Voter Fraud!? | A Reimagining of Precinct Matching for Analyzing Absentee Ballots in the 2020 Presidential Election

Sean J. Birch

2023-04-24

Special Thanks to Dr. Jonathan Cervas for his sponsorship, guidance, and enthusiastic support.

Introduction

Allegations of voter fraud in both Allegheny County, Pennsylvania and Fulton County, Georgia arose after the results of the 2020 Presidential Election were released. This paper investigates the existence of statistical evidence using tangential precinct matching and linear regression. To compute the differences between election results across county lines, precincts on the border of Allegheny County, Pennsylvania, and Fulton County, Georgia were analyzed.

Allegheny County, Pennsylvania

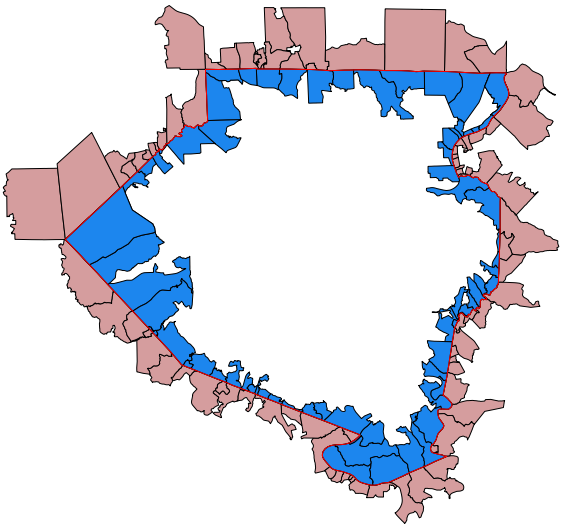


Figure i

Fulton County, Georgia

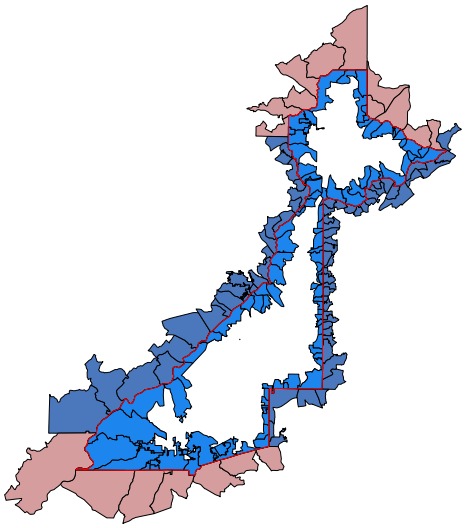


Figure ii

The results of the 2020 Presidential Election were subsetted for precincts tangential to the county border (which is shown by the **red** outline on figures i & ii). Precincts within the county border were defined as *interior* whereas precincts outside the county border were defined as *exterior*. Both counties of interest voted democrat, and their (*interior*) border precincts are, therefore, shaded in **dodger blue** (shown in figures i & ii). Border precincts outside the county of interest are shaded in **gray-blue** if they are within a county that voted democrat or in **pastel red** if they are within a county that voted republican, and are denoted *democratic exterior* and *republican exterior* precincts, respectively.

Data

Voter data for the 2020 Presidential Election was acquired individually at the precinct level for the following counties (where D denotes a Democratic county, and R denotes a Republican county):

Allegheny (D), Beaver (R), Butler (R), Washington (R), and Westmoreland (R) County, Pennsylvania.

Fulton (D), Cobb (D), Douglas (D), Gwinnett (D), DeKalb (D), Clayton (D), Cherokee (R), Forsyth (R), Carroll (R), Coweta (R), and Fayette (R) County, Georgia.

Multiple racial variables were collected from the Decennial Census for the year 2020 and were combined into a single variable:

- *Minority* - Percentage of population 18 and older with race 100% black or African American, Asian or Hispanic or Latino $\left(\frac{\text{P3\_004N+P3\_006N+P4\_002N}}{\text{P3\_001N}}\right)$.

