</th>
<th>Obs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turnout</td>
<td>4.117</td>
<td>4.508</td>
<td>0.175</td>
<td>0.203</td>
<td>0</td>
<td>18801</td>
</tr>
<tr>
<td>Democratic Party</td>
<td>4.129</td>
<td>4.473</td>
<td>0.198</td>
<td>0.2</td>
<td>0.993</td>
<td>18801</td>
</tr>
<tr>
<td>Constituency Leader</td>
<td>4.102</td>
<td>4.514</td>
<td>0.199</td>
<td>0.202</td>
<td>1</td>
<td>18801</td>
</tr>
</tbody>
</table>

Note: statistics and tests based on aggregation unit 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; “C05s,” mean of variable indicating whether the last digit of the vote count is zero or five. “Obs,” number of aggregation unit observations. Values in parentheses are 95% nonparametric bootstrap confidence intervals. Point estimates in red differ significantly from the values expected if there are no anomalies.

Given the SMD election rules, an approach that potentially produces sharper insights
regarding the constituency contests is to compute the EFT statistics separately for each constituency. The counterbalancing concern is statistical power: overall there are \( n = 18801 \) aggregation unit observations, but the median size of constituencies is 66 aggregation units with sizes ranging from a minimum of 38 units to a maximum of 183 units. In most cases with such sample sizes bootstrap confidence intervals for the EFT statistics (Hicken and Mebane 2015; Mebane 2015) are too wide to support finding significant differences from the values of the statistics that are expected if there are no anomalies.

Nonetheless Figure 6 shows that many constituencies have significantly anomalous values for the 2BL, LastC and P05s statistics. The plots of the 2BL statistic show several values that are either too big or too small to explain as results of electors’ strategic behavior (Mebane 2013a). The LastC statistics, motivated by Beber and Scacco (2012), show many too-large or too-small values. The P05s statistics show a few constituencies with excessively high values, in line with the usual understanding of how the percentages are often used to signal, but many more have significantly small values. Moreover both Turnout and the votes for the candidates exhibit many significantly anomalous P05s statistics. The many significant P05s statistics suggest the data are artificial (cf. Mebane 2013b).

---

8The C05s statistics, not shown, exhibit similar frequencies of significant anomalies. The significance test results used to color points in Figure 6 are obtained by checking whether the nonanomalous expected values are contained in 95% confidence intervals obtained using bootstrap methods. The EFT software is not designed to support finding exact \( p \)-values that might be used to implement adjustments for multiple testing.

9A qualification to note is that LastC does not exclude aggregation units that have counts less than 100. Beber and Scacco (2012) advocate excluding such counts. In the Democratic-Party-focused data, 1463 units have fewer than 100 votes for the Democratic Party candidate, with a median of one such unit per constituency (max 83). For the constituency-leader-focused data, 603 units have fewer than 100 votes for the constituency leading candidate, with a median of one such unit per constituency (max 22).
Note: statistics and tests based on aggregation unit observations, analyzed by constituency. Constituencies are matched to numbers in the Appendix. “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. Red points differ significantly at level $\alpha = .05$ from the values expected if there are no anomalies.
The spikes model tests for deviations in the proportions of votes for candidates in a more general way than does the P05s test, relative to a flexible and empirically grounded null distribution (Rozenas 2017). Figure 7 shows graphics that identify the ranges of polling stations the model estimates are fraudulent. Vertical gray bars in the graph indicate which aggregation units have suspicious votes: the aggregation units with vote proportions for (a) the Democratic Party or (b) the constituency leader that match the highlighted vote shares are suspicious. The results reinforce the findings for P05s in Table 3 and Figure 6 in that Figure 7 shows an excess of proportions near (a) .05 or (b) .07 and .38. The spike at 1.0 in Figure 7(b) matches the significantly high values of P05s in Figure 6(i), for constituency-leading candidates, but the spike at 1.0 in Figure 7(a) is not matched by the P05s findings for Democratic Party candidates. The spikes test in this case appears to be more sensitive, in that the P05s test is restricted to reporting only a single average value while the spikes test assesses an entire distribution.

Figure 7: Korea 2020 Spikes Tests Plots

(a) Democratic Party  (b) Constituency Leader

Note: vote shares are the proportion of valid votes for all (a) Democratic Party or (b) constituency leading candidates in each aggregation unit. n = 18801.
3 Conclusion

Taken together the eforensics estimates and EFT and spikes tests exhibit anomalies that strongly suggest the Korea 2020 legislative election data were fraudulently manipulated. “Such conclusions are always subject to the caveat that apparent frauds may really be consequences of strategic behavior, but that ambiguity can sometimes be mitigated by exploiting a multiplicity of statistics.... An election fraud will not necessarily trigger all of the statistics and tests, but we think a genuine fraud will in general set off many of them” (Hicken and Mebane 2015, 39).

