Final Project Proposal: A Comparative Study of Apportionment Algorithms Using Real U.S. Census Data

Mingqi Zhang , Zhaohua zheng , Anqi Li
CSE 5205: Computation in Economics and Social Choice
Fall 2025

1 Research Topic / Problem Statement

Apportionment determines how many representatives each state receives when a fixed number of legislative seats must be assigned based on population. Although the problem seems straightforward, different algorithms can produce noticeably different allocations. In the context of social choice and fair representation, the apportionment method itself is often as consequential as the population data.

This project will empirically evaluate four classical apportionment algorithms: Hamilton (Largest Remainder), Jefferson (D'Hondt), Webster (Sainte-Laguë), and Huntington–Hill. Using the 2020 United States Census population data to allocate 435 seats in the House of Representatives, we aim to answer the following question:

> *How does the choice of apportionment rule affect representational fairness, and do some rules systematically advantage large or small states?*

We also study sensitivity: *How stable are seat allocations when population numbers change slightly (e.g., $\pm 0.5\%$)?* Instability would indicate vulnerability to census measurement noise, migration patterns, or deliberate statistical manipulation.

2 Related Work

Apportionment has been studied in political science, economics, and algorithmic social choice for more than two centuries. Hamilton's method (1792) satisfies the quota rule but is vulnerable to paradoxes such as the Alabama paradox. Divisor methods such as Jefferson and Webster resolve quota violations but introduce distinct patterns of rounding that may benefit large or small states. Huntington–Hill, adopted for the U.S. House in 1941, was designed to minimize representational variance.

Balinski and Young (1982) characterize paradoxes and prove impossibility results: no method satisfies monotonicity, quota, and population consistency simultaneously. Snyder and Ting (2005) empirically demonstrate size-related bias depending on the divisor method chosen. However, existing empirical work often focuses on two methods or a single census cycle. Our project contributes a systematic comparison across four algorithms using a recent and complete real dataset, complemented with a robustness analysis under synthetic perturbations.

3 Proposed Approach

We will download the 2020 U.S. Census state population dataset (state-level totals). Let P_i denote the population of state i and $S = 435$ the total number of seats.

Algorithms to Implement

- **Hamilton (Largest Remainder)**: Allocate $\lfloor S \cdot P_i / \sum_j P_j \rfloor$ seats, then distribute remaining seats based on fractional remainders.
- **Jefferson (D'Hondt)**: Use divisors $1, 2, 3, \ldots$; tends to favor larger states.
- **Webster (Sainte-Laguë)**: Use divisors $1, 3, 5, \ldots$; tends to be more proportional.
- **Huntington–Hill**: Uses geometric mean rounding; currently used for U.S. House apportionment.

Evaluation Metrics

For each method, we will compute:

- **Representation ratio:** $\rho_i = \frac{P_i/s_i}{\sum_j P_j/S}$ where s_i is the number of seats for state i.
- **Fairness deviation** (lower is more equal): $\mathrm{FD} = \sum_i |\rho_i - 1|$.
- **Size advantage score:** correlation between s_i deviation and state population rank.
- **Sensitivity analysis:** Apply $\pm 0.5\%$ random noise to populations and count seat changes.

Expected Deliverables

- A reproducible Python implementation of all four algorithms.
- Comparison tables of seat allocations and representation ratios.
- Visualizations (bar plots and heatmaps) of deviation across states and across methods.
- A written discussion of fairness, bias, and robustness.

4 Backup Plan

If time does not permit full sensitivity analysis, we will instead include a simplified robustness measure (e.g., seat flipping count under uniform noise). If acquiring state-level population data becomes difficult, we will switch to preprocessed datasets from Kaggle or GitHub that contain identical values. The core contribution of comparing apportionment rules remains unchanged under either contingency.

5 Expected Outcomes

We expect to find that:

- All methods satisfy proportionality in aggregate but differ significantly in distribution.
- Jefferson likely benefits large states; Webster and Huntington–Hill may benefit medium or small states.
- Hamilton may produce quota-respecting yet paradox-sensitive outcomes.
- Even small population perturbations may cause nontrivial seat flips for some rules.

The results will illustrate that algorithmic design choices—even within well-defined mathematical frameworks—have real policy implications for democratic representation.

6 References

- Balinski, M. L., and Young, H. P. *Fair Representation: Meeting the Ideal of One Man, One Vote.* Yale University Press, 1982.
- Huntington, E. V. "Mathematical theory of apportionment." *PNAS*, 1921.
- U.S. Census Bureau. 2020 State Population Totals. `https://www.census.gov`
- Snyder, J., and Ting, M. "Why Roll Calls?" *Journal of Political Economy*, 2005.
- Balinski, M. L. and Young, H. P. "The quota method of apportionment." *American Mathematical Monthly*, 1975.