Other variables of interest were collected from the 2020 five-year American Community Survey (ACS). Missing entries were replaced with a random value drawn from a normal distribution with a mean equal to the median and a standard deviation equal to the standard deviation of the data.

The following variables were pulled from the five-year American Community Survey (ACS) for the year 2020:

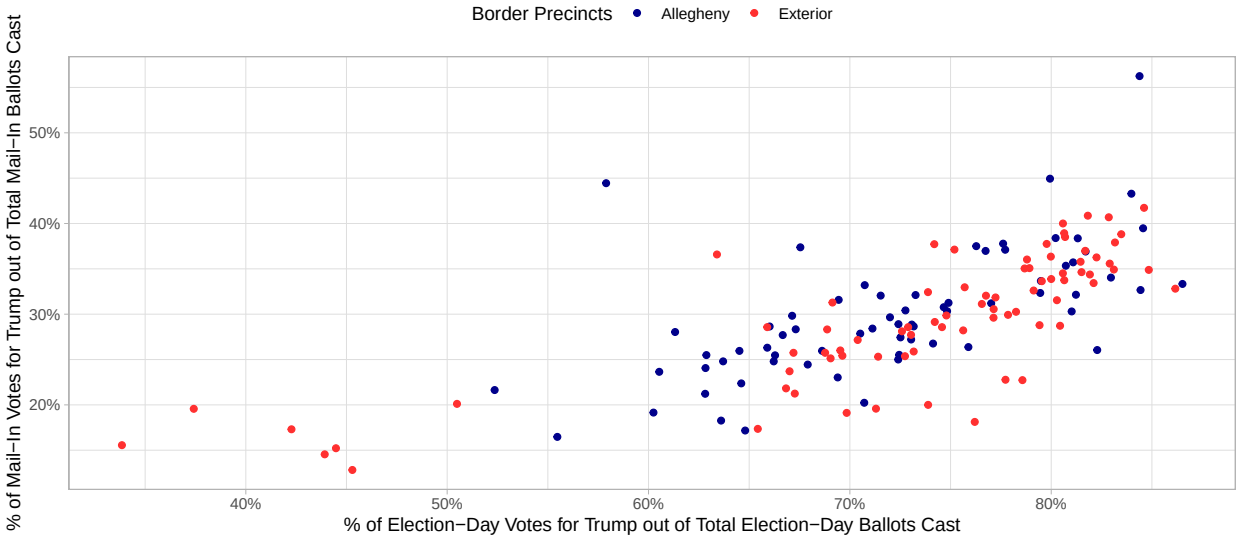
- *Internet* - Percentage of population without internet access $\left(\frac{\text{B28002\_013}}{\text{B28002\_001}}\right)$.
- *Age* - Population median age (B01002_001).
- *Income* - Median household income in the past 12 months for 2020 (B19013_001).
- *Vehicle* - Percentage of population without access to a vehicle $\left(\frac{B25045\_012+B25045\_0032}{B25045\_001}\right)$
- *School* - Percentage of population 3 years and over enrolled in school $\left(\frac{B14007\_002}{B14007\_001}\right)$

The socioeconomic variables were aggregated from block or block group to precinct level by assigning the total number of individuals who reported the variable under investigation weighted by the area of the block or block group that intersects the precinct. Proportion variables were calculated after performing this aggregation.

