The Effect of Candidate List Position on Vote Share: Improving Internal and External Validity *

Kentaro Fukumoto†
Class of 2018, Fast Track Course ‡

January 4, 2018

Abstract

When a candidate is located at a good position (e.g., at the top) on the ballot, the candidate tends to garner larger vote share. It is not, however, straight-forward to identify position effects unless positions are determined exogenously. An election to Japanese House of Councilors offers an ideal setup because candidate positons are randomized across municipalities. Nonetheless, the stable unit treatment value assumption does not hold. Accoringly, this paper proposes a new estimand and a new standard error under a relaxed assumption. We also show that a few estimators have nice properties. Our analysis implies position effects such as the primacy and recency effects. We also show that these effects are stronger for non-viable candidates than for viable ones.

*Paper prepared for delivery at the 1st Annual Meeting of the Japanese Society for Quantitative Political Science, University of Tokyo, January 8–9, 2018, and the 5th Asian Political Methodology Meeting, Seoul National University, January 11–12, 2018. Earlier versions of this paper were presented at the Annual Meeting of the Japanese Association of Elecotral Studies, Kagawa University, May 20–21, 2017, and a workshop at Hokkaido University, March 9, 2017. Fukumoto gratefully acknowledges the financial support of the Japan Society for the Promotion of Science (16K13340) and Sakuradakai.

†Professor, Department of Political Science, Gakushuin University. Mailing address: 1-5-1 Mejiro, Toshima, Tokyo 171-8588, Japan. Telephone: +81-3-5904-9258. Fax: +81-3-5992-1006. Email address: `Kentaro.Fukumoto@gakushuin.ac.jp`. Corresponding author.

‡Shuhei Fujita, Yuki Hotsumi, Eri Igarashi, Rei Kawasaki, Chihiro Konjiki, Kentaro Kuramoto, Yuji Miura, Yuri Nakane, Hinako Saito, Motoki Sasaki, Hiroki Sato, Mami Sato, Rena Wada, Yuki Wanibe, Akane Yamaki, and Ryujiro Yorimitsu. Undergraduate students, Department of Political Science, Gakushuin University.

INTRODUCTION

Politicians, judges, and scholars have belived that, when a candidate is located at a good position (e.g., at the top) on the ballot, the candidate tends to garner larger vote share. Theoretically, this "primacy effect" is expected because of cognitive costs of voting, confirmation bias, satisficing, or "up means good" heuristics (Miller and Krosnick, 1998; Tourangeau, Couper, and Conrad, 2013).[1] Some studies argue that the last position is beneficial because voters remember the last and forget the first ("recency effect") (Marcinkiewicz, 2014; Miller and Krosnick, 1998), while others claim that candidates in the middle positions have larger vote share (Bagley, 1965; Miller and Krosnick, 1998).

Nonetheless, as Darcy and McAllister (1990) and Miller and Krosnick (1998) well summarize the literature, evidences on these position effects are mixed, partly because candidates' positions are not (well) randomized and, thus, their effects on vote share are estimated with bias (see also Ho and Imai, 2008). In this sense, the Japanese upper house elections, which we study here, offer the most ideal design: the order of the candidates' names are randomized for every municipality in a district, which is effectively stratified randomization (Imbens and Rubin, 2015).

We also argue that, in the previous studies, the stable unit treatment value assumption (Imbens and Rubin, 2015) (SUTVA) is not plausible and, accordingly, the estimand is not well defined. This paper proposes a new estimand by relaxing SUTVA. Moreover, we show that, under stratified randomization, differences-in-means and fixed effects estimators have nice proporties for this estimand. We also simplfy variance of the differences-in-means estimator and invent the corresponding variance estimator. In these ways, we intend to improve

[1] Meredith and Salant (2013) provide evidences which doubt satisficing as a cause of name position effect. An eye-tracking experiment (Blom-Hansen et al., 2016, 176) discovers "almost no systematic reading from top to bottom or from left to right." Thus, they suggest that "the most likely process behind order effects is spatial [rather than temporal]."

internal validity of identifying position effects.

This paper also aims to improve external validity of position effects analysis. As we detail later, the Japanese setup is not only different from the other countries' but also, theoretically, makes it both easy and hard for position effects to work and, thus, be detected. Our findings of position effects add a new evidence on the debate. We also discover that minor candidates enjoy stronger position effects than major ones, which is a controversial topic.

The composition of this paper is as follows. The next section elaborates on how we improve internal and external validity. In the following section, we introduce a new estimand, estimators, and their variance estimators and show their properties. The fourth section applies our framework to a dataset of Japan. Finally, we conclude.

GOAL

Internal Validity

In many countries, candidates' names are placed in the alphabetical order (mostly in British ex-colonies), in the way a party nominates (mostly under PR system), in the decreasing order of vote share in a past election, in the order for candidates to file, by incumbency status, or whether candidates belong to major parites (e.g., Krosnick, Miller, and Tichy, 2004, Table 4.1). In these cases, since there are probably confounders, we cannot be sure of identifying position effects without bias even with many covariates controlled.

Recently, many studies exploit two types of variation in name order as natural experiments. One is rotation. For instance, in a dozen of American states, the order of candidates' names is rotated across sub-district areas such as precincts so that one candidate comes at the top almost as frequently as any other (Krosnick, Miller, and Tichy, 2004, 53). However, since rotation is conducted in a deterministic way, rotation "is clearly not the same as random assignment" (Miller and Krosnick, 1998, 299).[2] Thus, there might still remain

[2] When the name order is rotated from ballots to ballot, it is almost the same as ballot

confounders even after controlling pretreatments.

The other way to change the name order is randomization. For example, in Australia, the order of candidates' names is randomized for each district (King and Leigh, 2009). Since the candidates' names are printed on the ballot in the same order for all voters in the district unlike rotation, and pretreatments can be unblanced by chance, analysits have to control candidate's characteristics in order to improve efficiency of estimation and avoid potential bias. In particular, when they include candidate fixed effects in panel data, they end up with only those candidates who run more than once, which might result into selection bias.

Both rotation and randomization are used in California. For every even number year, a new order of the 26 alphabets is randomly drawn and candidates' names are reordered accordingly in a sub-district area. Then, in the other sub-district areas, the order of candidates' names is rotated subsequently. A similar procedure is employed in South Korea as well (Jun and Min, 2017). These randomization-rotation procedures enjoy best of both worlds, though analysts can estimate candidate specific effects alone but not average treatment effect and must make much effort to obtain standard errors (Ho and Imai, 2006, 2008) (these shortcomings hold for non-randomized rotation as well). Moreover, "[w]ith a single randomization for a single major candidate, ballot order can be highly confounded with observed and unobserved district characteristics" (Ho and Imai, 2008, 227) (this pitfall is applied to non-rotated randomization as well).

As Ho and Imai (2008, 227–228) recommend, "[i]f the order were randomized in *each* [Assembly] *district* [for the statewide election], the correlation between ballot order and . . . any . . . covariate should be zero" (emphasis original). This most ideal setup is exactly what we study: (not single but) repeated randomization. That is, in the Japanese upper house elections, the order of the candidates' names are randomized for every municipality in

---

level randomization (Darcy 1986, 648, Krosnick, Miller, and Tichy 2004, 53, Orr 2002, 577). However, we cannot observe for whom each voter casts which type of ballot.

the district.[3] Therefore, we improve internal validity of analysis to identify position effects.

External Validity

The case of Japan extends the reach of the literature in three directions. First, we examine the candidate position not on the ballot but on the list displayed at the voting booth. To our knowledge, all of the previous studies focus on the candidate name order on the ballot. In Japan, voters write in a candidate's name on a blank ballot. Instead, at the voting booth, candidates' names are displayed on the list.[4]

Second, we expect and find the primacy effect for the candidate at the top right corner, not left. This is because, in some situation including candidate lists, Japanese read and write vertically from right columns to left columns. Our case exemplifies that where is prime depends on culture.

Third, we document existence of position effects under the single non-transferable vote (SNTV) system for the first time. Prior research have found the position effect under the electoral systems of single member districts,[5] single transferable votes,[6] multiple votes,[7] and open or flexible list proportional representation (PR).[8] These three ways, we improve external

[3] Public Officers Election Act, Article 175.

[4] Public Officers Election Act, Article 175.

[5] U.S. and U.K. (Bagley, 1965).

[6] Australia (King and Leigh, 2009), Ireland (Robson and Walsh, 1974), Malta, and U.K. (Curtice and Marsh, 2008).

[7] Block vote in the U.S. (Meredith and Salant, 2013), cumulative voting in the U.S. (Brockington, 2003) and Germany (Marcinkiewicz and Jankowski, 2014), and limited vote in Spain (Lijphart and Pintor, 1988). Note that some of these works study not national but local elections.

