Matching and the Multi-valued Treatment Problem: An Application to the Estimation of the Ballot Order Effect

Betsy Sinclair†

September 7, 2006

† Ph.D. candidate, Division of Humanities and Social Sciences, California Institute of Technology (betsy@hss.caltech.edu). I thank R. Michael Alvarez, Jonathan Katz, Geralyn Miller, Melanie Goodrich, Rick Hasen, Delia Bailey and participants of the July 2005 Political Methodology poster session for helpful discussions about this paper.

1 Introduction

In recent years there has been much written about new methodologies for causal inference in political sciences, in particular the development of matching methods (see, for example, Imai 2005; Epstein et al. 2005). These methods are being widely applied in other social sciences, and of course have been discussed for some time in the statistics literature (for a review of recent research on matching methods outside political science see Diamond and Sekhon 2005). One of the important questions that has been the focus of recent methodological development has been how to best perform matching, with a variety of methods being now promoted in the literature. Some of the matching approaches utilize the popular propensity score matching technique, first elaborated by Rosenbaum and Rubin (1983); others advocate matching using a distance algorithm (see for example Cochran and Rubin 1973); and recent research by Diamond and Sekhon (2005) has proposed matching based on "evolutionary algorithms."

However, one of the limitations of the matching algorithms advanced in the political methodology literature is that they commonly focus on a relatively simple matching case — a situation where there is a single-valued treatment. That is, the applications in the literature to date examine situations where an individual participates in an experimental treatment or not (Imai 2005; Diamond and Sekhon 2005). The argument is that while these are of course important applications when the treatment is single-valued, that there are other situations that arise in political science research where the treatment will be multi-valued, taking a series of treatment values. The contribution to this literature is to provide a particular methodology – multi-valued treatment matching – to help in producing a balanced dataset.

There are obviously many situations in political science research when a treatment might take on multiple values; the substantive application to focus the attention on in this paper is the position that candidates are listed on a ballot, using data from two recent election cycles (1998 and 2002) in California. This substantive application arises from a long-standing, but recently revived, debate in the social science literature as to whether or not the candidate's position at the top of a ballot order enhances the number of votes that this candidate receives in an election. The recent research has centered on disputes over the proper methodology to measure the ballot order effect, whether the ballot order effect exists at all in general (as opposed to primary) elections, and what types of candidates (minor party, incumbent) would most benefit from a particular ballot placement.[1]

In the research reported below, a procedure is developed that allows for matching using multi-valued treatments. This procedure is used on data from these two elections in California, in order to test the hypothesis that candidates listed first on the ballot are necessarily advantaged over those listed in any other position in the 1998 and 2002 California *general* elections. I use a methodology for studying the ballot order effect that was first advanced in Alvarez, Sinclair and Hasen (2006); I then balance the data using the multi-valued treatment method and show that the substantive implications of the previous research are strengthened by the use of appropriately balanced data: I find little support for any claims of systematic bias towards candidates listed first on the ballot. Using the balanced data, the results indicate that there is no clear pattern to when candidates will be advantaged from any particular ballot position. Furthermore, when there is evidence that being first on the ballot does help candidates the effect is arbitrary, unpredictable and of small substantive magnitude. This paper develops a new statistical technique particularly suited to analyze the ballot order effect which eliminates many of the concerns associated with the existing literature.

[1]Whether the effect is observed: R. Michael Alvarez, Betsy Sinclair and Rick Hasen, "Ballot Order I", *Election Law Journal*, 2006. For primary elections: Jonathan GS Koppell and Jennifer A. Steen, The Effects of Ballot Position on Election Outcomes, 66 JOURNAL OF POLITICS 267 (2004). For minor party candidates: Daniel E. Ho and Kosuke Imai, The Impact of Partisan Electoral Regulation: Ballot Effects from the California Alphabet Lottery, 1978-2002, manuscript, http://www.princeton.edu/ kimai/research/alphabet.html.

2 Matching with Multi-valued Treatment

There has been great concern in the political methodology literature about empirical analysis on data that does not derive from a randomized experiment. Without the ability to compare the effects of a treatment on similar subjects, it is difficult to know whether or not the observed differences are the result of the treatment or instead the consequence of differences in the subjects. This problem is most visible with respect to model specification. Below is a small example to demonstrate the necessity of balancing a dataset. Suppose that there are only two groups – a treatment group and a control group. Suppose the hypothesis is that there is a linear and positive relationship between X and Y. Now suppose that there exist some values of X in the control group that are outside of the range within which the treatment group is observed. The graph below demonstrates this case – the circles are the treated units and the diamonds are the control units. Note that there are a series of diamonds (control units) outside the range of any of the circles (treated units). If I fit a line through the treated units and a line through the control units including these points that are outside the range of any of the treated units, I might wrongly conclude that there is a very strong "treatment" effect for small x values – that is, the line that best fits the treated units is horizontal while the line that best fits the control units is upward sloping. However, I have no ability to know for certain what the shape of the line would look for small X values that are treated, since there are none. Furthermore, I should not include these small X values in the fitted line for the control group since I have no treated values to compare them to. Thus, I plot new fitted lines for the "balanced" data – instances where there are both control and treat units within a certain distance of each other in their x values. Note that when I plot these new lines I do not observe a difference between treated and control.