Unpaired Precinct Analysis

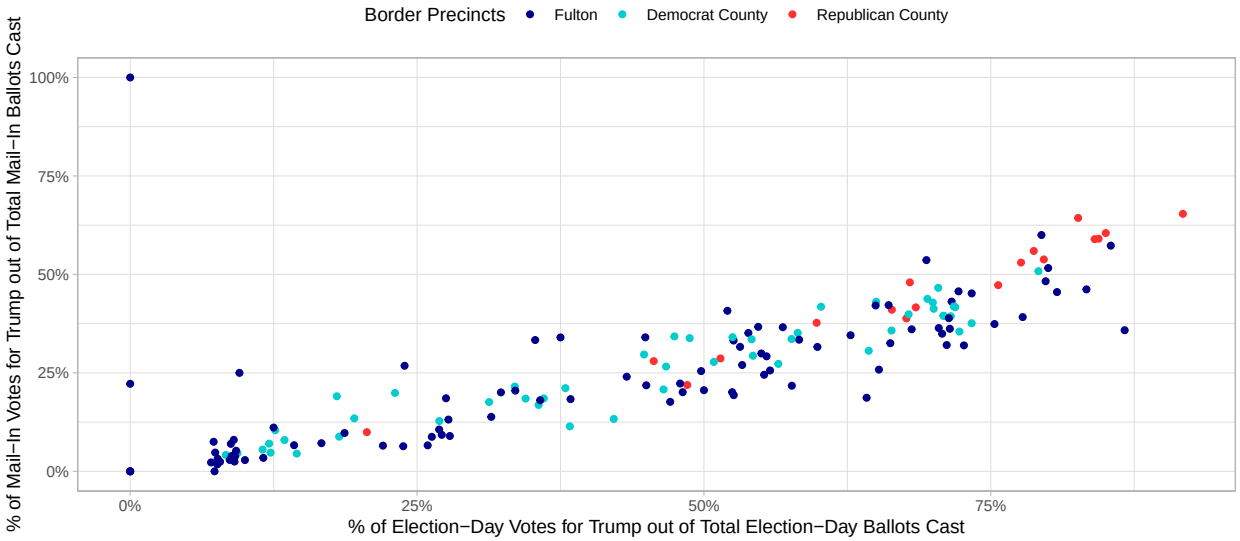
As seen in Figure 1, the precincts in Allegheny County appear to closely intermingle with the exterior precincts. Allegheny precincts maintain a continuous distribution; however, exterior precincts appear to exhibit a discontinuity for the Republican precincts stretching from approximately 50 to 65% of Trump's in-person vote. As the alleged fraud was in Allegheny County, this discontinuity does not appear to be indicative of the alleged voter fraud.

Figure 1: Trump % Absentee Vs Election−Day Vote; Allegheny & Surrounding County Border Precincts
2020 Presidential Election

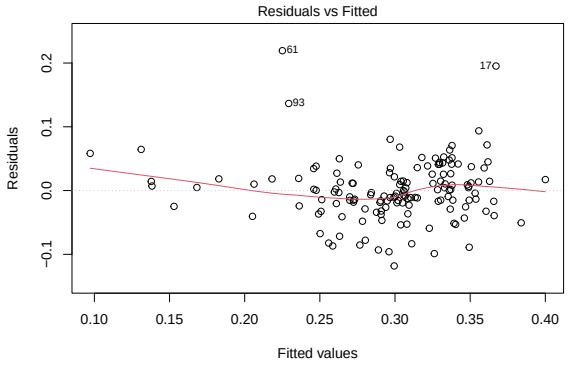


For Fulton County, the Republican County precincts appear to have, on average, a higher absentee share for Trump than the Democratic precincts. Although insufficient to prove voter fraud, compared to the distribution for Allegheny, Figure 2 appears more consistent with this hypothesis.

Figure 2: Trump % Absentee Vs Election−Day Vote; Fulton & Surrounding County Border Precincts
2020 Presidential Election



Results of the linear regression shown in Table 1 indicate insufficient statistical evidence to conclude (at a 95% level) that precincts across the border of Allegheny County had a higher absentee vote share for Trump, when controlling for socioeconomic variables. Furthermore, an ANOVA test failed to reject the null hypothesis that including the socioeconomic variables resulted in a more predictive model ($F_{6,138} = 1.47, p = 0.19$). Therefore, it cannot be concluded that such variables were relevant in analyzing voter fraud. As seen in the residuals plot, heteroskedasticity does not seem to be enough of an issue to invalidate the normality assumptions required for an ANOVA test to supply accurate results.