[8] Belgium (van Erkel and Thijssen, 2016; Geys and Heyndels, 2003), Czech (Marcinkiewicz and Stegmaier, 2015), Denmark (Blom-Hansen et al., 2016), Finland (Ortega Villodres,

validity of the position effect research.

We also call attention to the fact that the setup of the Japanese upper house election makes it both easy and hard for position effects to work and, thus, be detected. To begin, we look at facilitators. Under the SNTV system, intra-party competetion weakens party cue and strengthens position effects (cf. Alvarez, Sinclair, and Hasen, 2006; Ho and Imai, 2008; Miller and Krosnick, 1998; Pasek et al., 2014). In addition, the SNTV system has multiple member districts and, thus, tend to have larger number of candidates. In turn, the more candidates, the more cognitive cost voters must pay and, thus, the more likely position effects are to work (Brockington, 2003; Ho and Imai, 2008; Meredith and Salant, 2013).

Nonetheless, there are hurdles as well. Since the election we study is national and, thus, salient and since party affiliation is shown, voters are given much information about candidates as well as partisan cues and thus, do not have to rely on candidate positon as heuristic. In addition, since voting is not compulsory, politically unsofisticated citizens, who are susceptible to position effects (Miller and Krosnick, 1998; Ho and Imai, 2008), are less likely to show up.

METHOD

Setup

We index individual canididate and jurisdiction by i and j and their numbers by n and m, respectively; $i \in \{1, 2, \ldots, n\}, j \in \{1, 2, \ldots, m\}$. Jurisdiction j assigns position indices from one to n to n candidates without replacement. We denote the position of candidate i in

2003), Germany (Faas and Schoen, 2006), Poland (Marcinkiewicz, 2014), and Switzerland (Lutz, 2010).

jurisdiction j (hereafter, unit ij) by X_{ij}. It follows

$$\begin{aligned}\boldsymbol{X}_{\cdot j} &\equiv (X_{1j}, X_{2j}, \ldots, X_{nj}) \\ &\in \mathbb{X} \\ &\equiv \left\{ (x_{1\cdot}, x_{2\cdot}, \ldots, x_{n\cdot}) \middle| \forall x \in \{1, 2, \ldots, n\}, \sum_{i=1}^{n} I(x_{i\cdot} = x) = 1 \right\}\end{aligned}$$

where $I(\cdot)$ is a dummy variable to indicate whether the argument is true.

Generally, a treatment variable W_{ij} is a dummy variable dependent on X_{ij};

$$W_{ij} = w(X_{ij}) \in \{0, 1\}$$

For instance, the primacy treatment is defined by $W_{ij}^{(1)}$ where $W_{ij}^{(x)} = I(X_{ij} = x)$. Denote the numbers of treated and control units by

$$\sum_{i=1}^{n} W_{ij} \equiv n^T \in \{0, 1, \ldots, n\}$$
$$\sum_{i=1}^{n} (1 - W_{ij}) \equiv n^C = n - n^T,$$

respectively. In particular, when $W_{ij} = W_{ij}^{(x)}, x \in \{1, 2, \ldots, n\}$, it follows that $n^T = 1$.

Denote the (logged or not logged) vote share of unit ij by Y_{ij}.

Estimand

Importantly, the two components of the SUTVA are not plausible in our setup. First, the SUTVA would posit that, for every *treatment* vector $\boldsymbol{w} \equiv (\boldsymbol{w}_{\cdot 1}, \boldsymbol{w}_{\cdot 2}, \ldots, \boldsymbol{w}_{\cdot m})$, there is only one value of potential outcome, $y_{ij}(\boldsymbol{w})$. This is, however, unlikely to be true in our context. For instance, Y_{ij} tends to be larger when $X_{ij} = 2$ than when $X_{ij} = n$ (in both cases, $W_{ij}^{(1)} = 0$) if the earlier order tends to give a candidate the larger vote share. Instead, we make the following assumption:

Assumption 1 (Single Value). *For every assignment vector* $\boldsymbol{x} \equiv (\boldsymbol{x}_{\cdot 1}, \boldsymbol{x}_{\cdot 2}, \ldots, \boldsymbol{x}_{\cdot m})$*, there is only one value of potential outcome,* $y_{ij}(\boldsymbol{x})$*.*

Second, the SUTVA would require that the potential outcome of Y_{ij} is independent of the other candidates' positions: $y_{ij}(\boldsymbol{x}) = y_{ij}(\boldsymbol{x}')$ if $x_{ij} = x'_{ij}$. This is not, however, plausible in our setup. For instance, suppose candidates i and i' belong to the same party, but candidate i'' belongs to another party. Then, Y_{ij} will be smaller when $W^{(1)}_{i'j} = 1, W^{(1)}_{i''j} = 0$ than when $W^{(1)}_{i'j} = 0, W^{(1)}_{i''j} = 1$ (in both cases, $W^{(1)}_{ij} = 0$) if candidate i' is more likely to take away supporters of the same party from candidate i than candidate i'' is. Instead, we make the following assumption:

Assumption 2 (No Interference). *The potential outcome of* Y_{ij} *is independent of the candidates' positions in the other jurisdictions:*

$$y_{ij}(\boldsymbol{x}) = y_{ij}(\boldsymbol{x}') \quad \text{if } \boldsymbol{x}_{\cdot j} = \boldsymbol{x}'_{\cdot j}.$$

This assumption is reasonable because voters do not see candidate lists in the other jurisdictions.

Hereafter, we denote the potential outcome by $y_{ij}(\boldsymbol{x}_{\cdot j})$, not by $y_{ij}(w_{ij}) \in \{y_{ij}(0), y_{ij}(1)\}$ as under the SUTVA. Note that we cannot define a treatment effect by $y_{ij}(1) - y_{ij}(0)$. Instead, to begin, we define the marginal potential outcomes by

$$y^T_{ij} \equiv \frac{1}{|\mathbb{X}^T_{i\cdot}|} \sum_{\boldsymbol{x}_{\cdot j} \in \mathbb{X}^T_{i\cdot}} y_{ij}(\boldsymbol{x}_{\cdot j})$$
$$y^C_{ij} \equiv \frac{1}{|\mathbb{X}^C_{i\cdot}|} \sum_{\boldsymbol{x}_{\cdot j} \in \mathbb{X}^C_{i\cdot}} y_{ij}(\boldsymbol{x}_{\cdot j}),$$

where

$$\mathbb{X}^T_{i\cdot} \equiv \{\boldsymbol{x}_{\cdot j} | W_{ij} = 1, \boldsymbol{x}_{\cdot j} \in \mathbb{X}\}$$
$$\mathbb{X}^C_{i\cdot} \equiv \{\boldsymbol{x}_{\cdot j} | W_{ij} = 0, \boldsymbol{x}_{\cdot j} \in \mathbb{X}\}.$$

Then, we define the marginal treatment effect for unit ij by

$$\tau_{ij}^{\mathrm{MTE}} \equiv y_{ij}^{T} - y_{ij}^{C}$$

and the average marginal treatment effect (AMTE) by

$$\tau^{\mathrm{AMTE}} \equiv \frac{1}{mn} \sum_{j=1}^{m} \sum_{i=1}^{n} \tau_{ij}^{\mathrm{MTE}}.$$

We denote the super-population average marginal treatment effect by $\tau_{\mathrm{sp}}^{\mathrm{AMTE}}$.

Estimator

Now we presuppose stratified randomization (Imbens and Rubin, 2015), namely, that each jurisdiction randomly assigns its position vector independently of each other:

$$\Pr(\boldsymbol{X} = \boldsymbol{x} \in \mathbb{X}^m) = \prod_{j=1}^{m} \Pr(\boldsymbol{X}_{\cdot j} = \boldsymbol{x}_{\cdot j} \in \mathbb{X}) = \left(\frac{1}{|\mathbb{X}|}\right)^{m} = \frac{1}{n!^{m}}, \tag{1}$$

where

$$\mathbb{X}^m \equiv \{\boldsymbol{x} | \forall j \in \{1, 2, \ldots, m\}, \boldsymbol{x}_{\cdot j} \in \mathbb{X}\}.$$

Differnce-in-Means Estimator. We denote the difference-in-means estimator by

$$\hat{\tau}^{\mathrm{dif}} \equiv \frac{1}{mn^{T}} \sum_{j=1}^{m} \sum_{i=1}^{n} W_{ij} Y_{ij} - \frac{1}{mn^{C}} \sum_{j=1}^{m} \sum_{i=1}^{n} (1 - W_{ij}) Y_{ij}.$$

Generally, it is not guaranteed that $\hat{\tau}^{\mathrm{dif}}$ is unbiased for the average treatment effect under stratified randomization (Imbens and Rubin, 2015, 201). Nonetheless, in our design, since the numbers of treated and control candidates are equal to n^T and n^C in all jurisdictions, respectively, the following proposition holds:

Proposition 1 (Unbiasedness)**.**

$$\mathbb{E}(\hat{\tau}^{dif}) = \tau^{AMTE}.$$

Note that $\mathbb{E}(\cdot)$ is expectation operator with respect to $\boldsymbol{X}$. Proof of all propositions will be available in to-be appendix.