Figure 1 Goes Here

In the study, since the ordering assignment is not random, but instead rotates across the state, there is a concern that the specific rotation may be related to an observed ballot order effect. Specifically, because the interior of California tends to be more conservative than the coast, and because the ordering assignment scrolls through the assembly districts beginning in northern California and continuing east-to-west as it moves south, it would be possible for a liberal candidate to be first in many conservative assembly districts, for example.

In the analysis, I consider each candidate name ordering equivalent to one treatment in each contest. Since the candidate name order rotates through the candidate list, note then that the number of candidate orderings is equivalent to the number of candidates. The quantity of interest here is the effect of each treatment (ordering) on each candidate's vote shares. Furthermore, I would like to able to compare across treatments (orderings), so that I could then examine the effects of being first against the effects of being last, for example. However, most of the existing literature that examines treatment effects deals only with a binary treatment case – where the subjects are either treated or not.[2] In this case the binary treatment processes are not applicable since in most of the elections in this study consider there are more than two candidates. Therefore, I modify two existing processes to produce a new algorithm to incorporate multiple treatments.[3] The intuition is to produce something similar to the processes which perform matching based upon a single propensity score, and is discussed for subclassification with the binary case in

[2]For a good review of the literature, please see Alexis Diamond and Jasjeet Sekhon, "Genetic Matching for Estimating Causal Effects". Unpublished manuscript. http://sekhon.polisci.berkeley.edu/matching/.

[3]Two studies for multivalued treatments include one using balancing in Imbens (Imbens, Guido. "The Role of the Propensity Score in Estimating Dose-Response Functions", *Biometrika*, Vol. 87. No. 3 (Sep 2000), 706-710) and using subclassification instead of matching for multivalued treatments in Imai and Van Dyk (Imai, Kosuke and Van Dyk, David. "Causal Inference with General Treatment Regimes: Generalizing the Propensity Score", *Journal of the American Statistical Association*, September 2004, Vol. 99, No. 467, Theory and Methods.)

Rosembaum and Rubin (1984) [4] and introduced with matching in Rosenbaum and Rubin (1983). [5] The methodology is an extension of the generalized propensity score literature where instead in this case the matching will be based upon a multivalued propensity score. The goal is to produce a new dataset for each contest, where observations are selected into the new dataset based upon the similarities of their propensity vectors (a multi-valued propensity score). This process should reduce any bias in the estimates that would have occurred from the lack of a randomized experiment. Note that if California had randomly assigned ballot order by precinct or by census tract as opposed to rotating an order across 80 districts this would have allowed for the observation of data from a randomized experiment. Since, however, this did not occur I compensate as best I can for the lack of randomness by producing a new dataset which attempts to mimic data from a randomized experiment.

I begin by considering the number of treatments (orderings) equal to the number of candidates in that contest. These treatments reflect the possible ballot locations for each candidate. Note that because the number of candidates is relatively small (the largest number of candidates is 7), each candidate receives each treatment.

I next produce a propensity vector. This vector is a vector of conditional probability of assignment for each particular treatment given the vector of covariates described in the previous section. In the binary case we were interested in a propensity score, but given that we have multiple treatments we can still reduce the dimensionality of the covariates down into several values, the sole difference is that this reduction is now a vector instead of a single value. Each entry in the vector is the probability of getting that treatment giving the observed covariates (i.e. the first entry in the vector is the probability of getting

[4] Rosenbaum, Paul and Rubin, Donald. "Reducing Bias in Observational Studies Using Subclassification on the Propensity Score", *Journal of the American Statistical Association*, Vol. 79, No. 387 (Sep. 1984)

[5] Rosenbaum, Paul and Donald Rubin. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika*, Vol 70, 1983, 41-55.

the first treatment). We know from Rosenbaum and Rubin that if we can match observations based upon their propensity vector this is equivalent to matching on all observed covariates.[6] To produce the propensity vector I use treatment as the independent variable and, using all the covariates included above as independent variables, fit the data using multinomial logit. I then calculate the fitted values for all observations. This produces a vector which has the the same number of values as there are treatments. This vector represents a multi-valued propensity score.

I then sort the observations into different treatment groups. Note that each observation will have a propensity vector associated with it. I randomly select a single observation from the data in the first treatment. I find the observation in the second treatment that minimizes the distance between this observation and an observation in the second treatment in terms of the propensity vectors. I call this distance *pair distance*. Formally, it is defined as:

$d_{PAIR}(x_1, x_2) = \sum_{i=1}^{c}(x_{1i} - x_{2i})^2$ where c = the number of values in the propensity vector.

This notion of the distance between two pairs is intuitive; the distance between the two propensity vectors is the sum of squared distances between each entry in the propensity vector. Note here that these calculations below were also examined using absolute distance as well and that this notion of paired distance produced similar results.

I next look for an observation in the third treatment that is also close (in terms of the propensity vector) to the first two observations in the group. In order to find the "closest" observation, I must first define a notion of group distance. Formally it is defined as:

$d_{GROUP}(x_1, x_2, \ldots, x_g) = max[(d_{PAIR}(x_i, x_j))\forall(x_i, x_j) \in (x_1, x_2, \ldots, x_g)$ where $i \neq j]$

[6]Rosenbaum, Paul and Donald Rubin. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika*, Vol 70, 1983, 41-55.