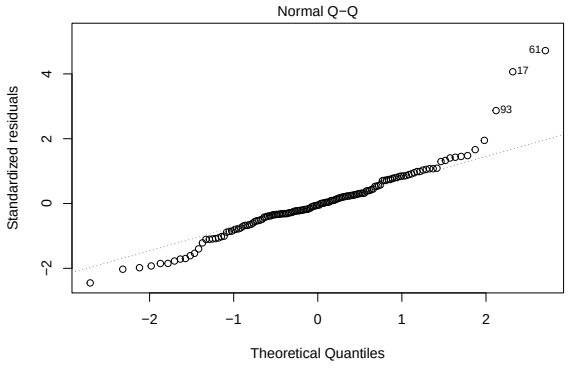
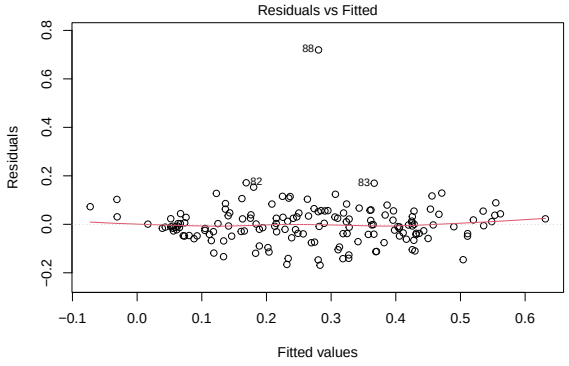


Table 1: Allegheny, PA Unpaired Model

	Dependent variable:
	Trump Absentee Vote Share
Trump Election-Day Vote Share	0.514^{***}
	(0.073)
Exterior Precinct	-0.014^{*}
	(0.009)
Minority	−0.102
	(0.087)
Age	−0.0001
	(0.0003)
Income	0.00000
	(0.00000)
Internet	0.010
	(0.076)
Vehicle	0.152^{*}
	(0.084)
School	−0.001
	(0.084)
Intercept	−0.080
	(0.057)
Observations	147
R^2	0.535
Adjusted R^2	0.508
Residual Std. Error	0.049 (df = 138)
F Statistic	19.817^{***} (df = 8; 138)
Note:	*p<0.1; **p<0.05; ***p<0.01

Results of the linear regression shown in Table 2 indicate sufficient statistical evidence to conclude (at a 95% level) that precincts across the border of Fulton County had a higher absentee vote share for Trump, even when controlling for numerous relevant socioeconomic variables. An ANOVA test found that there was statistical evidence that these socioeconomic variables resulted in a more predictive model ($F_{6,151} = 6.22, p \approx 0$). As seen in the residuals plot, heteroskedasticity does not seem to be enough of an issue to invalidate the normality assumptions required for an ANOVA test to supply accurate results.



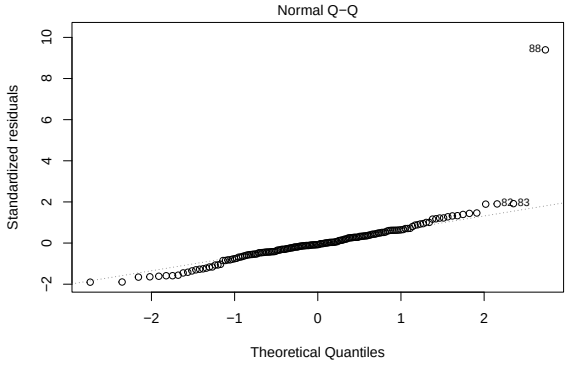


Table 2: Fulton, GA Unpaired Model

	Dependent variable:
	Trump Absentee Vote Share
Trump Election-Day Vote Share	0.374^{***}
	(0.057)
Democrat Exterior Precinct	0.009
	(0.018)
Republican Exterior Precinct	0.067^{**}
	(0.028)
Minority	-0.222^{***}
	(0.048)
Age	0.0004
	(0.0003)
Income	-0.00000^{**}
	(0.00000)
Internet	0.138
	(0.127)
Vehicle	−0.084
	(0.125)
School	0.344^{***}
	(0.101)
Intercept	0.113^{**}
	(0.055)
Observations	161
R^2	0.737
Adjusted R^2	0.721
Residual Std. Error	0.091 (df = 151)
F Statistic	46.902^{***} (df = 9; 151)
Note:	*p<0.1; **p<0.05; ***p<0.01

Precinct Matching Analysis

Precinct matching analysis assumes that tangential precincts should not be significantly different socioeconomically. In this analysis a comparison is performed between the difference in precincts across the border in different counties and the difference between border precincts within the county in question and the inner tangential precincts within that same county.