By contrast, variance estimator for $\hat{\tau}^{\mathrm{dif}}$ is not straight-forward. The variance of $\hat{\tau}^{\mathrm{dif}}$ is derived as

$$\mathbb{V}(\hat{\tau}^{\mathrm{dif}}) = \frac{1}{m^2}\sum_{j=1}^{m}\Big(\frac{S_j^T}{n^T} + \frac{S_j^C}{n^C} - \frac{S_j^{TC}}{n}\Big),$$

where

$$\begin{aligned} S_j^T &\equiv \frac{1}{n^T}\mathbb{V}\Big(\sum_{i=1}^{n} W_{ij}Y_{ij}\Big) \\ S_j^C &\equiv \frac{1}{n^C}\mathbb{V}\Big(\sum_{i=1}^{n}(1-W_{ij})Y_{ij}\Big) \\ S_j^{TC} &\equiv \frac{2n}{n^T n^C}\mathbb{V}\Big(\sum_{i=1}^{n} W_{ij}Y_{ij}, \quad \sum_{i=1}^{n}(1-W_{ij})Y_{ij}\Big) \end{aligned}$$

and $\mathbb{V}(\cdot)$ is variance operator with respect to $\boldsymbol{X}$. Neyman variance estimator for $\hat{\tau}^{\mathrm{dif}}$ under stratified randomization is (Imbens and Rubin, 2015, 202)

$$\hat{\mathbb{V}}^{\mathrm{Neyman}}(\hat{\tau}^{\mathrm{dif}}) \equiv \frac{1}{m^2}\sum_{j=1}^{m}\Big(\frac{\hat{S}_j^T}{n^T} + \frac{\hat{S}_j^C}{n^C}\Big),$$

where $\hat{S}_j^T$ and $\hat{S}_j^C$ are the sample variance of treated and control candidates within jurisdiction j, respectively:

$$\begin{aligned} \hat{S}_j^T &\equiv \frac{1}{n^T-1}\sum_{i=1}^{n} W_{ij}\Big(Y_{ij} - \frac{1}{n^T}\sum_{i'=1}^{n} W_{i'j}Y_{i'j}\Big)^2 \\ \hat{S}_j^C &\equiv \frac{1}{n^C-1}\sum_{i=1}^{n}(1-W_{ij})\Big(Y_{ij} - \frac{1}{n^C}\sum_{i'=1}^{n}(1-W_{i'j})Y_{i'j}\Big)^2. \end{aligned}$$

In general, a problem of $\hat{\mathbb{V}}^{\mathrm{Neyman}}(\hat{\tau}^{\mathrm{dif}})$ is its failure to estimate covariance *between* treated and control values in a jurisdiction, $S_j^{TC} < 0$, due to "the fundamental problem of causal inference." In addition, in our setup, $\hat{\mathbb{V}}^{\text{Neyman}}(\hat{\tau}^{\text{dif}})$ does not take into consideration covariance *among* treated (or control) values in a jurisdiction because of ignorance to interference between candidates: $\mathbb{V}(Y_{ij}, Y_{i'j} | W_{ij} = W_{i'j} = w), w \in \{0, 1\}$. Since they can be either positive or negative, we do not know the direction of bias of $\hat{\mathbb{V}}^{\text{Neyman}}(\hat{\tau}^{\text{dif}})$. Furthermore, in practice, when $n^T = 1$, we can calculate neither $\hat{S}_j^T$ nor $\hat{\mathbb{V}}^{\text{Neyman}}(\hat{\tau}^{\text{dif}})$ anyway.

Note that, when $\boldsymbol{Y}$ is not-logged vote share,

$$\sum_{i=1}^{n} Y_{ij} = 1.$$

Therefore,

$$\begin{aligned} \hat{\tau}^{\text{dif}} &= \frac{n}{n^C} \frac{1}{mn^T} \sum_{j=1}^{m} \sum_{i=1}^{n} W_{ij} Y_{ij} - \frac{1}{n^C} \\ \therefore \mathbb{V}(\hat{\tau}^{\text{dif}}) &= \left(\frac{n}{n^C} \right)^2 \frac{S^T}{mn^T}, \end{aligned}$$

where

$$S^T \equiv \frac{1}{m} \sum_{j=1}^{m} S_j^T.$$

In a nutshell, we only have to care about variation in treated values. In particular, when $n^T = 1$,

$$S_j^T = \underbrace{\frac{1}{n} \sum_{i=1}^{n} \left(y_{ij}^T - \frac{1}{n} \sum_{i'=1}^{n} y_{i'j}^T \right)^2}_{\text{variance of mean}} + \underbrace{\frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\{(Y_{ij} - y_{ij}^T)^2 | W_{ij} = 1\}}_{\text{mean of variance}}.$$

Thus, when $\boldsymbol{Y}$ is not-logged vote share and $n^T = 1$, we propose between variance estimator:

$$\hat{\mathbb{V}}^{\text{B}}(\hat{\tau}^{\text{dif}}) = \left(\frac{n}{n-1} \right)^2 \frac{\hat{S}^T}{m},$$

where adjusted sample variance of treated candidates between jurisdictions is estimated by

$$\hat{S}^T \equiv \underbrace{\frac{1}{N^{(1)}-1}\sum_{i=1}^{n}\left(\hat{y}_{i\cdot}^T - \frac{1}{N^{(1)}}\sum_{i'=1}^{n}\hat{y}_{i'\cdot}^T\right)^2}_{\text{variance of mean}} + \underbrace{\frac{1}{N^{(2)}}\sum_{i=1}^{n}\frac{1}{M_i-1}\sum_{j=1}^{m}W_{ij}(Y_{ij}-\hat{y}_{i\cdot}^T)^2}_{\text{mean of variance}},$$

the average marginal treated potential outcome of candidate i is estimated by

$$\hat{y}_{i\cdot}^T \equiv \frac{1}{M_i}\sum_{j=1}^{m}W_{ij}Y_{ij} \qquad \text{if } M_i \geq 1,$$

and $M_i \equiv \sum_{j=1}^{m} W_{ij}, N^{(d)} \equiv \sum_{i=1}^{n} I(M_i \geq d)$.

$\hat{\mathbb{V}}^{\mathrm{B}}(\hat{\tau}^{\mathrm{dif}})$ solves the problems of $\hat{\mathbb{V}}^{\mathrm{Neyman}}(\hat{\tau}^{\mathrm{dif}})$; it suffers from failure to estimate neither S_j^{TC} nor $\mathbb{V}(Y_{ij}, Y_{i'j}|W_{ij} = W_{i'j} = w), w \in \{0,1\}$. Above all, we can calculate $\hat{\mathbb{V}}^{\mathrm{B}}(\hat{\tau}^{\mathrm{dif}})$ even when $n^T = 1$.

When $\boldsymbol{Y}$ *is* logged vote share and $n^T = 1$, we regress $\boldsymbol{Y}$ on $\boldsymbol{W}$ alone and calculate the cluster robust standard error of $\boldsymbol{W}$'s coefficient where the cluster is jurisdiction j:

$$(\hat{\tau}^{\mathrm{ols}}, \hat{\alpha}^{\mathrm{ols}}) \equiv \arg\min_{\tau^{\mathrm{ols}},\alpha}\sum_{j=1}^{m}\sum_{i=1}^{n}(\epsilon_{ij}^{\mathrm{ols}})^2,$$

where

$$Y_{ij} = \alpha + \tau^{\mathrm{ols}}W_{ij} + \epsilon_{ij}^{\mathrm{ols}}.$$

Note that $\hat{\tau}^{\mathrm{ols}} = \hat{\tau}^{\mathrm{dif}}$ (Imbens and Rubin, 2015, 118). We use t-distribution, where the degree of freedom is equal to m.