I search for the observation in treatment three that minimizes group distance. I repeat this process for each additional treatment group until I have a complete set of observations which I consider ”matched”. I then discard that set of observations from the group and repeat the process, drawing one observation randomly from the group which received treatment one and finding the next set of observations in the remaining treatments which minimize group distance.

I continue selecting observations which are ”matched” until the paired-t-tests on the means of the covariates begin to produce values which indicate that there are statistically significant differences. In the particular study, I fixed a specific stopping value which appeared to eliminate most of the statistically significant differences in means, so that most of the t-statistics have a value less than 2. I note here that it is possible that these tests may be misleading – for example, it is conceivable that randomly discarding data might result in acceptable t-statistics while yet not improving balance.[7] Therefore, another method of evaluation is necessary to compare whether or not the new sample has indeed improved balance and will help in eliminating bias.

I next compare difference in means to notice the improvement in the new datasets. Each covariate for each contest is examined. Here I plot the mean and plus or minus 1/2 the standard deviation for a single covariate for each treatment in both the original sample and the balanced sample for the percent decline-to-state registration in the 2002 Controller contest. This is a particularly simple summary of the data – note, however, that the mean and 1/2 standard deviation interval of the covariates in the original sample often do not even overlap. This indicates that there are significant differences across particular orderings which may result in biased coefficients. Examining the balanced covariates, the means are often grouped closely together and generally the 1/2 standard deviation inter-

[7] Imai, Kosuke, Gary King and Elizabeth Stuart. 2005. ”The Balance Test Fallacy.” http://gking.harvard.edu/files/balfal.pdf.

vals overlap across treatments.[8] However, the means are not identical across treatments, they are simply closer than they were before the matching algorithm was applied to the data. This implies that examining the treatment effects will not be sufficient – it is still necessary to examine the parametric statistics in order to examine the ballot order effect. Also note that because I select a smaller sample in the balanced sample, the variances are larger in the balanced sample.[9]

Figure 2 Goes Here

Figure 3 Goes Here

3 California's Ballot Order Experiment

California provides a particularly good opportunity to study the ballot order effect. While the candidate rotations do not generate a random sample, the system of rotations (which began with a random alphabetic ordering) produce a unique opportunity to analyze the effects of different ballot placements. In California, the same system for ordering candidate names on the ballot has been in place since 1975. Prior to that, state courts held that the previous system, based on alphabetical order or listing incumbents first was unconstitutional due to an alleged 5% ballot order effect among undecided voters.[10] Legislation passed in 1975 details procedures to be followed for the placement of candidate names on any election ballot.[11]

[8]Ho, Daniel E., Kosuke Imai, Gary King and Elizabeth Stuart. "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference." Unpublished manuscript, 2005. http://gking.harvard.edu/projects/cause.shtml. According to the authors "One particularly simple low dimensional summary compares the mean of each variable in X (the set of covariates) for the treated group with the mean of each variable in the control group. The smaller these differences are the better. On rule of thumb that has been offered is if one or more of these differ by more than half a standard deviation of the respective X variable, then better balance is needed, but finding "small" imbalance in the original units is the real goal."

[9]Graphs of all covariate means and 1/2 standard deviation intervals are available upon request from Sinclair.

[10]CITE

[11]Codified in California's Election Code Sections 13111 through 13114.

The general principles in place in the California election code since 1975 are randomization and rotation. The Secretary of State, before an election, conducts a random drawing of letters of the alphabet.[12] This randomized list is used for all candidate races in the upcoming election. In statewide races, the randomized alphabet list drawn by the Secretary of State determines the order of candidate names on all ballots in Assembly District 1. The order is then rotated, so that in Assembly District 2 the candidate who appeared first in Assembly District 1 is moved to the bottom of the ballot list and the second candidate is moved up to be the first candidate on the ballot in Assembly District 2. This rotation process continues throughout all of the 80 Assembly Districts in California for statewide races. The hope is that this process would obliterate any ballot order effect.

One possibility is that this system of randomization and rotation may help alleviate the ballot order effect but not eliminate it. It seems possible that with a large ballot order effect this system may not be sufficient – therefore, even given the California ballot order experiment it should be possible to observe an effect. One question a student of statistics might ask about the California system is whether there is a difference between taking 80 random draws from all the possible candidate orderings ($c!$ where c is the number of candidates) or rotating the order across 80 districts after randomly picking a single ordering. The difference between these two methods may impact the ability to draw statistical inferences about the size of the ballot order effect depending upon whether we can conclude that the rotation system produces a random sample. Note that with only 80 districts it may not be possible to observe a ballot order effect depending upon the acceptable levels of type I and type II errors.[13]

We know at the outset that the rotation system does not produce a random sample. For the sample to be random it would be necessary to draw out an entire order 80 times,

[12]Cal. Elec. Code 13112.
[13]See Appendix for details.

where each possible order is $\frac{1}{c!}$ likely. In the California system, drawing the first ordering is just like drawing one order and is random. The problem arises from the next 79 districts. There are three conditions under which it does not matter that the California system is not random.

