

Inappropriate Applications of Benford's Law
Regularities to Some Data from the 2020 Presidential
Election in the United States*

Walter R. Mebane, Jr.[†]

November 10, 2020

*Several queries prompted me to write this.

[†]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).

As vote counting is drawing to a close in the 2020 presidential election in the United States, some¹ are claiming that application of Benford's Law to the precinct vote counts from a few counties and cities give evidence of election fraud. **The displays shown at those sources using the first digits of precinct vote counts data from Fulton County, GA, Allegheny County, PA, Milwaukee, WI, and Chicago, IL, say nothing about possible frauds.** Using data provided at https://github.com/cjph8914/2020_benford and obtained from <https://county.milwaukee.gov/EN/County-Clerk/Off-Nav/Election-Results/Election-Results-Fall-2020>, which are reportedly the data used to produce the displays, I'll here briefly try to explain what's going on in the data.

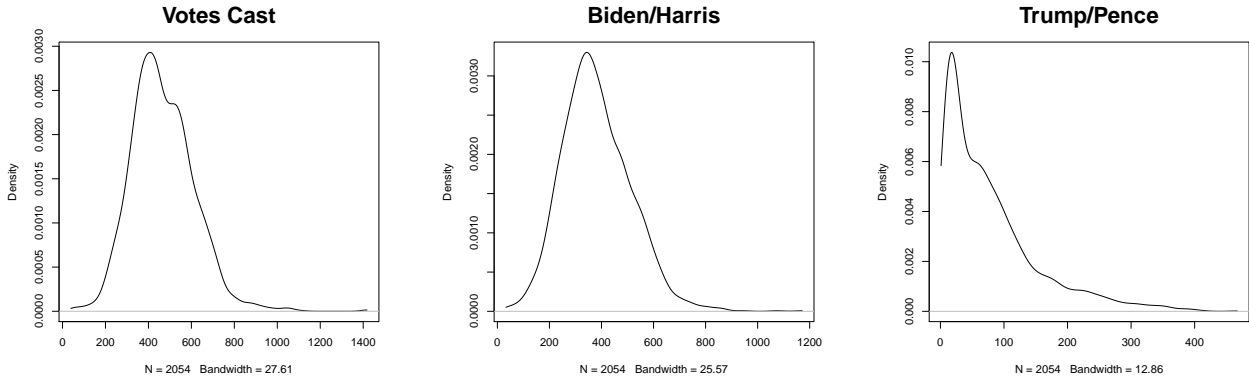
It is widely understood that the first digits of precinct vote counts are not useful for trying to diagnose election frauds. See for example the discussion in Carter Center (2005) and Pericchi and Torres (2011). The first digit is largely determined by the number of voters in each precinct, as usually—and especially in small jurisdictions such as individual cities and counties—the share of the votes received by parties or candidates does not vary all that greatly across precincts. Consider for example the densities in Figure 1 for votes from Chicago. The Biden/Harris ticket on average received a proportion of about .82 of the votes, and the Figure shows that the shape of the Biden/Harris vote count distribution pretty closely mirrors the shape of the distribution of votes cast. Trump/Pence received on average about .17 of the votes, but the Trump/Pence vote count distribution has a couple of hitches: for low counts the distribution reflects that in many precincts Trump/Pence vote counts are single digits.

Clearly the first digits of the Biden/Harris counts will most frequently be 3, 4 or 5. That non-Benford's Law pattern simply reflects the distribution of precinct sizes (presuming turnout did not vary that much across the city),² given the strong support for

¹E.g., <https://gnews.org/534248/>, https://www.youtube.com/watch?v=GLdPRwvwc2Y&feature=youtu.be&ab_channel=Nyar and https://twitter.com/Harrison_of_TX/status/1324536420992733185.

²Registered voter counts are not readily available.

Figure 1: Chicago 2020 Presidential Vote Distributions



Note: plots of empirical densities.

Biden/Harris across the whole city. The first-digit distribution has nothing whatsoever to do with any kind of election fraud.

The vote counts from the four jurisdictions are not final, so one should treat them cautiously. Nonetheless preliminary analysis shows little that suggests there are problems. For instance, consider results from the Election Forensics Toolkit (Hicken and Mebane 2015; Mebane 2015). For Milwaukee (Figure 2) and Allegheny County (Figure 3), no statistic exhibits a value that is statistically distinguishable from the value usually thought to occur in an election in which nothing unusual happens. For Fulton County (Figure 4) a few statistics have statistically distinguishable values, but these are probably benign: the DipT results are due to bimodality that is evident in the densities shown in Figure 5—some precincts lean slightly in favor of Trump/Pence while more overwhelmingly favor Biden/Harris. The P05s value for Dem.Total.Votes might be a concern, but notice that the P05s value for Rep.Total.Votes nearly exceeds the expected value of .2 as well. For Chicago (Figure 6) P05s is slightly elevated for Biden_Harris, and C05s is slightly elevated for Trump_Pence: the differences from the expected values of .2 are small. The 2BL test (based on the second digits and Benford's Law digit probabilities, (Mebane 2014)) shows second-digit means that differ significantly from 4.187 for both Biden_Harris and

Trump_Pence: the Trump_Pence result is perfectly compatible with nonstrategic votes (see Mebane 2013, Figure 2), while the result for Biden_Harris is harder to explain—but it matches results observed in some German elections (Mebane 2013, Figure 22) that are generally not considered to be problematic.

