An Excerpt of "The Ballot Order Effect in Taiwan's Legislative Elections, 2008–2020"

September 30, 2025
Ting-Chih Hung
GitHub Repo

2 Context and Data

2.1 Institutional Context

Since 2008, Taiwan's Legislative Yuan elections have operated under a mixed-member majoritarian system. This study focuses on the 73 single-member districts (SMDs), where legislators are elected via a first-past-the-post (FPTP) system. According to Article 34 of the Civil Servants Election and Recall Act, the ballot order for candidates in each district is determined by a public lottery held by the electoral commission. This procedure creates a natural experiment within each electoral district. Unlike the vertical ballots common in many Western countries, Taiwan's ballots are arranged horizontally, with candidates listed from left to right.

2.2 Data

Our data come from the official legislative election results published by the Central Election Commission (CEC) of Taiwan. The dataset contains detailed information for each candidate in every district from 2008 to 2020, including electoral district, ballot number, party affiliation, vote share, and incumbency status. Our treatment variable is whether a candidate is listed first on the ballot, and our outcome variable is the candidate's vote share.

3 Empirical Strategies

3.1 Notation and Setup

Let the index $k \in \{1, \dots, K\}$ denote an electoral district, where $K = 73$ for each election year. Within each district k, there are $n_{[k]}$ candidates, indexed by $i \in \{1, \dots, n_{[k]}\}$. For each candidate i in district k, we define a binary treatment indicator, $D_{[k]i}$, which equals 1 if the candidate is listed first on the ballot and 0 otherwise. The main outcome of interest is the candidate's vote share, denoted by $Y_{[k]i}$.

A naive approach would be to compare the average vote share of all candidates listed first with that of all candidates not listed first:

$$\hat{\tau}_{\text{naive}} = \frac{1}{n_1} \sum_{k=1}^{K} \sum_{i:D_{[k]i}=1} Y_{[k]i} - \frac{1}{n_0} \sum_{k=1}^{K} \sum_{i:D_{[k]i}=0} Y_{[k]i}, \tag{1}$$

where n_1 and n_0 are the total counts of treated and control candidates, respectively. This estimator, however, is biased because the number of candidates $n_{[k]}$ is a confounder that affects both the treatment assignment probability and the vote share distribution.

3.2 Conventional Approaches and Their Limitations

Two conventional methods address the confounding caused by district size. The first approach includes fixed effectss (FEs) for the number of candidates in a linear regression model:

$$Y_{[k]i} = \alpha + \beta_{\text{fe}} D_{[k]i} + \sum_{s=3}^{S} \gamma_s \mathbf{1}(n_{[k]} = s) + \varepsilon_{[k]i}, \tag{2}$$

where districts with two candidates serve as the baseline reference group.[1]

The second approach relies on inverse probability weighting (IPW), which creates a pseudo-population where the treatment assignment is independent of district size (Horvitz and Thompson, 1952; Hernán and Robins, 2020). It is convenient to implement IPW using a weighted linear regression model:

$$Y_{[k]i} = \alpha + \beta_{\text{ipw}} D_{[k]i} + \varepsilon_{[k]i}, \tag{3}$$

where each observation is weighted by the inverse of its treatment assignment probability:

$$w_{[k]i} = D_{[k]i} n_{[k]} + (1 - D_{[k]i}) \frac{n_{[k]}}{n_{[k]} - 1}.$$

While both approaches adjust for district size confounding, their causal interpretation requires the stable unit treatment value assumption (SUTVA) (Cox, 1958; Rubin, 1980). In our context, this assumption is likely violated since candidates within the same district compete for the same pool of votes, and vote shares must sum to 100%. When interference is present, the coefficients lose their interpretation as an average treatment effect (ATE), and conventional standard error estimates become invalid. These limitations motivate us to seek a design-based inference approach, which leverages the known randomization mechanism and does not rely on SUTVA.

3.3 Fisher Randomization Test

Instead of relying on model-based inference and questionable stability assumptions, we can exploit the known randomization mechanism to draw causal inferences via Fisher randomization test (FRT) (Fisher, 1935). We conceptualize the ballot order assignment process, as a stratified randomized experiment (SRE), where each electoral district constitutes a stratum. Within each district k, one candidate is randomly assigned, to treatment (listed first), and the remaining $n_{[k]} - 1$, candidates are assigned to control.