First, assume that the ballot effect occurs only because of the placement of being first or last but has nothing to do with which neighbors each candidate name has on the ballot. Suppose that there is only an effect of being first. Since the entire order does not matter, but only a single placement within that order, it is not necessary to redraw the entire order over and over again, but instead redraw only the candidate who will get to be first. In this case, each candidate is $\frac{1}{c}$ likely to be first if the order was redrawn n times. Rotating the order n times after a single draw in which each candidate is $\frac{1}{c}$ likely is then identical in the limit as $n \rightarrow \infty$.

However, we know that it *does* sometimes matter who the candidate's neighbor is on the ballot.[14] For example, in the California recall election, the candidate George B. Schwartzman who was placed next to Arnold Schwarzenegger won a surprisingly high number of votes considering his previous ambiguity (.2% of the statewide total). Although in this instance there could be an additional effect that the names are similar, it seems likely that unknown candidates who are placed next to more famous candidates will win more votes, either because they are more easily seen on the ballot or that they are chosen because of voting error. Furthermore, the order is rotated only 80 times. Since the number of rotations is fixed, this can produce a set of differences between randomization and rotation.

The second major concern about the lack of randomization is if c does not divide perfectly into 80. Suppose that if you divide 80 by c you produce a remainder of r. This means that some candidates will get to be first more times than others. If the ballot or-

[14] Alvarez et al., "The Complexity of the California Recall Election", PSOnline January 2004.

der effect implies that the candidates who are first win more votes, then the candidates that are included in the remainder will be more likely to do better overall. This effect is examined in detail with the data set in the next section.

Third, the lack of randomization matters if the districts are not completely homogeneous. Suppose that c does divide perfectly into 80 and suppose there is no ordering effect. However, the system of rotation can itself produce something that could alter the outcome of the election. Districts may vary in size, political preferences, and demographics. For example, consider all the assembly districts that are a multiple of 6 (i.e. 1, 6, 12, 18, 24, etc.). Each one of these districts, based upon the numbering system that begins in Northern California and scrolls across the state, occurs almost always near the coast (with the exception of Assembly District 60). One voting pattern that is observed in California is often that the coastal regions tend to be more liberal than the interior. Since the population of the districts is not homogeneous in its political preferences, the process of rotation can give an advantage to one particular candidate after the first order is drawn.

4 Previous Literature on Ballot Order Effects

This section should be prefaced by noting that there are several methodological concerns about the most recent and most significant previous literature. Many of these papers address one or two of the concerns, but none address all of them and thus it is possible that each study, due to a particular assumption or statistical technique, has drawn inferences which the work could disagree with. In particular, there are three methodological concerns. The first concern is about the compositional nature of the data. It is crucial to note that an increase in any party's vote share in a district due to a ballot order effect will affect the other party's vote shares. The correlation between the party vote shares helps to produce more precise coefficient estimates by improving efficiency. A second concern is to note that inferences are made using data which comes from samples which are not balanced in terms of the covariates. Using unbalanced samples could produce biased estimates of the ballot order effect. Third, much of the previous literature fails to examine the effect of multiple ballot positions. It is possible that while candidates appear to get an increased vote share from being first that they might also get an increased vote share from being last. Below several of the most significant studies in this area are discussed and the conclusions and methodological assumptions are examined in detail, with particular concern as to which methodological assumption may in fact drive the results.

Recent examination of the ballot order effect began with Miller and Krosnick's 1998 analysis.[15] The authors began by examining 26 earlier published studies on ballot ordering effects. They found significant methodological flaws in 18 of these studies, rendering their conclusions moot (though 14 of the 18 did find that candidates listed early did better than candidates listed later on ballots). Six of the previous studies had stronger underlying methodologies, and these produced mixed results. Some of these six found that candidates listed early did better but others found that candidates listed last did better. Miller and Krosnick found only two previous studies with reliable methodologies; neither study found any ordering effect.

Miller and Krosnick then undertook their own analysis focused on 1992 elections in the three largest counties in Ohio – Franklin, Cuyahoga and Hamilton Counties – each of which had different procedures for listing candidate names on the ballot. Of the 182 county-elections studied by Miller and Krosnick, only 40.7% of these observations had a statistically significant ballot order effect, with 1.6% of cases having a negative effect. Thus, 39.0% of cases in their study had a statistically significant positive ballot order effect, where by being listed first the candidate received more votes than by being listed last. Notice, furthermore, that the ballot order effect varied dramatically across the three

[15] Joanne E. Miller and Jon A. Krosnick, The Impact of Candidate Name Order on Election Outcomes, 62 PUBLIC OPINION QUARTERLY 291 (1998).

counties: 72.2% of the cases in Franklin County were statistically significant, 31.3% in Cuyahoga, and 20.0% in Hamilton County; factoring out the cases that were negative and significant, the heterogeneity across counties is more dramatic, as Cuyahoga's estimate drops to 27.7%.