Allegheny County, Pennsylvania

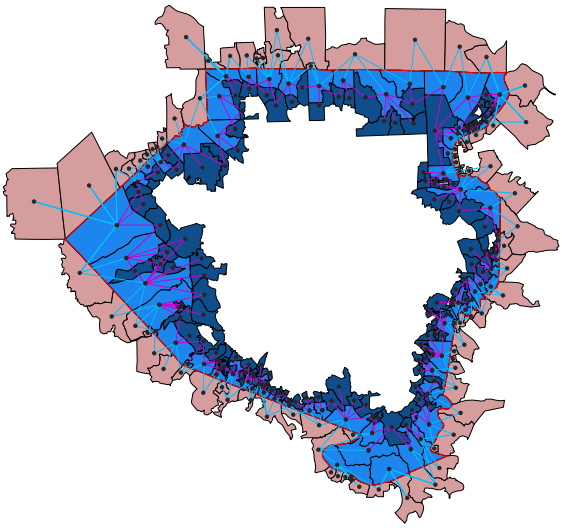


Figure iii

Fulton County, Georgia

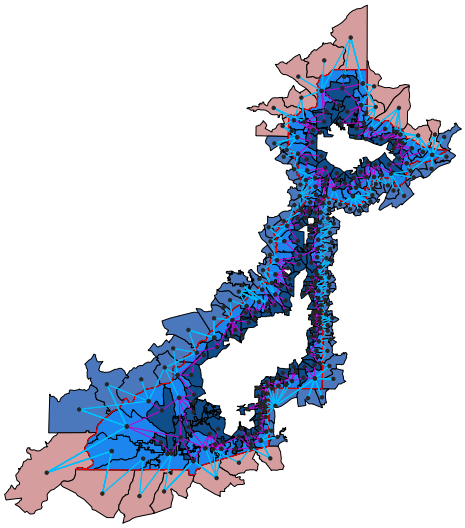
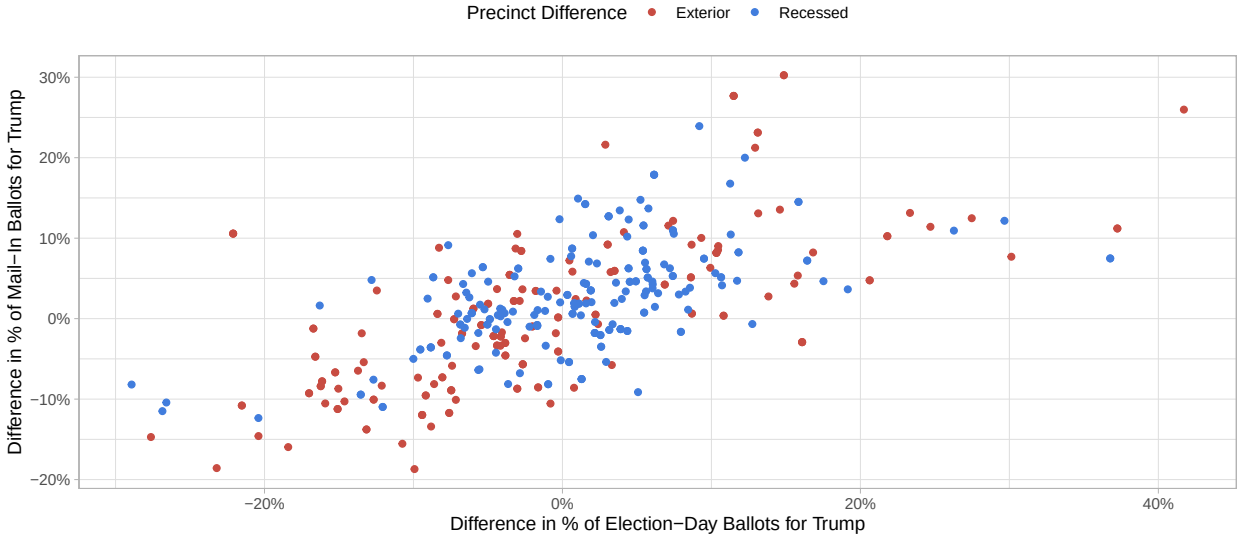


