Written Testimony for Ohio Redistricting Commission

August 27, 2021

Akron, OH

Witness: Dr. Casey Boyd-Swan

This commission is tasked with an important job: redrawing Ohio's House and Senate districts. While this job is one which must be completed every ten years, yours is a particularly unique one as this is the first redistricting to take place under new legislation approved by the people of Ohio. The language of this legislation appears relatively straightforward:

The Ohio redistricting commission shall attempt to draw a general assembly district plan that meets all of the following standards:

(A) No general assembly district plan shall be drawn primarily to favor or disfavor a political party.

(C) General assembly districts shall be compact.

Part B of this legislation is what my comments today will focus on interpreting – from the perspective of an econometrician or analyst. (B) The statewide proportion of districts whose voters, based on statewide state and federal partisan general election results during the last ten years, favor each political party shall correspond closely to the statewide preferences of the voters of Ohio.

Developing an analytic strategy as it pertains to determining the proportion of districts which correspond to each political party requires several decisions regarding the use of statewide state and federal partisan general election results from the previous 10 years. There are three strategies which I will explore today.

One strategy would be to use the population of eligible elections. More specifically, examining the proportion of voters which correspond to each political party across all statewide state and federal partisan general elections and create a general assembly district plan which aligns with these proportions. The proportion of voters aligning with each political party can be calculated in a variety of ways. A simple approach would be to average the proportions across all individual observations

(eligible elections); one drawback to this approach is that it assigns equal weight to each election distribution regardless of proximity to the present. This could be problematic if there were a significant shift in the state population in more recent years which alters the proportion of voter preferences by political party. There are two ways in which this could be addressed.

The first approach would be to narrow case selection to only those elections which most closely reflect current voter preferences. This approach would entail inclusion of only the most recent statewide state and federal partisan election distributions when calculating voter party preferences. In other words, instead of including all statewide state elections in the calculation of proportions, only the most recent election for each office would be used. This is relatively easy to calculate – the following estimates are calculated to only include the most recent statewide state and federal partisan election for each office; this reduces the sample to 7 elections – Governor (2018), Attorney General (2018), Auditor of State (2018), Secretary of State (2018), Treasurer (2018), President (2020), and US Senate (2018). The party-level average using this reduced sample estimates that 47.59% of voters indicate a preference for the Democratic party while 50.88% of voters indicate a preference for the Republican party. A more complicated approach would be to weight the elections based on the time proximity to the most recent census; this calculation approach could be legitimized as more reflective of the current state population. However, determining the most accurate weighting would require the creation of population models using data from non-census years (for example, annual American Community Survey data). Data collected during non-census years is drawn from representative samples, not the population, and may not perfectly represent population shifts, especially in non-urban geographic locations. More pressingly, the legislative language does not indicate that proximity to the current time is a criterion for data weighting or sample selection. Specifically, the legislation stipulates "based on statewide state and federal partisan general election results *during the last ten years* [emphasis added]", which indicates that elections in all years should be given identical weight to those in other years.

A second strategy would be to randomly sample the population of eligible elections. Random sampling strategies are typically used under the following circumstances: 1. Collection of data on the population of interest is not possible, oftentimes due to inability to accurately identify the individual members of said population; 2. Collection of data on the population of interest is not feasible, oftentimes due to the cost prohibitive nature of collecting data on all individuals of interest. Neither of these circumstances is salient in the current case. The population of eligible elections is already

observed and rather small in terms of the number of observations (N=16). Relying on random samples from the population would run the risk of creating significant bias in the estimated distributions. For example, in two random samples where 4 of the 16 elections were selected for use in calculating party proportions, significantly different party averages were estimated. In the first random draw, the Democratic candidates' average voter percentage was 42.63% while the Republican candidates' average voter percentage was 54.63%; in the second random draw, the Democratic candidates' average voter percentage was 47.23% while the Republican candidates' average voter percentage was 50.24%. Between the two random sample draws, the party-level difference is almost 5 percentage points for both political parties. This suggests that a random sampling approach would significantly bias the estimated distribution of voter party preferences.

A third strategy would be to target specific elections to include in the calculation of voter party preferences. For example, to use only gubernatorial elections as indicators of voter party preference. This is similar to the sampling approach discussed previously, where the goal was to narrow the sample to only the most recent of each office in statewide state and federal partisan elections. Justifying the targeted sample approach would be difficult. First, there is no legislative language to support such an approach that only certain state or federal elections are eligible for inclusion in the sample. Specifically, the legislation stipulates "based on *statewide state and federal partisan general election* [emphasis added] results during the last ten years", which indicates that elections in all statewide state and federal partisan elections should be given identical weight/consideration. Second, there is no methodological justification for excluding specific elections. For example, it is common for analysts to exclude observations from analyses in cases where the data was collected incorrectly or divergently to the way in which other data was collected. There is no reason to suspect this is the case in any of the 16 elections observed. Alternatively, in cases where analysts have too much data, it may be appropriate to set rules of exclusion/inclusion to narrow down the sample to fit the research question at-hand; setting these rules must completed prior to observation of the data in order to avoid "cherry-picking" observations that the analysts believe will reveal the estimates they are anticipating. Again, there does not appear to be any justification for excluding any of the 16 observed elections currently included in the analytic sample as each election observed was held for either a statewide state election or a federal partisan election.