In general, Miller and Krosnick find little overall difference in the ballot order effect in two-candidate races relative to three-candidate races. They find that 42.7% of the two-candidate elections have statistically significant ballot order effects relative to 38.4% of the three-candidate races.[16] In Franklin County, the ballot order effect appears significant in 71% of two-candidate and 73.9% of three-candidate races. However, in Cuyahoga County the rate is slightly higher for two-candidate races, but the opposite is true in Hamilton County.[17]

Miller and Krosnick use linear regression with the vote percentage for each candidate in a precinct as the dependent variable and a variable indicating the name order as the independent variable. This method ignores the compositional nature of the data and does not address whether or not they have a balanced sample.

More recent research has been published by Koppell and Steen (2004).[18] They focus on precinct-level data from the 1998 Democratic primary in New York City. Their results indicate that in the four statewide primaries in their sample, candidates received a statistically greater fraction of votes when listed first. For example, their analysis indicates that Democratic gubernatorial candidates received a vote advantage of 2.3% when in the first ballot position.[19] Koppell and Steen also looked at the local races, finding that 67 of

[16]Note that after factoring out the three negative and significant cases, the percentage drops to 39.6%

[17]Krosnick has extended his analysis to include the 2000 elections in Ohio, North Dakota, and California (J.A. Krosnick, J.M. Miller, and M.P. Tichy, An Unrecognized Need For Ballot Reform: Effects of Candidate Name Order, in A.N. Crigler, M.R. Just, E.J. McCaffery, *Rethinking the Vote: The Politics and Prospects of American Election Reform*, N.Y., N.Y., Oxford University Press.).

[18]Jonathan GS Koppell and Jennifer A. Steen, The Effects of Ballot Position on Election Outcomes, 66 JOURNAL OF POLITICS 267 (2004).

[19]See Table 1 of their paper.

75 candidates in the first ballot position received more votes than they expected, while in 17 of 75 cases the advantage was statistically significant. Last, their analysis indicates that the magnitude of the effect of being first on the ballot seems greater for down-ballot races than for the top-of-the-ticket races (they estimate that the average effect of being first on the ballot for the state committeewoman race was almost 4.5%, compared to 2.3% for candidates for governor).

Thus, Koppell and Steen present a straight-forward and compelling case, finding that candidates in the first position on the ballot in the 1998 Democratic primaries appear to have received more votes than expected, and that sometimes the advantage to being first on the ballot was both statistically and substantively relevant. Significantly, their analysis focuses on primary elections, and only considers the advantage that might come to a candidate if she appears first on the ballot — not considering the advantage that might arise if she appears in other positions on the ballot, especially last. Furthermore, they do not address the compositional nature of the data nor whether or not they have a balanced sample.

More recently, Ho and Imai (2004) have examined California data, including state wide primaries and general elections from 1978 through 2002.[20] Employing a sophisticated statistical analysis of these data, Ho and Imai find little systematic evidence that major party candidates in general elections are favored by being first on the ballot.[21] They find that only minor party or nonpartisan candidates are advantaged by being listed first on the ballot. But they do interpret their results as showing a stronger advantage for

[20] Daniel E. Ho and Kosuke Imai, The Impact of Partisan Electoral Regulation: Ballot Effects from the California Alphabet Lottery, 1978-2002, manuscript, http://www.princeton.edu/ kimai/research/alphabet.html. Table 2 of their paper lists the races they concentrate on, which are most of the presidential, senate and gubernatorial primary and general elections from 1978-2002, and the primary and general elections for the other other statewide constitutional offices in 1998 and 2002.

[21] Ho and Imai also examine the effect of being on the first page of the ballot in the 2003 California gubernatorial recall election and again find no effect for major party candidates. Daniel E. Ho and Kosuke Imai, "Randomization Inference with Natural Experiments: An Analysis of Ballot Effects in the 2003 California Recall Election", *Journal of the American Statistical Association*, v. 2004/12/09.

major party candidates when listed first on the ballot in *primary* elections.[22] Again, as in much of the contemporary literature in this area, Ho and Imai are interested in the effects that being first on the ballot might provide to a candidate, and like the other contemporary research papers in this area do find evidence they argue supports the hypothesis that being listed first on the ballot does provide an advantage to those candidates.

The analysis in Ho and Imai comes closest to addressing some of the methodological concerns. They address the lack of a balanced sample (and check for balance in the covariates before performing their analysis) and acknowledge the lack of a random sample. Given that they observe what they term ”systematic random treatment assignment” as opposed to ”random treatment assignment”, they examine all of their covariates and find that despite the systematic rotations their covariates are close to being balanced. They then proceed to look at the average treatment effect of each position relative to the other positions. Thus, they do succeed in examining each ballot position. However, their analysis hinges upon bivariate treatments (first or not, second or not, etc). This means that they cannot compare across treatments (the effect of being first against the effect of being second second against the effect of being third, etc.) because they are able to only examine treatments that have binary values. Finally, their analysis does not incorporate any measure to address the compositional nature of the election data.