Figure iv

As seen in figure iii, the county border is again outlined in **red**, (*interior*) border precincts (which are all democratic) are shaded in **dodger blue**, border precincts outside the county of interest are shaded in **gray-blue** if they are within a county that voted democrat or in **pastel red** if they are within a county that voted republican (denoted *democratic exterior* and *republican exterior* precincts, respectively), and the newly added recessed (democratic) precincts are shaded in **dark blue**. The difference between precinct pairs are plotted with **neon blue** and **neon purple** lines to denote exterior and recessed pairs, respectively. Difference calculations are consistently performed by subtracting from the border precincts residing in the county under investigation. Therefore, in the plots (of figures 3 and 4), the region to the left of 0% on the x-axis indicates that the proportion in the county under investigation (i.e., Allegheny or Fulton) was lower; and the region to the right of 0% indicates that the proportion in county under investigation (i.e., Allegheny or Fulton) was higher. For simplicity, differences between the county under investigation (i.e., Allegheny or Fulton) and the exterior border precinct will be referred to as the exterior difference; and differences between the county under investigation and the recessed border precinct will be referred to as the recessed difference.

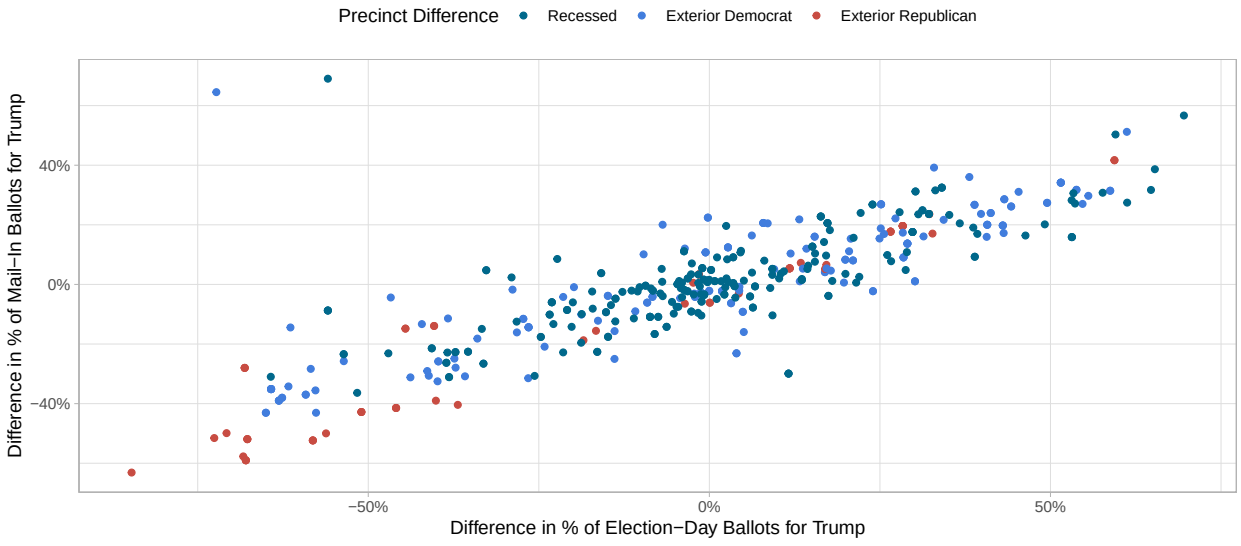
As seen in Figure 3, for Allegheny County there does not appear to be a significant difference in the relationship between Trump's absentee and election day vote shares for exterior and recessed precinct pairs, as their distribution appears substantially intermixed.