Fixed Effects Estimator. Difference in vote shares among candidates will be substantial. For instance, major party candidates tend to obtain larger vote share than minor party candidates. Thus, we may control candidate fixed effects:

$$Y_{ij} = \tau^{\mathrm{FE}}W_{ij} + \sum_{i'=1}^{n}\beta_{i'}I(i=i') + \sum_{j'=1}^{m-1}\gamma_{j'}I(j=j') + \epsilon_{ij}^{\mathrm{FE}}. \tag{2}$$

As a result, we effectively control usual covariates which affect vote share (e.g. party affiliation, incumbency, seniority, and age). We also control jurisdiction fixed effects. Define the fixed effects estimator as

$$(\hat{\tau}^{\text{FE}}, \hat{\boldsymbol{\beta}}^{\text{FE}}, \hat{\boldsymbol{\gamma}}^{\text{FE}}) \equiv \arg\min_{\tau^{\text{FE}}, \boldsymbol{\beta}, \boldsymbol{\gamma}} \sum_{j=1}^{m} \sum_{i=1}^{n} (\epsilon_{ij}^{\text{FE}})^2.$$

It turns out that, though $\hat{\tau}^{\text{FE}}$ is not necessarily unbiased for τ^{AMTE} in a finite sample any more (Imbens and Rubin, 2015, Theorem 9.1), the following proposition holds under standard regularity conditions:[9]

Proposition 2 (Consistency). *$\hat{\tau}^{FE}$ is consistent for τ_{sp}^{AMTE}.*

When $\boldsymbol{Y}$ is not-logged vote share, it holds $\gamma_j = 0$; thus we do not include jurisdiction fixed effects in Equation 2.

Since Y_{ij}'s are dependent on each other within a jurisdiction, we report cluster robust standard error (where the cluster is jurisdiction j).[10] $\hat{\tau}^{\text{FE}}$ is more efficient than $\hat{\tau}^{\text{dif}}$ thanks to candidate fixed effects.

ANALYSIS

Data

We analyze the election to the House of Councillors in Japan held on July 10, 2016. The election consisted of the national PR tier and the local tier.[11] In the latter, each of 47

[9] This is because the numbers of treated or control candidates are constant across jurisdictions (n^T and n^C, respectively). For the same reason, we do not have to include interaction terms between W_{ij} and $I(j = j')$'s so as to obtain a consistent estimator (cf. Imbens and Rubin, 2015, Theorem 9.2). Note that these hold even if Equation 2 is misspecified.

[10] Instead, we may employ blocked bootstrap, where the block is jurisdiction j.

[11] Unlike the House of Representatives, there is no link betweeen the two tiers.

参議院（東京都選出）議員選挙候補者氏名等掲示

稲城市選挙管理委員会

党派名	氏名	党派名	氏名
日本のこころを大切にする党	鈴木まりこ	国民怒りの声	小林こうき
おおさか維新の会	田中康夫	無所属	原田きみあき
日本共産党	山添拓	無所属	いわさかゆきお
無所属	三宅洋平	公明党	竹谷とし子
支持政党なし	さめじま良司	無所属	浜田かずゆき
民進党	蓮舫	維新政党・新風	鈴木信行
無所属	よこぼり喜久	無所属	よこくめ勝仁
社会民主党	増山れな	チャレンジド日本	ふじしろ洋行
支持政党なし	おおつき文彦	支持政党なし	佐藤ひとし
支持政党なし	深江孝	新党改革	たかぎさや
地球平和党	ひめじけんじ	無所属	鈴木たつお
無所属	柳沢秀敏	無所属	佐藤かおり
犬丸勝子と共和党	犬丸勝子	世界経済共同体党	マタヨシ光雄
民進党	小川敏夫	自由民主党	朝日けんたろう
自由民主党	中川まさはる	無所属	川上晃司
幸福実現党	トクマ		

FIGURE 1: The Candidate List in Inagi City.

prefectures composes a SNTV district, where every voter casts a ballot for a candidate. We study Tokyo prefecure district, which had the most (six) seats and the most ($n = 31$) candidates among prefecure districts. In the district, no other election was held on the same day in any municipality.[12]

We collect candidate lists at the voting booth from all of 62 municipalities in Tokyo, sometimes by way of its freedom of information ordinance. An example of the list (Inagi City) is displayed in Figure 1. The list is displayed in row-horizontal format. The list have one row in seven municipalities, two rows in 51 municipalities (including Inagi City), and three rows in four municipalities. Since the position effects will vary across the numbers of rows, this study focuses on two row lists (thus, $m = 51$).[13]

A candidate's party and name are shown vertically within each cell, though candidates

[12] `http://www.senkyo.metro.tokyo.jp/election/schedule/senkyo2016/` (access December 13, 2017).

[13] The results for one- or three-row lists will be demonstrated in to-be appendix.

are not sorted by party as in the PR tier. Candidates' occupation, including incumbency status, race, ethnicity, and gender are not indicated. We refer to each position (or cell) by row r and column c or, simply, (r, c). In the cases of the top and bottom rows, $r = 1$ and $r = 2$, respectively. Two-row lists have 16 columns. Since Japanese count and read the vertical format from right columns to left columns, $c = 1$ and $c = 16$ in the cases of the most right and left columns in the top row, respectively. Note that cell $(2, 16)$ is empty.

The board of election in each municipality assigns candidates' positions by lots independently of each other. Thus, we suppose that candidate i in municipality j is placed in position (R_{ij}, C_{ij}) where

$$R_{ij} \equiv r(X_{ij}) \equiv 1 + I\{X_{ij} > 16\}$$
$$C_{ij} \equiv c(X_{ij}) \equiv X_{ij} - I(X_{ij} > 16) \times 16.$$

In order to examine not only the primacy and recency effects but also effects of all positions, this paper focuses on n treatments of position (r, c)'s:

$$W_{ij}^{(r,c)} \equiv I(R_{ij} = r, C_{ij} = c).$$

We denote the corresponding AMTE by $\tau^{(r,c)}$. For instance, the primacy effect is $\tau^{(1,1)}$. Note that, for all treatmetns, it always holds that $n^T = 1$.[14]

The dependent variable is either vote share (percentage, $0 \leq Y_{ij}^{\text{not log}} \leq 100$) or logarithm of vote share ($Y_{ij}^{\text{log}} \leq 0$).[15]

[14] In the elections of Brussel Regional Parliament, 75 candidates are displayed in one, four, or five colomns. Geys and Heyndels (2003) anaylze the effect of top or last position in each column.

[15] The data source is `http://www.senkyo.metro.tokyo.jp/election/sanngiin-all/sanngiin-sokuhou2016/` (access December 13, 2017). The details and descriptive statistics of the dependent variables are displayed in to-be appendix.

We will compare candidates by their group affiliation. One group is incumbents. Since the number of seats increased from five to six and one incumbent retired, four incumbents ran. We also divide candidates into three party groups.[16] First, (10) major party candidates are defined as those who belong to parties which are eligible to receive public funding. Second, (11) minor party candidates are defined as those who belong to parties which are not eligible to receive public funding. Last, (10) independent candidates are defined as those who do not belong to any parties.

Check of Randomization

Even if the candidate list order is randomized or rotated mechanically, researchers suspect that its precedure is sometimes politically manipulated (e.g. U.S. from 1917 to 1984 (Darcy and McAllister, 1990, 9), Russia in 2007 (Meredith and Salant, 2013, 194)). Thus, we check how plausible stratified randomization of $\boldsymbol{X}$ (or Equation 1) is.

Group Proportion. We adapt the method by Meredith and Salant (2013, 179–180). Let $b_i^{(g)}$ be a dummy variable to indicate whether candidate i Belongs to group g (e.g. incumbents). Among candidates who are assigned to position (r, c) in all jurisdictions given position vector $\boldsymbol{x}$, the proportion of candidates who belong to group g is calculated as

$$\pi^{(g)}(r, c|\boldsymbol{x}) \equiv \frac{1}{m} \sum_{j=1}^{m} b^{(g)}_{i(r,c|\boldsymbol{x}_{\cdot j})},$$

where $i(r, c|\boldsymbol{x}_{\cdot j})$ refers to the index of the candidate who is placed in position (r, c) given position vector $\boldsymbol{x}_{\cdot j}$.[17] Under the null hypothesis that candidates of group g are randomly as-

[16] For details of parties, see to-be appendix.

[17] Formally,

$$i(r, c|\boldsymbol{x}_{\cdot j}) \equiv \sum_{i'=1}^{n} I\{r(x_{i'j}) = r, c(x_{i'j}) = c\} i'.$$

signed to position (r, c), thanks to the Central Limit Theorem, the distribution of $\pi^{(g)}(r, c|\boldsymbol{X})$ should be approximated by the normal distribution whose mean and standard deviation are

$$\mu^{(g)} \equiv \frac{1}{n}\sum_{i=1}^{n} b_i^{(g)}, \qquad \sigma^{(g)} \equiv \sqrt{\frac{\mu^{(g)}(1-\mu^{(g)})}{m}},$$

respectively. Let $\boldsymbol{x}^{\text{obs}}$ denote the observed value of $\boldsymbol{X}$. When $\pi^{(g)}(r, c|\boldsymbol{x}^{\text{obs}})$ does not fall in the 95% interval of the above normal distribution, we reject the null hypothesis for position (r, c) at the five percentage significance level.