Statistical findings such as are reported here should be followed up with additional information and further investigation into what happened. The statistical findings alone cannot stand as definitive evidence about what happened in the election.
List of constituencies

1, Busan 1; 2, Busan 10; 3, Busan 11; 4, Busan 12; 5, Busan 13; 6, Busan 14; 7, Busan 15; 8, Busan 16; 9, Busan 17; 10, Busan 18; 11, Busan 2; 12, Busan 3; 13, Busan 4; 14, Busan 5; 15, Busan 6; 16, Busan 7; 17, Busan 8; 18, Busan 9; 19, Chung-cheong bukdo 1; 20, Chung-cheong bukdo 2; 21, Chung-cheong bukdo 3; 22, Chung-cheong bukdo 4; 23, Chung-cheong bukdo 5; 24, Chung-cheong bukdo 6; 25, Chung-cheong bukdo 7; 26, Chung-cheong bukdo 8; 27, Chungcheongnam-do 1; 28, Chungcheongnam-do 10; 29, Chungcheongnam-do 11; 30, Chungcheongnam-do 2; 31, Chungcheongnam-do 3; 32, Chungcheongnam-do 4; 33, Chungcheongnam-do 5; 34, Chungcheongnam-do 6; 35, Chungcheongnam-do 7; 36, Chungcheongnam-do 8; 37, Chungcheongnam-do 9; 38, Daegu Metropolitan City 1; 39, Daegu Metropolitan City 10; 40, Daegu Metropolitan City 11; 41, Daegu Metropolitan City 12; 42, Daegu Metropolitan City 2; 43, Daegu Metropolitan City 3; 44, Daegu Metropolitan City 4; 45, Daegu Metropolitan City 5; 46, Daegu Metropolitan City 6; 47, Daegu Metropolitan City 7; 48, Daegu Metropolitan City 8; 49, Daegu Metropolitan City 9; 50, Daejeon 1; 51, Daejeon 2; 52, Daejeon 3; 53, Daejeon 4; 54, Daejeon 5; 55, Daejeon 6; 56, Daejeon 7; 57, Gangwon-do 1; 58, Gangwon-do 2; 59, Gangwon-do 3; 60, Gangwon-do 4; 61, Gangwon-do 5; 62, Gangwon-do 6; 63, Gangwon-do 7; 64, Gangwon-do 8; 65, Gwangju 1; 66, Gwangju 2; 67, Gwangju 3; 68, Gwangju 4; 69, Gwangju 5; 70, Gwangju 6; 71, Gwangju 7; 72, Gwangju 8; 73, Gyeonggi-do 1; 74, Gyeonggi-do 10; 75, Gyeonggi-do 11; 76, Gyeonggi-do 12; 77, Gyeonggi-do 13; 78, Gyeonggi-do 14; 79, Gyeonggi-do 15; 80, Gyeonggi-do 16; 81, Gyeonggi-do 17; 82, Gyeonggi-do 18; 83, Gyeonggi-do 19; 84, Gyeonggi-do 2; 85, Gyeonggi-do 20; 86, Gyeonggi-do 21; 87, Gyeonggi-do 22; 88, Gyeonggi-do 23; 89, Gyeonggi-do 24; 90, Gyeonggi-do 25; 91, Gyeonggi-do 26; 92, Gyeonggi-do 27; 93, Gyeonggi-do 28; 94, Gyeonggi-do 29; 95, Gyeonggi-do 3; 96, Gyeonggi-do 30; 97,