Figure 3: Difference in Trump Absentee Vs In−Person Share; Allegheny Exterior & Recessed Paired Precincts
Across Adjacent Precincts, 2020 Presidential Election



As seen in Figure 4, it appears that on average the exterior Republican precincts have a lower difference of absentee vote share for Trump than the Democratic or recessed precincts.

Figure 4: Difference in Trump Absentee Vs In−Person Share; Fulton Exterior & Recessed Paired Precincts
Across Adjacent Precincts, 2020 Presidential Election



The results of linear regression shown in Table 3 indicate insufficient statistical evidence to conclude (at a 95% level) that the difference in Trump's absentee vote share is different between exterior and recessed pairs for Allegheny County. An ANOVA test found that there was statistical evidence that the inclusion of socioeconomic variables resulted in a more predictive model ($F_{6,376} = 4.84, p \approx 0$). As seen in the residuals plot, heteroskedasticity does not seem to be enough of an issue to invalidate the normality assumptions required for an ANOVA test to supply accurate results.

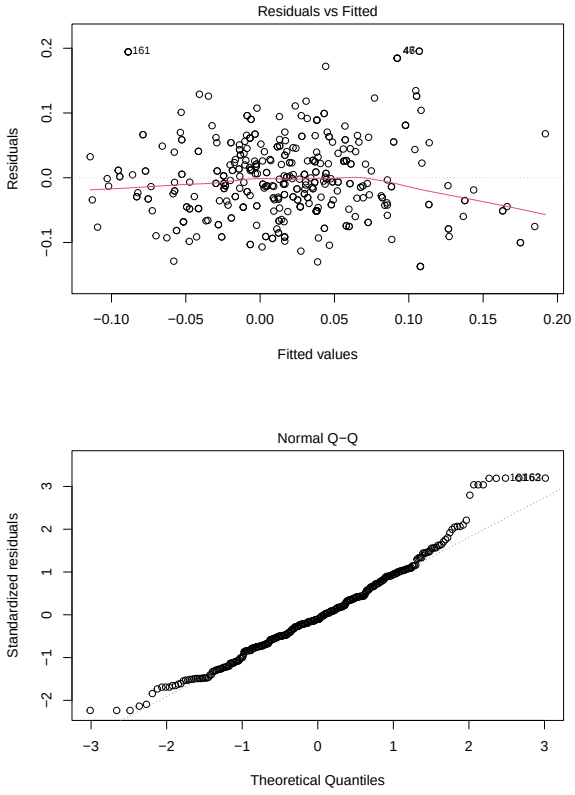
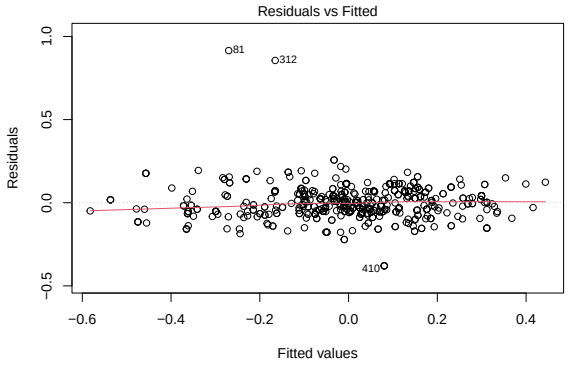


Table 3: Allegheny, PA Difference Model

	Dependent variable:
	Δ Trump Absentee Vote Share
Δ Trump Election-Day Vote Share	0.493^{***}
	(0.042)
Exterior Precinct	-0.007
	(0.007)
Δ Minority	-0.153^{***}
	(0.048)
Δ Age	-0.0002
	(0.0002)
Δ Income	0.00000
	(0.00000)
Δ Internet	-0.012
	(0.049)
Δ Vehicle	0.189^{***}
	(0.053)
Δ School	-0.022
	(0.049)
Intercept	0.021^{***}
	(0.005)
Observations	385
R^2	0.445
Adjusted R^2	0.433
Residual Std. Error	0.062 (df = 376)
F Statistic	37.678^{***} (df = 8; 376)
Note:	*p<0.1; **p<0.05; ***p<0.01