Figure 2 shows results where groups (g) are incumbents (top left), major party (top right), minor party (bottom left), and incumbent candidates (bottom right). In each panel, the horizonal axis indicates column c, the vertical axis represents group proportion $\pi^{(g)}(r, c|\boldsymbol{X})$. The thick and thin solid lines show $\pi^{(g)}(1, c|\boldsymbol{x}^{\text{obs}})$ and $\pi^{(g)}(2, c|\boldsymbol{x}^{\text{obs}})$, respectively. The top, middle, and bottom horizontal dotted lines correspond to $\mu^{(g)} + 1.96\sigma^{(g)}$, $\mu^{(g)}$, and $\mu^{(g)} - 1.96\sigma^{(g)}$, respectively. We find that four of 124 ($=n$ positions $\times$ 4 groups) or 3.2% of $\pi^{(g)}(r, c|\boldsymbol{x}^{\text{obs}})$'s are outside of the 95% interval (between the top and bottom dotted lines), which is not surprising under the null hypothesis.

We also test the joint null hypothesis that candidates of group g are randomly assigned across all positions. We define the group proportion test statistic as the average absolute value of the difference between $\pi^{(g)}(r, c|\boldsymbol{X})$ and $\mu^{(g)}$ in each position (r, c):

$$\theta^{(g)}(\boldsymbol{X}) \equiv \frac{1}{n}\sum_{x=1}^{n} |\pi^{(g)}\{r(x), c(x)|\boldsymbol{X}\} - \mu^{(g)}|.$$

The first column of Table 1 displays $\theta^{(g)}(\boldsymbol{x}^{\text{obs}})$ for the four groups (in the first through fourth rows). Analytically, under the joint null hypothesis, the distribution of $\theta^{(g)}(\boldsymbol{X})$ can be calculated exactly by employing Equation 1. Computationally, however, since we have to consider all elements of $\mathbb{X}^m$, namely, $|\mathbb{X}^m| = n!^m \approx 4.6 \times 10^{1729}$ sets of m lists of n candidates, it takes extraordinary time. Instead, we approximate the distribution of $\theta^{(g)}(\boldsymbol{X})$ by Monte Carlo simulation. In the s-th simulation, we randomly draw $\boldsymbol{x}^{(s)}$ from $\mathbb{X}^m$ and calculate

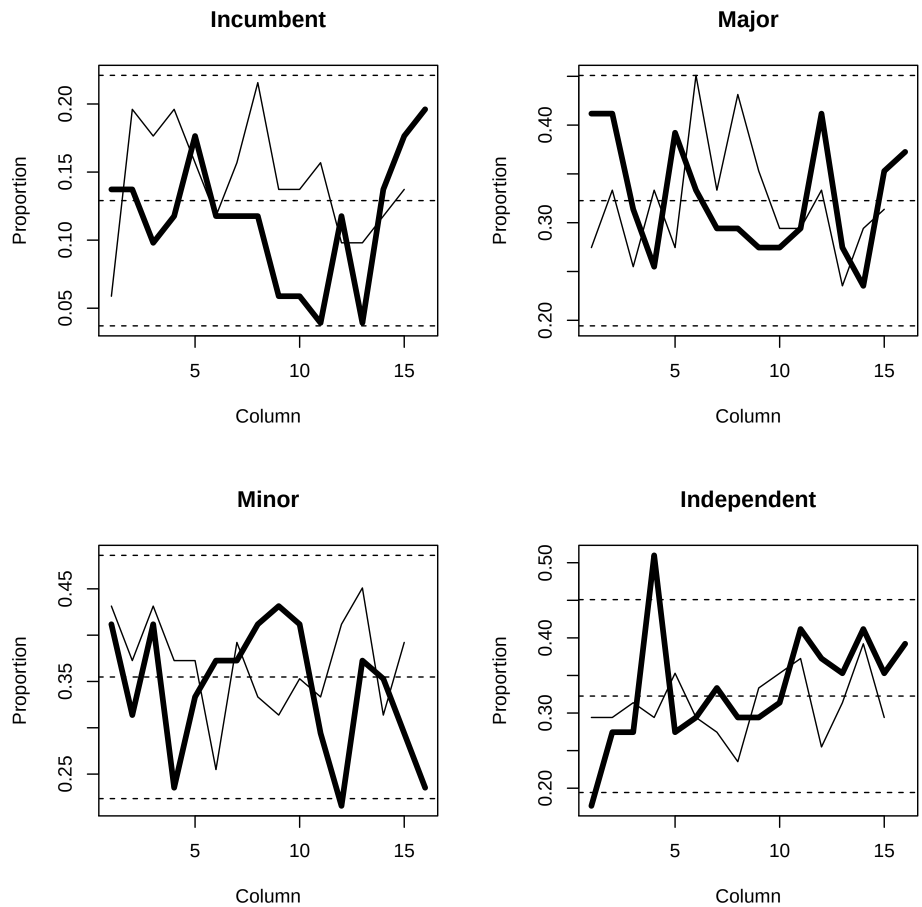

FIGURE 2: In each panel, the horizonal axis indicates column c, the vertical axis represents group proportion $\pi^{(g)}(r, c|\boldsymbol{X})$. The thick and thin solid lines show $\pi^{(g)}(1, c|\boldsymbol{x}^{\text{obs}})$ and $\pi^{(g)}(2, c|\boldsymbol{x}^{\text{obs}})$, respectively. The top, middle, and bottom horizontal dotted lines correspond to $\mu^{(g)} + 1.96\sigma^{(g)}, \mu^{(g)}$, and $\mu^{(g)} - 1.96\sigma^{(g)}$, respectively.

$\theta^{(g)}(\boldsymbol{x}^{(s)})$. We repeat this simulation $s^* = 10,000$ times. Then, we estimate one-tail p-value, $p^{(g)} \equiv \Pr\{\theta^{(g)}(\boldsymbol{X}) > \theta^{(g)}(\boldsymbol{x}^{\text{obs}})\}$, by

$$\hat{p}^{(g)} \equiv \frac{1}{s^*} \sum_{s=1}^{s^*} I\{\theta^{(g)}(\boldsymbol{x}^{(s)}) > \theta^{(g)}(\boldsymbol{x}^{\text{obs}})\}.$$

When $\hat{p}^{(g)} < 0.05$, we reject the joint null hypothesis at the five percentage significance level. The second column of Table 1 reports p-values, which imply that we do not have to reject the joint null hypothesis.

TABLE 1: Group Proportion Test Statistics

Group (g)	$\theta^{(g)}(\boldsymbol{x}^{\text{obs}})$	$\hat{p}^{(g)}$
Incumbent	0.037	0.592
Major	0.046	0.759
Minor	0.049	0.623
Independent	0.056	0.721

Note: The first clumn reports the value of the group proportion test statistics based on the observed candidate position. The second column shows one-tail p-values which are estimated by simulation.

Rank. We adapt the method by Ho and Imai (2008, 222–224) with slight modification. We calculate the "rank" statistics:

$$\theta^{\text{Rank}}(\boldsymbol{X}) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{n-1}\sum_{i'\neq i}\frac{1}{m}\Big|\sum_{j=1}^{m}(X_{ij}-X_{i'j})\Big|.$$

We can estimate the p-value in the same way as $\theta^{(g)}$. It turns out that $\theta^{\text{Rank}}(\boldsymbol{x}^{\text{obs}}) = 1.647, \hat{p}^{\text{Rank}} = 0.133$. Again, there is little evidence that stratified randomization fails.

Treatment Effects

Causal Inference. Figure 3 demonstrate the results of the fixed effect estimator ($\hat{\tau}^{\text{FE}}$) where the dependent variable is the logged vote share ($\boldsymbol{Y}^{\log}$). The left and right panels report the cases of the top and bottom rows ($\tau^{(1,c)}$ and $\tau^{(2,c)}$), respectively. In each panel, the horizonal axis indicates column (c), while the vertical axis represents exponentiated treatment effects ($\exp(\tau^{(r,c)})$). The thick solid line corresponds to point estimates ($\exp(\hat{\tau}^{\text{FE}(r,c)})$), while two thin solid lines represent the upper and lower bounds of the 95 % confidence interval based on the cluster robust standard error. The horizontal dotted line indicates a reference line of $\exp(\tau^{(r,c)}) = \exp(0) = 1$.