Constituencies can be identified fully by matching constituencies sequentially using “list of winners” tables available from http://info.nec.go.kr/main/showDocument.xhtml?electionId=0020200415&topMenuId=EP&secondMenuId=EPEI01.
Gyeonggi-do 31; 98, Gyeonggi-do 32; 99, Gyeonggi-do 33; 100, Gyeonggi-do 34; 101, Gyeonggi-do 35; 102, Gyeonggi-do 36; 103, Gyeonggi-do 37; 104, Gyeonggi-do 38; 105, Gyeonggi-do 39; 106, Gyeonggi-do 4; 107, Gyeonggi-do 40; 108, Gyeonggi-do 41; 109, Gyeonggi-do 42; 110, Gyeonggi-do 43; 111, Gyeonggi-do 44; 112, Gyeonggi-do 45; 113, Gyeonggi-do 46; 114, Gyeonggi-do 47; 115, Gyeonggi-do 48; 116, Gyeonggi-do 49; 117, Gyeonggi-do 5; 118, Gyeonggi-do 50; 119, Gyeonggi-do 51; 120, Gyeonggi-do 52; 121, Gyeonggi-do 53; 122, Gyeonggi-do 54; 123, Gyeonggi-do 55; 124, Gyeonggi-do 56; 125, Gyeonggi-do 57; 126, Gyeonggi-do 58; 127, Gyeonggi-do 59; 128, Gyeonggi-do 6; 129, Gyeonggi-do 7; 130, Gyeonggi-do 8; 131, Gyeonggi-do 9; 132, Gyeongsangbuk-do 1; 133, Gyeongsangbuk-do 10; 134, Gyeongsangbuk-do 11; 135, Gyeongsangbuk-do 12; 136, Gyeongsangbuk-do 13; 137, Gyeongsangbuk-do 2; 138, Gyeongsangbuk-do 3; 139, Gyeongsangbuk-do 4; 140, Gyeongsangbuk-do 5; 141, Gyeongsangbuk-do 6; 142, Gyeongsangbuk-do 7; 143, Gyeongsangbuk-do 8; 144, Gyeongsangbuk-do 9; 145, Gyeongsangnam-do 1; 146, Gyeongsangnam-do 10; 147, Gyeongsangnam-do 11; 148, Gyeongsangnam-do 12; 149, Gyeongsangnam-do 13; 150, Gyeongsangnam-do 14; 151, Gyeongsangnam-do 15; 152, Gyeongsangnam-do 16; 153, Gyeongsangnam-do 2; 154, Gyeongsangnam-do 3; 155, Gyeongsangnam-do 4; 156, Gyeongsangnam-do 5; 157, Gyeongsangnam-do 6; 158, Gyeongsangnam-do 7; 159, Gyeongsangnam-do 8; 160, Gyeongsangnam-do 9; 161, Incheon Metropolitan City 1; 162, Incheon Metropolitan City 10; 163, Incheon Metropolitan City 11; 164, Incheon Metropolitan City 12; 165, Incheon Metropolitan City 13; 166, Incheon Metropolitan City 2; 167, Incheon Metropolitan City 3; 168, Incheon Metropolitan City 4; 169, Incheon Metropolitan City 5; 170, Incheon Metropolitan City 6; 171, Incheon Metropolitan City 7; 172, Incheon Metropolitan City 8; 173, Incheon Metropolitan City 9; 174, Jeju Special Self-Governing Province 1; 175, Jeju Special Self-Governing Province 2; 176, Jeju Special Self-Governing Province 3; 177, Jeollabuk do 1; 178, Jeollabuk do 10; 179, Jeollabuk do 2; 180, Jeollabuk do 3; 181, Jeollabuk do 4; 182, Jeollabuk do 5; 183, Jeollabuk do 6; 184, Jeollabuk do 7; 185,
Jeollabuk do 8; 186, Jeollabuk do 9; 187, Jeollanam-do 1; 188, Jeollanam-do 10; 189, Jeollanam-do 2; 190, Jeollanam-do 3; 191, Jeollanam-do 4; 192, Jeollanam-do 5; 193, Jeollanam-do 6; 194, Jeollanam-do 7; 195, Jeollanam-do 8; 196, Jeollanam-do 9; 197, Sejong Special Self-governing City 1; 198, Sejong Special Self-governing City 2; 199, Seoul 1; 200, Seoul 10; 201, Seoul 11; 202, Seoul 12; 203, Seoul 13; 204, Seoul 14; 205, Seoul 15; 206, Seoul 16; 207, Seoul 17; 208, Seoul 18; 209, Seoul 19; 210, Seoul 2; 211, Seoul 20; 212, Seoul 21; 213, Seoul 22; 214, Seoul 23; 215, Seoul 24; 216, Seoul 25; 217, Seoul 26; 218, Seoul 27; 219, Seoul 28; 220, Seoul 29; 221, Seoul 3; 222, Seoul 30; 223, Seoul 31; 224, Seoul 32; 225, Seoul 33; 226, Seoul 34; 227, Seoul 35; 228, Seoul 36; 229, Seoul 37; 230, Seoul 38; 231, Seoul 39; 232, Seoul 4; 233, Seoul 40; 234, Seoul 41; 235, Seoul 42; 236, Seoul 43; 237, Seoul 44; 238, Seoul 45; 239, Seoul 46; 240, Seoul 47; 241, Seoul 48; 242, Seoul 49; 243, Seoul 5; 244, Seoul 6; 245, Seoul 7; 246, Seoul 8; 247, Seoul 9; 248, Ulsan Metropolitan City 1; 249, Ulsan Metropolitan City 2; 250, Ulsan Metropolitan City 3; 251, Ulsan Metropolitan City 4; 252, Ulsan Metropolitan City 5; 253, Ulsan Metropolitan City 6.
References