In summary, the recent literature has focused on the effect of being first. It is possible, however, that there are multiple locations on the ballot which result in a positive or negative change in vote shares; the first ballot position may in fact be the most preferable position but it is not sufficient to simply compare being first to being located elsewhere

[22]Methodologically, Ho and Imai’s technique relies on the assumption of “no interference among units”, which here means ignoring the compositional nature of the data. As Ho and Imai state, “the assumption is violated in an analysis that pools candidates, typical in this literature: since the candidate vote shares in one district must sum to 1, a ballot order effect on one candidate necessarily affects the remaining candidates” (page 13). Ho and Imai state that they have reproduced their analysis using another technique that relaxes this assumption, and that those “results are largely consistent with those presented here” (page 14). They do not report those results in the current version of their manuscript.

unless being located elsewhere is guaranteed of producing a lower vote share. In order to make that assumption it is necessary to evaluate multiple positions on the ballot, especially the effect of being last. Also, these papers fail to account for the fact that an increase for one candidate in terms of vote share is likely to imply a decrease for another candidate's vote share. This implies that the individual candidate's vote shares are not independent, an assumption which is crucial in OLS. Furthermore, with the exception of the Imai and Ho paper, none of these studies examine whether or not they have a random sample. Without a balanced sample it is not possible to directly compare the electoral outcome of one geography to another, thus possibly invalidating the counterfactuals: "this candidate would have done better if she had come first in this district". While Imai and Ho do succeed in creating a balanced sample and evaluating average treatment effects, their analysis also hinges upon the assumption that all candidate's vote shares are independent.

This paper does employ a more sophisticated set of tools for the analysis of this problem. In particular, the analysis includes a step to pre-process the data to produce a balanced sample. Given that the treatment (rotation location) is multi-valued, the data is balanced by matching using a group distance metric. On both the balanced and unbalanced data SUR is employed after transforming the vote shares to log ratios so as to account for the compositional nature of the data. These techniques are discussed in detail in the methods section.

5 A New Methodology for Studying the "Ballot Order Effect"

This paper makes several contributions to the study of the ballot order effect. First, 15 contests are examined encompassing 94 candidates in two general elections (1998 and 2002) in California. This permits us to examine whether or not the ballot order effect has any variance across party or type of contest. Next, this paper contributes several methodological advances to the study of the ballot order effect. A statistical model is used based on recent advances in the study of multiparty and multicandidate elections by both Katz and King as well as Tomz, Tucker and Wittenberg.[23] Also, an algorithm is generated for producing a matched dataset on multi-treatment data which relies heavily upon the advances Imai and Van Dyke.[24] This next section will discuss each of these contributions, with the matching algorithm described in detail at the end.

The 1998 and 2002 California general elections were chosen for three reasons. First, each was a California state-wide election year, where the visibility of the contests ranged from California governor to the state controller. These two elections differ only in 1 contest – the 1998 election included a contest for a U.S. Senate seat – and are likely to have attracted a similar electorate. Using this data provides a total of ninety-four races to examine across the 80 Assembly Districts in California. Second, the all election returns are readily available in a format suitable for analysis. There is a very detailed data set available to study the impact of candidate order on the ballot, as it is possible to employ a census tract-level (and precinct level) database with a wide array of control variables. Third, as California's primary process has been in flux in recent years, the focus is on recent general elections.

Specifically, data was obtained at the state-wide election level from the 1998 and 2002 elections from the Statewide Database, which collects and archives data from California elections.[25] The database includes the census tract election return file in the 1998 general election and the precinct election return file in the 2002 election. While although one of

[23]Jonathan Katz and Gary King, A Statistical Model for Multiparty Electoral Data, 93 AM. POL. SCI. REV.15 (1999). Michael Tomz, Michael, Joshua A. Tucker and Jason Wittenberg, An Easy and Accurate Regression Model for Multiparty Electoral Data, 10 POLITICAL ANALYSIS 66 (2002).

[24]Imai, Kosuke and Van Dyk, David. "Causal Inference with General Treatment Regimes: Generalizing the Propensity Score", *Journal of the American Statistical Association*, September 2004, Vol. 99, No. 467, Theory and Methods.

[25]Statewide Database, http://swdb.berkeley.edu/index.html.

the datasets is organized by census tract and the other by precinct, fortunately each observation is located within an Assembly District. Any observation that is not completely contained within a single Assembly District is discarded.

The dependent variables of the analysis are the vote shares in each census tract/precinct received by every candidate on the statewide ballot for each contest. Thus, to take the governor's race as an example, we had seven dependent variables, one for the respective percentage of the total vote received by the candidates running as Democrats, Republicans, American Independence Party (AIP), Green, Libertarian, National Law Party, and Peace and Freedom Party. Our goal is to test whether or not the appearance of each party's candidate first or last on the ballot, relative to appearing in the middle of the ballot, produces a marginal increase or decrease in the party candidate's vote share.

This paper provides four methodological improvements to introduce, relative to previous work in this area. The first improvement is that to produce a balanced dataset for each contest and ensure that the lack of randomness in assigning orderings does not affect the results. That is, it is possible that without accounting for the lack of randomness some candidate orderings could occur in precincts or census tracts that are significantly different than others. One consequence of this could be that we observe Republican candidates getting additional votes when they are placed first on the ballot and erroneously assume that Republican candidates are more likely to benefit from being first, when instead the reality was that Republicans happened to be first in highly-Republican districts. Thus, it is crucial to check the sample to ensure that the data itself will not lead to biased estimates of the ballot order effect. The second improvement is to incorporate control variables into a multivariate statistical model to account for differences in vote shares across census tracts and assembly districts. The third improvement is that both primacy and latency are controlled for here, where much of the previous literature has focused entirely upon primacy. The fourth improvement is to utilize a more appropriate statistical framework for estimating the marginal effects of candidate name order using a method appropriate for multiparty data, holding constant the control variables.