There may be an argument to include additional elections aside from the 16 currently considered in the sample. Any additional elections would need to indicate voter party preference. I examine the

available election data posted on the Ohio Secretary of State <u>website</u> to determine if there may be missing elections that could be added to the analytic sample. In 2011 there were statewide elections, but none were explicitly partisan in nature (e.g. all were state issues up for consideration). In 2012 there were two federal partisan elections (one for President and one for US Senate), both of which are currently in the analytic sample. In 2013 there were no statewide state elections. In 2014 there were five statewide state elections (gubernatorial, Attorney General, Auditor of State, Secretary of State, State Treasurer), all of which were included in the current analytic sample. In 2015 there were no explicitly partisan elections (e.g. only state issues). In 2016 there were two federal partisan elections (one for President and one for US Senate), both of which are currently in the analytic sample. In 2017 there were no statewide state or federal partisan elections. In 2018 there were five statewide state elections (gubernatorial, Attorney General, Auditor of State, Secretary of State, State Treasurer) and one federal partisan election (US Senate), all of which were included in the current analytic sample. In 2019 there were no statewide state or federal partisan elections. In 2020 there was one federal partisan election (President), and it is included in the current analytic sample. The analytic sample total, based on this audit of all Ohio elections located in the Secretary of State's database, reveals that the total election count should equal 16; this is the number of observations currently in the analytic sample. The exercises completed above support the validity of the estimations presented below as being representative of the population of Ohio voters between the years 2011-2020.

Given the above outlined considerations, a relatively straightforward determination of voter preferences by political party could be completed by using all eligible elections from 2011-2020 and calculating averages based on individual election distributions. I present these data below (Table 1). Importantly, there are several distinct ways in which you can calculate averages. First, averages can be determined by summing the total count of voters for each political party affiliated candidate, then calculate the proportion of voters for each party (Calculation 1). This approach is problematic in that it ignores voters who voted for candidates not in one of the two main political parties. However, given the language in the legislation regarding redistricting, it may be deemed appropriate to calculate proportions in this manner.

Second, averages can be determined with averaging by subgroupings first. For example, the specific office that the election is being held for (Calculation 2), the year in which the election took place (Calculation 3), or the level of seat (state versus federal) that the election was intended to fill

(Calculation 4). The problem with this approach is that you are taking averages of averages, which means there is some data loss due to rounding. Third, averages can be determined by summing all proportions across all elections, then dividing by the total election count (Calculation 5). Regardless of approach, the proportions of voters aligning with each political party have a relatively small range (3.03 percentage points for Democratic party and 1.61 percentage points for Republican party); the average proportions by political party across all five types of presented analytic approaches are 44.35% Democratic and 53.59% Republican. It is important to note that these proportions do not sum to 100% since none of the approaches except Calculation 1 exclude the proportion of voters whose preferences were not for a candidate aligned with one of the two primary political parties.

I additionally perform a falsification check on shifting voter preferences by examining the proportion of voters by political party preferences over time (Figure 1). There does appear to be some volatility over time, but the plurality of this volatility is reflected in a significantly lower (higher) proportion of voters showing preferences for Democratic (Republican) candidates in 2014. In the most current time periods, there is a convergence of voter party preference, where voters on average have approximately equal preference between Democratic and Republican candidates.

In conclusion, I would like to offer some words of caution regarding the use of data and statistics as pieces of evidence for those making decisions on behalf of the public. Mark Twain popularized the phrase: "There are three kinds of lies: lies, damned lies, and statistics." Data and statistics in and of themselves are important tools for public policymakers to be able to make decisions which have a positive impact on the lives of those they are charged with serving. However, these same tools can be wielded in non-transparent and improper ways. Academics producing research, which is often used by policymakers to aid in decision-making, understand the importance of this – and it is one driving force behind the peer-review process. As you can probably discern from my testimony, an abundance of analytic choices sit behind any statistics you consume. The legitimacy of these rests on whether or not those analytic choices were made explicit and were justifiable. Hopefully I have demonstrated both of these characteristics in my discussion for you today. Thank you.

TABLE 1: Calculations by Analytic Approaches

	Democratic	Republican	Total
Calculation 1	45.90	54.09	99.99
Calculation 2	42.87	54.51	97.38
Calculation 3	44.41	52.90	97.31
Calculation 4	44.56	52.90	97.46
Calculation 5	44.00	53.55	97.55

Note: Calculation 1 uses voter count data by party. Calculations 2-4 average within subgroups first then across subgroups; Calculation 2 uses office, Calculation 3 uses years, and Calculation 4 uses level (state versus federal). Calculation 5 averages across individual election results.

FIGURE 1: Proportions by Party over time