In the top row ($r = 1$, left panel), the primacy effect is significant and substantial: $\exp(\hat{\tau}^{\text{FE}(1,1)}) = 1.520$. That is, a candidate in position $(1,1)$ garners one and half as

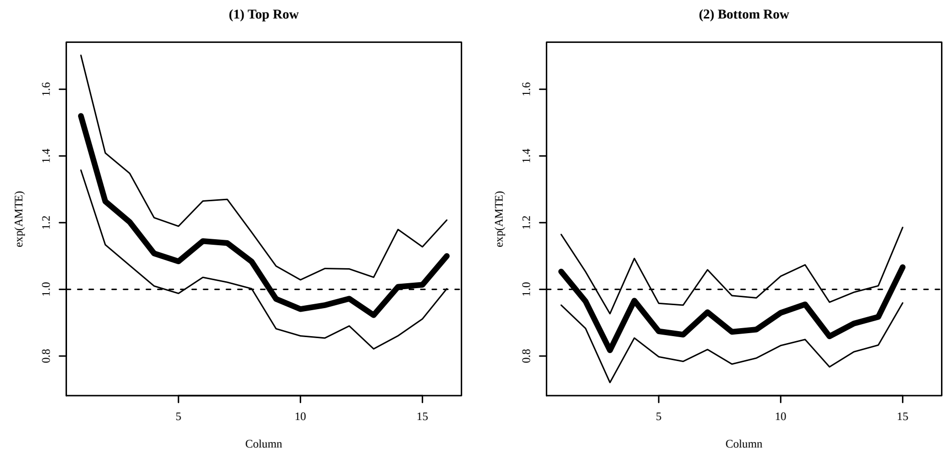

FIGURE 3: The results of the fixed effect estimator where $\boldsymbol{Y} = \boldsymbol{Y}^{\text{log}}$. In each panel, the horizonal axis indicates c, while the vertical axis represents $\exp(\tau^{(r,c)})$. The thick solid line corresponds to $\exp(\hat{\tau}^{\text{FE}(r,c)})$, while two thin solid lines represent the upper and lower bounds of the 95 % confidence interval based on $\hat{\mathbb{V}}^{\text{robust}}(\hat{\tau}^{\text{FE}(r,c)})$.

many votes as otherwise. In addition, for the right half columns ($c \leq 8$), point estimates ($\exp(\hat{\tau}^{\text{FE}(1,c)})$'s) are larger than one and, except for the fifth column ($\exp(\hat{\tau}^{\text{FE}(1,5)})$), they are significant. The recency effect ($\exp(\hat{\tau}^{\text{FE}(1,16)})$) is also significantly larger than one. When it comes to the bottom row ($r = 2$, right panel), treatment effects ($\exp(\hat{\tau}^{\text{FE}(2,c)})$'s) are insignificant or, if significant, smaller than one.

Figure 4 switches to the cases of non-logged vote share ($\boldsymbol{Y}^{\text{not log}}$) where the vertical axis represents non-exponentiated treatment effects ($\tau^{(r,c)}$) and the horizontal dotted line indicates a reference line of $\tau^{(r,c)} = 0$. The primacy effect is significantly positive ($\hat{\tau}^{\text{FE}(1,1)} = 0.560$). This is comparable to the vote share margin between the last winner and the runner-up, 0.605 %. The recency effect ($\hat{\tau}^{\text{FE}(1,16)} = 0.807$) is larger than the primacy effect, though not signficantly positive any more. In the top row ($r = 1$, left panel), the other treatment effects ($\hat{\tau}^{\text{FE}(1,c)}$'s) are insignificant except for the sixth column ($\hat{\tau}^{\text{FE}(1,6)}$). In the bottom row ($r = 2$, right panel), treatment effects ($\hat{\tau}^{\text{FE}(2,c)}$'s) are insignificant or, if significant, negative.

At first glance, results (e.g., the size of the primacy effect) in Figures 3 and 4 seem to be inconsistent with each other. Figure 5 addresses such a concern. In both panels, every dot

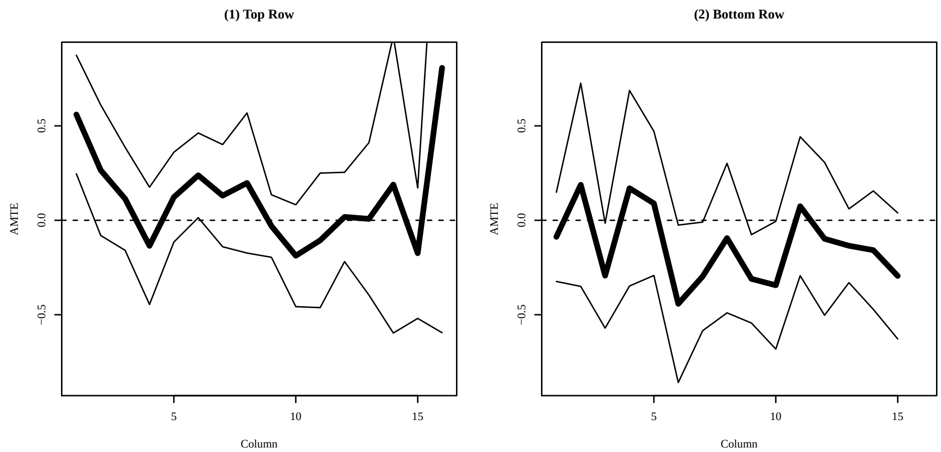

FIGURE 4: The results of the fixed effect estimator where $\boldsymbol{Y} = \boldsymbol{Y}^{\text{not log}}$. In each panel, the horizonal axis indicates c, while the vertical axis represents $\tau^{(r,c)}$. The thick solid line corresponds to $\hat{\tau}^{\text{FE}(r,c)}$, while two thin solid lines represent the upper and lower bounds of the 95 % confidence interval based on $\hat{\mathbb{V}}^{\text{robust}}(\hat{\tau}^{\text{FE}(r,c)})$.

represents each candidate, while the vertical and horizontal axes indicate each candidate's average vote shares in the treated and control groups ($\hat{y}_{i\cdot}^{T(1,1)}$ and $\hat{y}_{i\cdot}^{C(1,1)}$), respectively, where the treament variable is $\boldsymbol{W}^{(1,1)}$ and [18]

$$\hat{y}_{i\cdot}^{C(1,1)} \equiv \frac{1}{m - M_i^{(1,1)}} \sum_{j=1}^{m} (1 - W_{ij}^{(1,1)}) Y_{ij} \qquad \text{if } M_i^{(1,1)} \leq m - 1.$$

The distance between a dot for candidate i and the 45 degree line is equal to[19]

$$\hat{\tau}_{i\cdot}^{MTE(1,1)} \equiv \hat{y}_{i\cdot}^{T(1,1)} - \hat{y}_{i\cdot}^{C(1,1)}.$$

[18] Since $N^{(1)(1,1)} = 26$ and $0 \leq M_i^{(1,1)} \leq 4 < m - 1$, there are 26 dots in each panel.

[19] We can interpret $\hat{\tau}_{i\cdot}^{MTE(1,1)}$ as a natural estimate of the candidate i's average MTE:

$$\tau_{i\cdot}^{MTE(1,1)} \equiv \frac{1}{m} \sum_{j=1}^{m} \tau_{ij}^{MTE(1,1)}.$$

In the left panel, the dependent variable is the non-logged vote share ($\boldsymbol{Y}^{\text{not log}}$). It turns out that $\hat{\tau}_{i\cdot}^{MTE(1,1)}$ increases in $\hat{y}_{i\cdot}^{T(1,1)}$ or $\hat{y}_{i\cdot}^{C(1,1)}$. In the right panel, the dependent variable is the logged vote share ($\boldsymbol{Y}^{\text{log}}$). Now, $\hat{\tau}_{i\cdot}^{MTE(1,1)}$ *de*creases in $\hat{y}_{i\cdot}^{T(1,1)}$ or $\hat{y}_{i\cdot}^{C(1,1)}$. In essence, a certain percentage points of the primacy effect is relatively large for non-viable candidates than viable ones.

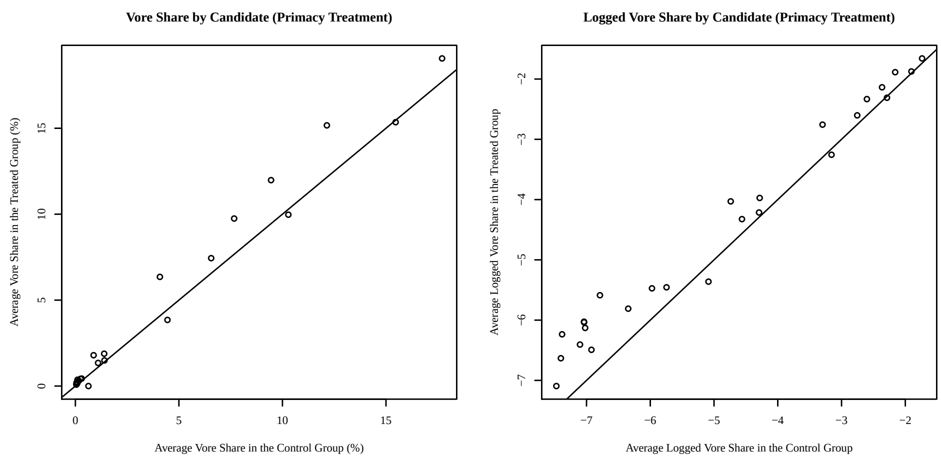

FIGURE 5: Average vote shares by candidate. Every dot represents each candidate, while the vertical and horizontal axes indicate $\hat{y}_{i\cdot}^{T(1,1)}$ and $\hat{y}_{i\cdot}^{C(1,1)}$, respectively. In the left and right panels, the dependent variable is $\boldsymbol{Y}^{\text{not log}}$ and $\boldsymbol{Y}^{\text{log}}$, respectively.