In the available tract-level and precinct-level data, there are an important array of control variables for each observation. First, in each general election there were a number of initiatives and propositions included on the statewide ballot. The voting patterns on these initiatives provide an opportunity to include a measure of the observation's ideology. Accordingly, I undertake a factor analysis of each observation's vote on all eleven proposition and initiatives, and the factor score from this analysis is used as a measure of ideology in the multivariate statistical model.[26] Second, there is partisan registration data for each observation – which enables variables to be constructed that measure the Democratic, Republican, Decline-to-state, and non-partisan registration in each observation. Third, there is some demographic information for each observation, including variables that measure the percentage of Hispanic registered voters in each observation and the percentage of registered female voters in each observation. Fourth, data is available to indicate which type of voting technology was being used in each particular assembly district. This data is included an indicator variable as to whether or not the the voters we using a punch card machine or some alternative technology. This set of independent variables is used as controls in the multivariate statistical model. These are all variables which may affect the percentage of voters who vote for a particular candidate or party. It seems likely that with the addition of these control variables it is more likely to be able to differentiate the percent of the voting population who was choosing a candidate based upon their preferences and the percent of the voting population who was choosing a candidate based upon that candidate's location. Note that all observations are not homogeneous and by controlling for the differences that affect electoral outcomes it is more likely to observe the ballot order effect.

[26]The results from this factor analysis are available from Sinclair.

Data is available on the randomized order of candidate names as they were used in Assembly District 1, and the pattern of rotation for the remaining Assembly Districts, it is simple to construct two dummy variables for each Assembly District and each candidate and to account for the effect of candidate name order on the ballot. There is one dummy variable for each candidate in a race, coded 1 for Assembly Districts where that candidate was first on the ballot, and zero otherwise. There is a second dummy variable for each candidate in a race, coded 1 for Assembly Districts where that candidate was last on the ballot, and zero otherwise. These two dummy variables for each candidate will allow us to estimate the effect of candidate ballot name order, across Assembly Districts, holding the control variables constant.

Given how these two indicator variables for the candidate's position on the ballot have been operationalized, it is now possible to determine if their relative position on the ballot had any statistical effect on their vote shares. The regression coefficients on these two dummy variables allow for the clear differentiation of four different types of ordering effects (here, for simplicity I call the dummy variable indicating whether or not the particular candidate was first on the ballot "first" and the dummy variable for whether or not the candidate was last on the ballot "last"). If there is a statistically significant and positive coefficient on the "first" dummy, that is evidence for primacy (as this indicates that the candidate received a statistically significant increase in votes where he or she was first on the ballot). If there is a statistically significant and positive coefficient on the "last" dummy, that is evidence for latency (as the candidate receive a statistically significant increase in votes where he or she was listed last on the ballot). If there is a statistically significant and negative coefficient for the "first" dummy variable, that is evidence for anti-primacy – here the candidate received fewer votes in precincts where he or she was listed first on the ballot. Last, if there is a statistically significant but negative coefficient on the "last" dummy variable, there is evidence for anti-latency, as here the candidate is receiving fewer votes in precincts where he or she is listed last on the ballot. The expectation is to observe both primacy and anti-latency based upon the survey literature that indicates that respondents are more likely to check the items listed first.[27]

Our statistical approach is the same as that recently developed by Tomz, Tucker and Wittenberg, and is similar to that earlier exposited by Katz and King. Katz and King noted that when estimating models with dependent variable involving multiparty election outcomes, important assumptions that underlay ordinary least squares (OLS) regression are untenable and thus OLS regression is likely to produce incorrect results.[28] One of these two problematic assumptions is that as the quantity of interest is the percentage vote that each candidate received in a tract; those percentages are necessarily bounded between 0 and 1. Since OLS assumes that the dependent variable is unbounded and continuous, OLS is likely to produce incorrect results in this setting. The second problematic assumption is the independence of each candidate's vote share. The application of OLS sequentially to each candidate's vote share assumes that these vote shares are independent, when in fact they are certainly not independent as the vote shares must all sum to 100%.[29]

Katz and King develop a methodology that is suitable for data involving three candidates or parties; Tomz, Tucker and Wittenberg have developed a more general procedure for data that include more than three parties or candidates. The Tomz, Tucker, and Wittenberg approach is simple and intuitive, involving first the transformation of candidate vote shares into log-odds ratios, and second, the use of seemingly unrelated regression

[27] Don Dillman, *Mail and Internet Surveys*, John Wiley and Associates, New York, 2000.

[28] Jonathan Katz and Gary King, A Statistical Model for Multiparty Electoral Data, 93 AM. POL. SCI. REV.15 (1999).