Figure 2: Milwaukee 2020 Presidential EFT

11/8/2020

Milwaukee2020p_EFT.html

x

Level	Candidate's Name	_2BL	LastC	P05s	C05s	DipT	Obs
National	Turnout	3.962	4.331	0.211	0.211	0.982	475
		(3.716, 4.211)	(4.077, 4.573)	(0.176, 0.247)	(0.174, 0.249)	--	
National	Joseph.R..Biden...Kamala.D..Harris	4.353	4.395	0.227	0.197	0.376	475
		(4.095, 4.603)	(4.138, 4.657)	(0.19, 0.264)	(0.159, 0.228)	--	
National	Donald.J..Trump...Michael.R..Pence	4.287	4.613	0.181	0.218	0.432	475
		(4.034, 4.55)	(4.343, 4.872)	(0.144, 0.215)	(0.184, 0.255)	--	

Note: results from Election Forensics Toolkit (Hicken and Mebane 2015).

For two jurisdictions there is information about the number of registered voters at each precinct, so I can use `eforensics` (Ferrari, McAlister, Mebane and Wu 2019; Mebane 2019*b,a*, 2020) to estimate the occurrence of what the `eforensics` model considers to be frauds. For Milwaukee the probabilities of no fraud, incremental fraud and extreme fraud are, respectively, .9797, .0180 and .0023, and for Allegheny County the probabilities are .9839, .0154 and .0006. These probabilities correspond to estimates of 1003.1 `eforensics`-fraudulent votes in Milwaukee and 599.3 in Allegheny County—negligible amounts that probably neither reflect nor result from bad acts.

Figure 3: Allegheny County 2020 Presidential EFT

11/8/2020

Allegheny2020p_EFT.html

X

Level	Candidate's Name	_2BL	LastC	P05s	C05s	DipT	Obs
National	Turnout	4.334	4.435	0.191	0.198	1	1323
		(4.179, 4.488)	(4.28, 4.581)	(0.169, 0.212)	(0.175, 0.221)	--	
National	DEM.Total.Votes	4.34	4.506	0.208	0.221	0.519	1323
		(4.185, 4.492)	(4.339, 4.675)	(0.185, 0.231)	(0.198, 0.244)	--	
National	REP.Total.Votes	4.122	4.458	0.197	0.205	0.388	1323
		(3.957, 4.287)	(4.305, 4.621)	(0.174, 0.218)	(0.183, 0.228)	--	

Note: results from Election Forensics Toolkit (Hicken and Mebane 2015).

Final verdicts regarding the elections in these and other jurisdictions should await the production of completed vote counts and should draw on additional information about election processes that go beyond mere vote count data. To date I've not heard of any substantial irregularities having occurred anywhere, and the particular datasets examined in this paper give essentially no evidence that election frauds occurred.