To formally define the FRT in our context, we first introduce the potential outcomes framework (Neyman, 1923; Rubin, 1974; Rosenbaum, 2007; Imbens and Rubin, 2015). Let $\mathbf{D}_{[k]} = (D_{[k]1}, \dots, D_{[k]n_{[k]}})$, be a random vector representing the treatment assignments, for all $n_{[k]}$ candidates in district k, and let $\mathbf{d}_{[k]}$ be a specific realization of $\mathbf{D}_{[k]}$. The entire set of possible ballot assignments, for district k is denoted by $\mathcal{D}_{[k]}$, defined as

$$\mathcal{D}_{[k]} = \left\{ \mathbf{d}_{[k]} \in \{0,1\}^{n_{[k]}} : \sum_{i=1}^{n_{[k]}} d_{[k]i} = 1 \right\}. \tag{4}$$

We denote the set of all possible assignments, across all districts by $\mathcal{D} = \mathcal{D}_{[1]} \times \cdots \times \mathcal{D}_{[K]}$. We define the potential outcome, $Y_{[k]i}(\mathbf{d}_{[k]})$, as the vote share that candidate i in district k would receive, if the ballot assignment were $\mathbf{d}_{[k]}$. This notation explicitly allows a candidate's

[1] The estimator $\widehat{\beta}_{\text{fe}}$ is a variance-weighted average of the difference-in-means estimators from each stratum, giving more influence to strata with larger variance in the treatment indicator.

vote share, to depend on which of their opponents received the treatment. These potential outcomes are constrained, by the institutional fact that vote shares must sum to one within a district:

$$\sum_{i=1}^{n_{[k]}} Y_{[k]i}(\mathbf{d}_{[k]}) = 1$$

for any valid assignment $\mathbf{d}_{[k]}$. Notice that the potential outcomes, are treated as fixed quantities, and the only source of randomness, comes from the random assignment of treatments. The observed outcome for candidate i in district k is then, $Y_{[k]i} = Y_{[k]i}(\mathbf{D}_{[k]})$.

The FRT tests the null hypothesis:

$$H_0 : \; Y_{[k]i}(\mathbf{d}_{[k]}) = Y_{[k]i}(\mathbf{d}'_{[k]}), \quad \forall k, \; i, \; \mathbf{d}_{[k]}, \; \mathbf{d}'_{[k]} \in \mathcal{D}_{[k]}, \tag{5}$$

which states that the ballot order has no effect (neither directly nor indirectly through interference) on any candidate's vote share. Under H_0, the SUTVA automatically holds because for each candidate, all potential outcomes are identical.

To conduct the FRT, we must choose a test statistic $T(\mathbf{D}, \mathbf{Y})$ a priori, where $\mathbf{D}$ and $\mathbf{Y}$ are treatment assignment and outcome vectors across all districts and candidates. For simplicity, we consider the weighted difference-in-means statistic:

$$T(\mathbf{D}, \mathbf{Y}) = \sum_{k=1}^{K} w_{[k]} \tau_{[k]},$$

where $\tau_{[k]}$ is the difference-in-means estimator within district k, and the weight

$$w_{[k]} = \frac{\frac{n_{[k]}}{n} \cdot \frac{1}{n_{[k]}} \cdot \frac{n_{[k]}-1}{n_{[k]}}}{\sum_{k'=1}^{K} \frac{n_{[k']}}{n} \cdot \frac{1}{n_{[k']}} \cdot \frac{n_{[k']}-1}{n_{[k']}}} = \frac{n_{[k]} - 1}{\sum_{k'=1}^{K} (n_{[k']} - 1)}$$

is a weight proportional to the variance of the treatment indicator in district k, which gives more influence to larger districts to improve efficiency.

Under SRE, the randomization distribution of $T(\mathbf{D}, \mathbf{Y})$ is uniform over $\mathcal{D}$. Since under H_0 the observed outcomes $\mathbf{Y}$ are fixed, we can simulate the null distribution of T by repeatedly drawing treatment assignments from $\mathcal{D}$ and computing the corresponding test statistic. The p-value, calculated as the proportion of simulated test statistics at least as extreme as the observed value, quantifies how surprising the observed data are under the null hypothesis. Given that $\mathcal{D}$ contains approximately 10^{52} possible assignments, we approximate the randomization distribution via Monte Carlo simulation with 10000 random draws. We also report the standard error of the p-value estimate to quantify the Monte Carlo approximation error.

4 Main Results

4.1 Conventional Estimates

Figure 1 shows estimated ballot order effects using naive comparison, fixed effects, and IPW approaches. Among the four election years, 2016 shows the largest estimated effect. While naive estimates suggest substantial effects, these disappear once we control for district size, yielding point estimates near zero with no statistical significance. However, these conventional methods rely on the questionable SUTVA assumption.

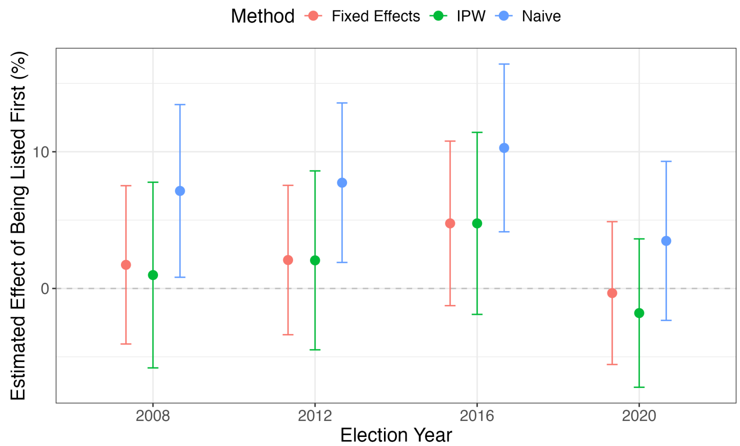

Figure 1: Estimated Ballot Order Effect by Election Year and Method.