In Figures 6 and 7, we apply the difference-in-means estimator $\hat{\tau}^{\text{dif}}$ to $\boldsymbol{Y}^{\text{log}}$ and $\boldsymbol{Y}^{\text{not log}}$, utilize $\hat{\mathbb{V}}^{\text{robust}}(\hat{\tau}^{\text{dif}(r,c)})$ and $\hat{\mathbb{V}}^{\text{B}}(\hat{\tau}^{\text{dif}(r,c)})$, and show the results in the same way as in Figures 3 and 4, respectively. Generally, as expected, the standard errors of $\hat{\tau}^{\text{dif}}$ are larger than those of $\hat{\tau}^{\text{FE}}$. In Figure 6, only $\exp(\hat{\tau}^{\text{dif}(1,1)})$, $\exp(\hat{\tau}^{\text{dif}(1,2)})$, and $\exp(\hat{\tau}^{\text{dif}(2,13)})$ are significantly different from one (the first two are larger than one, though the last is smaller than one). In particular, the primacy effect duplicates vote share ($\exp(\hat{\tau}^{\text{dif}(1,1)}) = 2.024$). In Figure 7, no treatment effect is significantly different from zero, though the primacy and recency effects are estimated as around one and two percentage points without bias, respectively ($\hat{\tau}^{\text{dif}(1,1)} = 1.144, \hat{\tau}^{\text{dif}(1,16)} = 1.800$).

Model-based Approach. To sum, we find the primacy and recency effects. Moreover, it seems that, even if we ignore these two, treatment effects in the top row are larger than those in

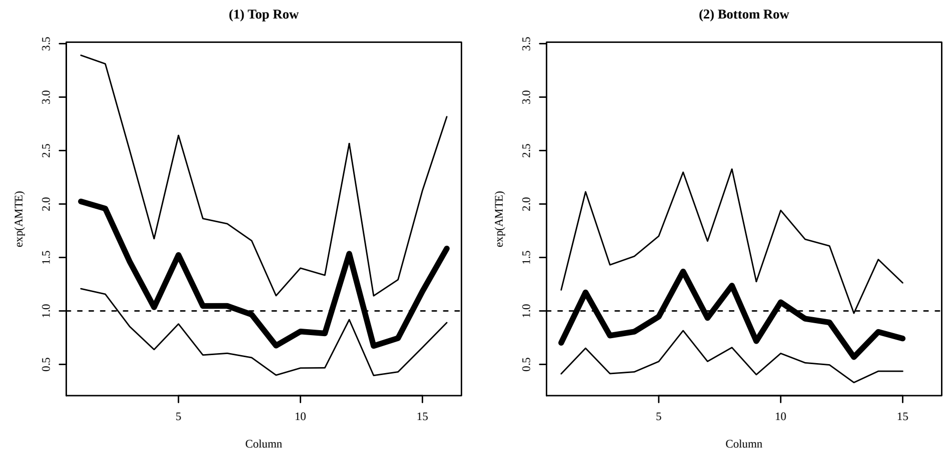

FIGURE 6: The results of the difference-in-means estimator where $\boldsymbol{Y} = \boldsymbol{Y}^{\log}$. In each panel, the horizonal axis indicates c, while the vertical axis represents $\exp(\tau^{(r,c)})$. The thick solid line corresponds to $\exp(\hat{\tau}^{\mathrm{dif}(r,c)})$, while two thin solid lines represent the upper and lower bounds of the 95 % confidence interval based on $\hat{\mathbb{V}}^{\mathrm{robust}}(\hat{\tau}^{\mathrm{dif}(r,c)})$.

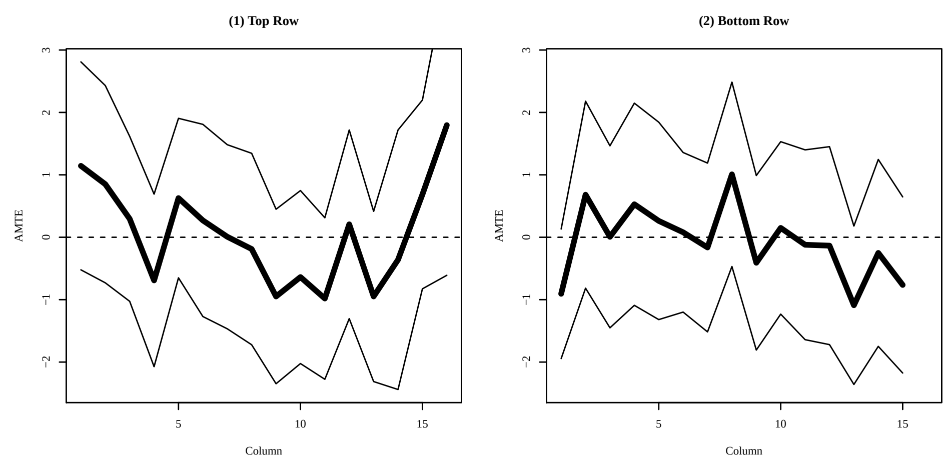

FIGURE 7: The results of the difference-in-means estimator where $\boldsymbol{Y} = \boldsymbol{Y}^{\text{not log}}$. In each panel, the horizonal axis indicates c, while the vertical axis represents $\tau^{(r,c)}$. The thick solid line corresponds to $\hat{\tau}^{\mathrm{dif}(r,c)}$, while two thin solid lines represent the upper and lower bounds of the 95 % confidence interval based on $\hat{\mathbb{V}}^{\mathrm{B}}(\hat{\tau}^{\mathrm{dif}(r,c)})$.

the bottom row ($\tau^{(1,c)} > \tau^{(2,c)}$);[20] the treatment effect for a right column is larger than that

[20] In the recall election of California Governor in 2003, 135 candidates are carried over several pages. Ho and Imai (2006) examine the effect of the first page.

for a left column in the top row, but not in the bottom row ($\tau^{(1,c)} > \tau^{(1,c')}$ if $c < c'$). In order to examine these tendencies, we analyze the following regression by OLS:

$$\begin{aligned} y_{ij}^{\text{log}} = {} & \tau^{\text{primacy}} I(R_{ij} = 1, C_{ij} = 1) + \tau^{\text{recency}} I(R_{ij} = 1, C_{ij} = 16) + \tau^{\text{top row}} I(R_{ij} = 1) \\ & + \tau^{\text{top row} \times \text{column}} I(R_{ij} = 1) \times C_{ij} + \tau^{\text{bottom row} \times \text{column}} I(R_{ij} = 2) \times C_{ij} \\ & + \sum_{i'=1}^{n} \beta_{i'}^{\text{model}} I(i = i') + \sum_{j'=1}^{m} \gamma_{j'}^{\text{model}} I(j = j') + \epsilon_{ij}^{\text{model}}. \end{aligned} \tag{3}$$

No coefficients have causal interpretation any more. We use cluster-robust standard error, where the cluster is jurisdiction (j). According to the first column of Table 2, the result follows our conjecture; not only $\hat{\tau}^{\text{primacy}}$ and $\hat{\tau}^{\text{recency}}$ but also $\hat{\tau}^{\text{top row}}$ and $\hat{\tau}^{\text{top row} \times \text{column}}$ are significant, though $\hat{\tau}^{\text{bottom row} \times \text{column}}$ is not.

	All		Incumbent		Non-incumbent	
Primacy	0.213	**	0.068		0.240	**
	(0.065)		(0.059)		(0.078)	
Recency	0.183	**	0.255	*	0.190	**
	(0.056)		(0.109)		(0.059)	
Top Row	0.294	**	0.073		0.316	**
	(0.061)		(0.057)		(0.066)	
Top Row × Column	−0.019	**	−0.014		−0.018	**
	(0.005)		(0.008)		(0.005)	
Bottom Row × Column	0.000		−0.008		0.001	
	(0.003)		(0.005)		(0.003)	
N	1581		204		1377	

TABLE 2: Results of Model-based Approach

Note: Entries are OLS estimates using Equation 3. Cluster-robust standard error are in parenthesis.

Exploiting benefits of a model-based approach, we divide the dataset by group affiliation and apply Equation 3 to each sub-dataset so as to study the above empirical regularity in each group. The second and third columns of Table 2 show the results for incumbents and non-incumbents, respectively. Clearly, we find the same tendencies for non-incumbents (third column) as for all candidates (first column), but not for incumbents (second column).