Figure 4: Fulton County 2020 Presidential EFT

11/8/2020

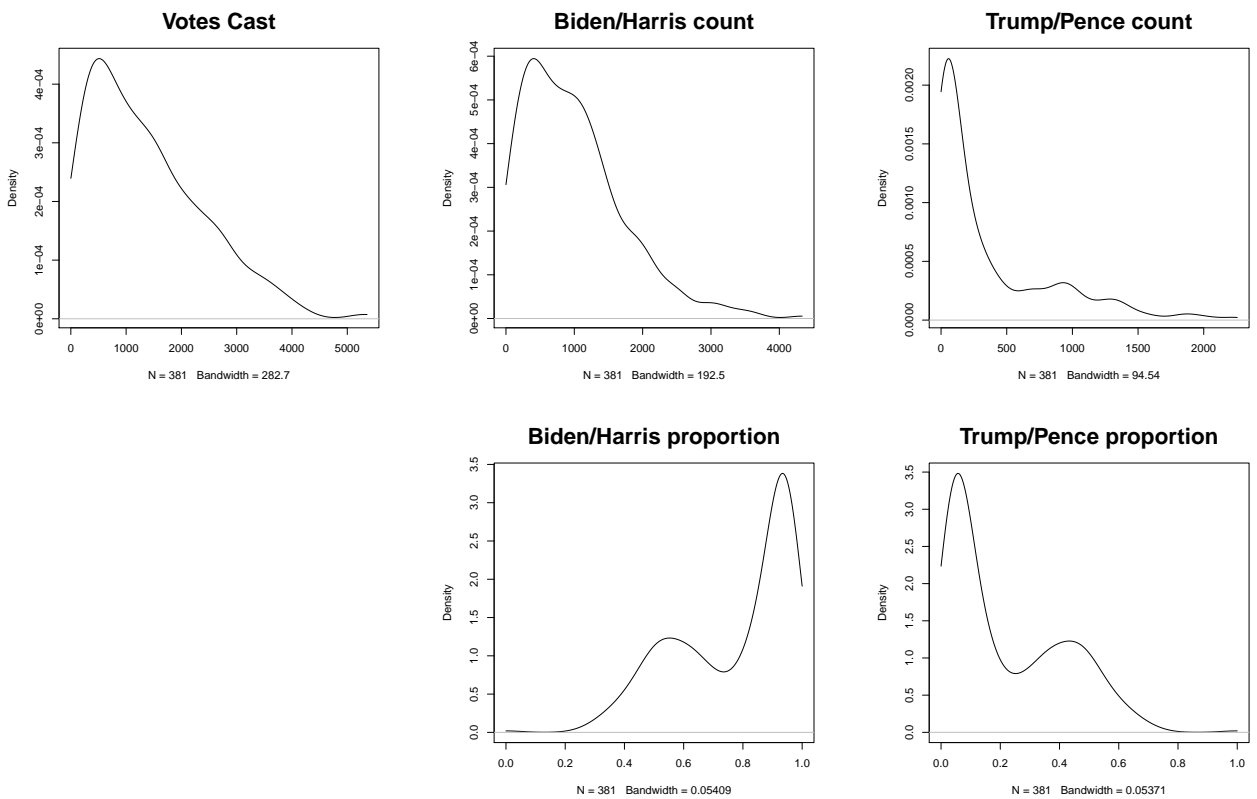
Fulton2020p_EFT.html

x

Level	Candidate's Name	_2BL	LastC	P05s	C05s	DipT	Obs
National	Rep.Total.Votes	4.278	4.234	0.241	0.213	0.038	381
		(3.996, 4.584)	(3.942, 4.512)	(0.197, 0.283)	(0.171, 0.252)	--	
National	Dem.Total.Votes	3.947	4.591	0.262	0.194	0.031	381
		(3.663, 4.221)	(4.307, 4.874)	(0.218, 0.307)	(0.155, 0.236)	--	

Note: results from Election Forensics Toolkit (Hicken and Mebane 2015).

Figure 5: Fulton County 2020 Presidential Vote Distributions



Note: plots of empirical densities.

Figure 6: Chicago 2020 Presidential EFT

11/8/2020

Chicago2020p_EFT.html

x

Level	Candidate's Name	_2BL	LastC	P05s	C05s	DipT	Obs
National	Biden_Harris	4.521	4.506	0.222	0.185	0.134	2054
		(4.39, 4.642)	(4.383, 4.623)	(0.203, 0.24)	(0.168, 0.204)	--	
National	Trump_Pence	3.895	4.392	0.192	0.221	0.912	2054
		(3.772, 4.025)	(4.271, 4.516)	(0.175, 0.209)	(0.202, 0.238)	--	

Note: results from Election Forensics Toolkit (Hicken and Mebane 2015).

References

- Carter Center. 2005. “Observing the Venezuela Presidential Recall Referendum: Comprehensive Report.”
- Ferrari, Diogo, Kevin McAlister, Walter R. Mebane, Jr. and Patrick Wu. 2019. “Election Forensics Package (eforensics).” **R** package, https://github.com/UMeforensics/eforensics_public.
- 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.” <http://www.umich.edu/~wmebane/USAID15/guide.pdf>.
- 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. 2014. Can Votes Counts’ Digits and Benford’s Law Diagnose Elections? In *The Theory and Applications of Benford’s Law*, ed. Steven J. Miller. Princeton: Princeton University Press pp. 206–216.
- Mebane, Jr., Walter R. 2015. “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.” <http://www.umich.edu/~wmebane/USAID15/report.pdf>.
- Mebane, Jr., Walter R. 2019a. “Evidence Against Fraudulent Votes Being Decisive in the Bolivia 2019 Election.” working paper, <http://www.umich.edu/~wmebane/Bolivia2019.pdf>.
- Mebane, Jr., Walter R. 2019b. “Notes Regarding Use of the **R** package `eforensics`: for POLSCI 485 in Fall 2019.” working paper, class notes.
- Mebane, Jr., Walter R. 2020. “Anomalies and Frauds in the Korea 2020 Parliamentary Election, SMD and PR Voting with Comparison to 2016 SMD.” working paper, <http://www.umich.edu/~wmebane/Korea2020.pdf>.

[//www.umich.edu/~wnebane/Korea2020.pdf](http://www.umich.edu/~wnebane/Korea2020.pdf).

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.