Notes: The plot displays point estimates and 95% confidence intervals for the ballot order effect on vote share (%) using three different methods: naive comparison (blue), FE regression (red), and IPW regression (green). All standard errors are calculated using the HC2 variant of the Eicker-Huber-White (EHW) sandard errors.

4.2 Fisher Randomization Test Results

Table 1 presents the FRT results, including the observed test statistics, p-values, and standard errors of the p-value estimates. The results are valid regardless of potential interference, as the FRT does not require SUTVA. For all four election years, the null hypothesis of no ballot order effect cannot be rejected at the conventional 0.05 significance level.

Table 1: Randomization Inference Results for Ballot Order Effects.

Year	Test Statistic	P-value	P-value SE	Num. of Permutations
2008	1.7288	0.629	0.005	10000
2012	2.0801	0.535	0.005	10000
2016	4.7613	0.150	0.004	10000
2020	−0.3366	0.910	0.003	10000

Notes: P-values computed via randomization inference with 10000 Monte Carlo permutations. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Test statistic is the variance-weighted difference in means.

Figure 2 displays the randomization distributions of the test statistic for each election year. Each panel shows the null distribution of the weighted difference-in-means statistic, with the observed statistic indicated by a vertical red line. The p-value is the proportion of simulated statistics at least as extreme as the observed value. In most years, the observed statistics lie near the center of the null distribution, yielding large p-values. In 2016, however, the observed statistic is displaced toward the tail, with a two-sided p-value of about 0.15. This remains above conventional significance thresholds, so we do not reject the null.

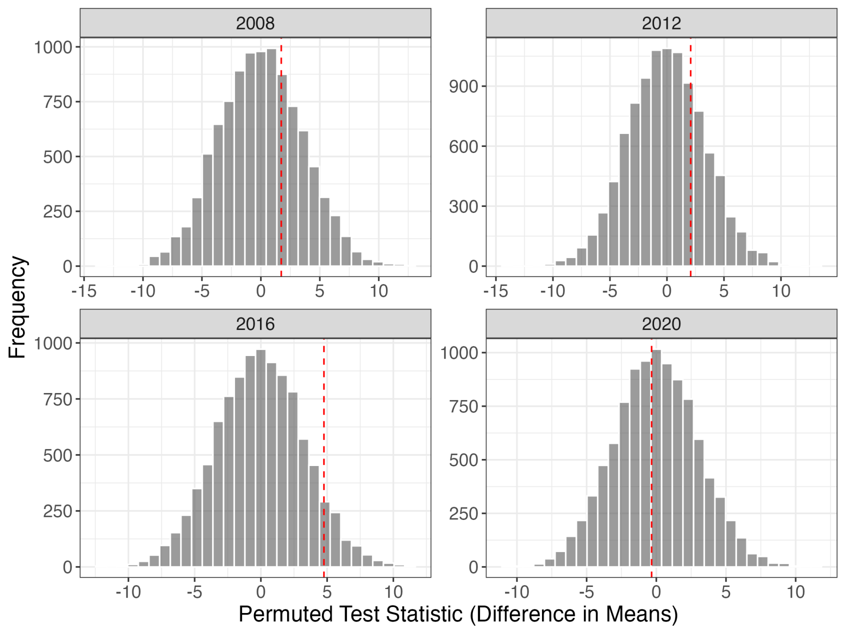

Figure 2: Randomization Distributions of the Test Statistic by Election Year.

Notes: The vertical red line indicates the observed test statistic.

References

Cox, D. R. (1958). *Planning of Experiments*. John Wiley and Sons.

Fisher, R. A. (1935). *The Design of Experiments*. Oliver and Boyd.

Hernán, M. A. and Robins, J. M. (2020). *Causal Inference: What If*. Chapman & Hall/CRC.

Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47(260):663–685.

Imbens, G. W. and Rubin, D. B. (2015). *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press.

Neyman, J. (1923). On the application of probability theory to agricultural experiments. essay on principles. section 9. reprinted edition.

Rosenbaum, P. R. (2007). Interference between units in randomized experiments. *Journal of the American Statistical Association*, 102(477):191–200.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701.

Rubin, D. B. (1980). Discussion of "Randomized analysis of experimental data: the fisher randomization test" by Basu D. *Journal of the American Statistical Association*, 75:591–593.