In Table 3, candidates are classified into major party, minor party, and independents. All the coefficients' sizes for independents (third column) are larger than those for all candidates, while even $\hat{\tau}^{\text{bottom row} \times \text{column}}$ is significant and positive contraty to $\hat{\tau}^{\text{top row} \times \text{column}}$ which is negative. All the coefficients' sizes (except for $\hat{\tau}^{\text{bottom row} \times \text{column}}$) for minor party candidates (second column) are larger than those for major party ones (first column), though $\hat{\tau}^{\text{primacy}}$ and $\hat{\tau}^{\text{recency}}$ are insignificant for minor party (for major party, $\hat{\tau}^{\text{top row}}$ and $\hat{\tau}^{\text{top row} \times \text{column}}$ are not significant). Therefore, it seems that candidate positions matter for independents most.

	Major		Minor		Independent	
Primacy	0.157	*	0.218		0.324	**
	(0.079)		(0.169)		(0.103)	
Recency	0.162	**	0.244		0.168	*
	(0.060		(0.145)		(0.073)	
Top Row	0.041		0.398	**	0.451	**
	(0.041)		(0.094)		(0.105)	
Top Row × Column	−0.004		−0.025	**	−0.025	*
	(0.005)		(0.006)		(0.010	
Bottom Row × Column	−0.003		−0.004		0.008	**
	(0.003)		(0.006)		(0.002)	
N	510		561		510	

TABLE 3: Results of Model-based Approach

Note: Entries are OLS estimates using Equation 3. Cluster-robust standard error are in parenthesis.

These results are consistent with some studies (Meredith and Salant, 2013; Krosnick, Miller, and Tichy, 2004; Pasek et al., 2014), but not others (Alvarez, Sinclair, and Hasen, 2006; Ho and Imai, 2008; King and Leigh, 2009; Marcinkiewicz and Jankowski, 2014).

CONCLUSION

We do not intend to arbitrate the debate on candidate position effects on vote share, while we hope to add a new evidence for further discussion by improving internal and external validitity.

On the side of internal validitity, taking into consideration violation of SUTVA, we propose a new estimand, τ^{AMTE}, and a new variance estimator, $\hat{\mathbb{V}}^{\text{B}}(\hat{\tau}^{\text{dif}})$. We show unbiasedness of $\hat{\tau}^{\text{dif}}$ for τ^{AMTE} (Proposition 1) and consistency of $\hat{\tau}^{\text{FE}}$ for $\tau^{\text{AMTE}}_{\text{sp}}$ (Proposition 2) under stratified randomzaition (Equation 1) and Assumptions 1 and 2 which are more relaxed than SUTVA.

On the side of external validitity, we find the primacy effect ($\tau^{(1,1)}$) and the recency effect ($\tau^{(1,16)}$) in Japan where candidates' names are displayed not on the ballot but on the vertical format list at the voting booth; primacy lies at the top right corner instead of top left; SNTV is employed. It also turns out that candidates in the top row, in particular in right columns, tend to obtain more votes. Furthermore, it seems that non-incumbent or independent candidates are likely to be affected by position effects. We believe accumulation of findings in various settings improve the literature.

A practical policy suggestion implied by this study is to randomize the list order not across municipalities as currently implemented but across precincts or even voting booths. For instance, Setagaya ward has over 749 thousands eligible voters, while Aogashima village (in Aogashima island) has just 140.[21] Benefit of the primacy effect should be more than five thousands times as large in the former as in the latter. Note that printing various lists is much less costly than printing various ballots. Our proposal for candidates list should not add to confusion in counting write-in ballots, either.

Finally, we hope that our study moves forward the research on position effects.

[21] `http://www.senkyo.metro.tokyo.jp/election/sanngiin-all/sanngiin-sokuhou2016/` (last access on December 13, 2017).

References

Alvarez, R. Michael, Betsy Sinclair, and Richard L. Hasen. 2006. "How Much Is Enough? The Ballot Order Effect and the Use of Social Science Research in Election Law Disputes." *Election Law Journal* 5(1): 40–56.

Bagley, C R. 1965. "Does Candidates' Position On The Ballot Paper Influence Voters' Choice? – A Study Of The 1959 and 1964 British General Elections." *Parliamentary Affairs* 19(2): 162–174.

Blom-Hansen, Jens, Jørgen Elklit, Søren Serritzlew, and Louise Riis Villadsen. 2016. "Ballot position and election results: Evidence from a natural experiment." *Electoral Studies* 44(dec): 172–183.

Brockington, David. 2003. "A low information theory of ballot position effect.".

Curtice, John, and Michael Marsh. 2008. "How Did They Vote? Voters' Use of the STV Ballot Paper in the 2007 Scottish Local Elections." *Representation* 44(4): 285–300.

Darcy, R. 1986. "Position Effects with Party Column Ballots." *The Western Political Quarterly* 39(4): 648–662.

Darcy, R., and Ian McAllister. 1990. "Ballot position effects." *Electoral Studies* 9(mar): 5–17.

Faas, Thorsten, and Harald Schoen. 2006. "The importance of being first: Effects of candidates' list positions in the 2003 Bavarian state election." *Electoral Studies* 25(mar): 91–102.

Geys, Benny, and Bruno Heyndels. 2003. "Ballot layout effects in the 1995 elections of the Brussels' government.".

Ho, Daniel E, and Kosuke Imai. 2006. "Randomization Inference With Natural Experiments." *Journal of the American Statistical Association* 101(sep): 888–900.

Ho, Daniel E., and Kosuke Imai. 2008. "Estimating causal effects of ballot order from a randomized natural experiment: The California alphabet lottery, 1978-2002.".

Imbens, Guido W., and Donald B. Rubin. 2015. *Causal Inference for Statistics, Social, and Biomedical Sciences.* Cambridge University Press.

Jun, Byung-hill, and Heechul Min. 2017. "What creates heterogeneity in ballot order effects? Evidence from Korea's local elections of education superintendent." *Electoral Studies* 46: 1–14.

King, Amy, and Andrew Leigh. 2009. "Are ballot order effects heterogeneous?" *Social Science Quarterly* 90(1): 71–87.

Krosnick, Jon A., Joanne Miller, and M. P. Tichy. 2004. "An unrecognized need for ballot reform: Effects of candidate name order." In *Rethinking the vote: The politics and prospects of American election reform.*, ed. A. N. Crigler, M. R. Just, and E. J. McCaffery. New York, NY:: Oxford University Press.

Lijphart, Arend, and Rafael L?pez Pintor. 1988. "Alphabetic bias in partisan elections: Patterns of voting for the Spanish Senate, 1982 and 1986." *Electoral Studies* 7(dec): 225–231.

Lutz, Georg. 2010. "First come, first served: The effect of ballot position on electoral success in open ballot PR elections." *Representation* 46(2): 167–181.

Marcinkiewicz, Kamil. 2014. "Electoral contexts that assist voter coordination: Ballot position effects in Poland." *Electoral Studies* 33: 322–334.

Marcinkiewicz, Kamil, and Mary Stegmaier. 2015. "Ballot Position Effects Under Compulsory and Optional Preferential-List PR Electoral Systems." *Political Behavior* 37(2): 465–486.

Marcinkiewicz, Kamil, and Michael Jankowski. 2014. "When There's No Easy Way Out: Electoral Law Reform and Ballot Position Effects in the 2011 Hamburg State Elections." *German Politics* 23(1-2): 103–117.

Meredith, Marc, and Yuval Salant. 2013. "On the Causes and Consequences of Ballot Order Effects." *Political Behavior* 35(1): 175–197.

Miller, Joanne M, and John Krosnick. 1998. "The Impact of Name Order on Election Outcomes." *Public Opinion Quarterly* 62(3): 291–330.

Orr, Graeme. 2002. "Ballot Order: Donkey Voting in Australia." *Election Law Journal: Rules, Politics, and Policy* 1(dec): 573–578.

Ortega Villodres, Carmen. 2003. "Intraparty Competition Under Preferential List Systems: The Case of Finland." *Representation* 40(1): 55–66.

Pasek, Josh, Daniel Schneider, Jon A. Krosnick, Alexander Tahk, Eyal Ophir, and Claire Milligan. 2014. "Prevalence and moderators of the candidate name-order effect evidence from statewide general elections in California." *Public Opinion Quarterly* 78(2).

Robson, Christopher, and Brendan Walsh. 1974. "The Importance of Positional Voting Bias in the Irish General Election of 1973." *Political Studies* 22(jun): 191–203.

Tourangeau, Roger, Mick P. Couper, and Frederick G. Conrad. 2013. "Up means good: The effect of screen position on evaluative ratings in web surveys.".

van Erkel, Patrick F.A., and Peter Thijssen. 2016. "The first one wins: Distilling the primacy effect." *Electoral Studies* 44(dec): 245–254.