Results of the linear regression shown in Table 4 indicate sufficient statistical evidence to conclude (at a 95% level) that the difference in Trump's absentee vote share is different between Republican exterior and recessed pairs for Fulton County. However, there is not sufficient statistical evidence to conclude a difference in Trump's absentee vote share is different between Democrat exterior and recessed pairs for Fulton County. An ANOVA test found that there was statistical evidence that the inclusion of socioeconomic variables resulted in a more predictive model ($F_{6,462} = 9.14, p \approx 0$). As seen in the residuals plot, heteroskedasticity does not seem to be enough of an issue to invalidate the normality assumptions required for an ANOVA test to supply accurate results.



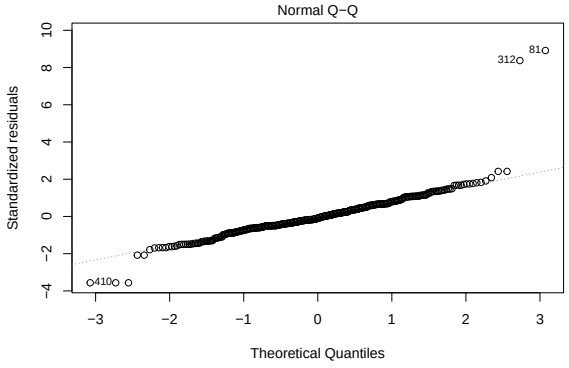


Table 4: Fulton, GA Difference Model

	Dependent variable:
	Δ Trump Absentee Vote Share
Δ Trump Election-Day Vote Share	0.505^{***}
	(0.031)
Democrat Exterior Precinct	0.009
	(0.011)
Republican Exterior Precinct	-0.047^{***}
	(0.017)
Δ Minority	-0.123^{***}
	(0.026)
Δ Age	0.0005^{***}
	(0.0001)
Δ Income	-0.00000^{***}
	(0.00000)
Δ Internet	0.143^{**}
	(0.059)
Δ Vehicle	−0.026
	(0.064)
Δ School	0.174^{***}
	(0.051)
Intercept	0.003
	(0.007)
Observations	472
R^2	0.755
Adjusted R^2	0.750
Residual Std. Error	0.107 (df = 462)
F Statistic	158.157^{***} (df = 9; 462)
Note:	*p<0.1; **p<0.05; ***p<0.01

Conclusion

This paper finds no significant evidence for voter fraud detectable by cross-county analysis for Allegheny County, Pennsylvania. There is statistically significant evidence that precincts from Republican Counties had a higher proportion of mail-in votes for Trump when controlling for socioeconomic variables. Additionally, there is statistical evidence that adding socioeconomic variables results in a more predictive model for Fulton County, Georgia. However, there is not enough evidence to claim the same for Allegheny County, Pennsylvania.

References

Lott, John R., Simple tests for the extent of vote fraud with absentee and provisional ballots in the 2020 US presidential election (December 21, 2020). Public Choice, forthcoming, Available at SSRN: https://ssrn.com/abstract=3756988 or http://dx.doi.org/10.2139/ssrn.3756988

Eggers, Andrew C, Haritz Garro and Justin Grimmer. 2021. "No evidence for systematic voter fraud: A guide to statistical claims about the 2020 election." Proceedings of the National Academy of Sciences 118(45).

Eggers and Justin Grimmer. 2022."Comment on 'Simple tests for the extent of vote fraud with absentee and provisional ballots in the 2020 US presidential election"' https://www.dropbox.com/s/ouxd7zinzv8l9o6/Fraud2.pdf?dl=0

Hlavac, Marek (2022). stargazer: Well-Formatted Regression and Summary Statistics Tables. R package version 5.2.3. https://CRAN.R-project.org/package=stargazer