[29] Similar work in this area is by John E. Jackson, A Seemingly Unrelated Regression Model for Analyzing Multiparty Elections, 10 Political Analysis 49 (2002); James Honaker, Jonathan N. Katz, and Gary King, A Fast, Easy, and Efficient Estimator for Multiparty Electoral Data, 10 Political Analysis 84 (2002); and Mikhailov, Niemi and Weimer, "Application of Theil Group Logit Methods to District-Level Vote Shares: Tests of Prospective and Retrospective Voting in the 1991, 1993, and 1997 Polish Elections", Electoral Studies (2002).

on a series of J-1 regressions (where J indexes the number of candidates in the particular application).[30] The use of the log ratios alleviates the problems associated with the bounded nature of this vote share data, and the use of SUR allows the error terms across the J-1 regressions to be correlated, which permits the vote share of one candidate to be correlated with another.

First a balanced dataset is produced and then coefficients are recalculated using the Tomz, Tucker and Wittenberg approach.[31] Even after processing the data through the matching algorithm that there are only improvements with respect to the data that would have come from a randomized experiment, and therefore it is necessary to continue to incorporate the control variables. Thus, after producing a balanced sample using the algorithm for multi-valued treatment data, it is necessary to acknowledge that the balance of the data has improved but it is still necessary to reproduce coefficients using the Tomz, Tucker and Wittenberg estimation procedure. Interestingly, the coefficients produced are very similar to the earlier estimates.

6 Results from the 1998 and 2002 California General Elections

Below two sets of analysis are presented– first the data is analyzed using the original unbalanced dataset. Then analysis is presented using the balanced dataset. These two sets of results are then compared. All estimation is done using the Tomz, Tucker and Wittenberg estimation procedure, thus estimating J-1 regressions for each contest. In this situation, the coefficients are normalized by excluding the Republican party from the SUR analysis.

[30] Arnold Zellner, "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias", Journal of the American Statistical Association, 57(298):348-368.(1962)

[31] The decision to pre-process with matching and then use the parametric model is consistent with the recommendations of Daniel E. Ho, Kosuke Imai, Gary King and Elizabeth Stuart. "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference". Unpublished manuscript. http://gking.harvard.edu/projects/cause.shtml

On the right-hand side of each of these regressions are variables measuring whether that particular candidate was first on the ballot, last on the ballot, ideology, partisanship, and some demographic attributes of the observation.

The full SUR results are not reported.[32] Instead, the analysis below begins by reporting a summary table of the results, which depicts the number of times in each contest that a significant primacy effect was found (that candidates who appeared first on the ballot in that contest had a significant and positive estimated coefficient on the dummy variable measuring if they were first at the $\alpha = .05$ level), the number of times in each contest that a significant latency effect was found (that candidates who appeared last on the ballot in that contest had a significant and positive estimated coefficient on the dummy variable measuring if they were last); the number of times in the each contest that a significant anti-primacy effect was found (candidates appearing first had a significant but negative estimated coefficient on the dummy variable measuring if they appeared first on the ballot); and the number of times in the contest that a significant anti-latency effect was found (candidates appearing last had a significant and negative estimated coefficient on the dummy variable measuring if they appeared last on the ballot). These results are given in Table 1 for the coefficients from the original data and in Table 2 for the coefficients from the balanced data. Note that because all coefficients are measured relative to the Republican party, that party is not included.

Table 1 Goes Here

Table 2 Goes Here

In Table 1 there is little systematic evidence that consistently supports the hypothesis that the primacy phenomenon prevails. There is evidence in every contest where some

[32] Available from Sinclair upon request. The calculations include nine independent variables on the right-hand side of each SUR regression; thus for each contest the data will produce J-1 equations of results, or 9 times J-1 coefficients and standard errors.

candidates appear to have received some marginal benefit by having their name at the top of the ballot, sometimes as many as 5 of 6 candidates in a contest. On the other hand, in Table 1 there is strong support for the hypothesis of latency – candidates whose name is at the bottom of the ballot sometimes also win additional votes – although this effect is not as strong as primacy.

The third and fourth columns of Table 1 provide results for anti-primacy and anti-latency. These are situations where the coefficients obtained from SUR produced statistically significant but negative estimates for the impact of the candidate's name at the top or bottom of the ballot. Notice here that anti-primacy and anti-latency effects occur nearly as often as primacy and latency. But anti-primacy and anti-latency do not prevail. In some contests (like the Treasurer in '02) there are few candidates who did better when they were first or last on the ballot.

The summary results presented in Table 1 do not provide support for the hypothesis that when candidates are listed first on the ballot they *necessarily* receive a greater vote share than when listed elsewhere on the ballot. There is evidence for a complicated picture: sometimes being first helps some candidates, but sometimes being last increases a candidate's vote share. In other cases, being first or last appears to statistically decrease the candidate's vote share, a result that is completely opposite of what was expected based upon the primacy and latency hypotheses.

In Table 2 there are results that are again similar in that there are no guarantees with respect to the direction or existence of the ballot order effect. There are more statistically significant coefficients overall, and the gap between the number of primacy and anti-primacy coefficients shrinks, as does the gap between latency and anti-latency.

Table 3 provides the same basic summary of the SUR analysis, but by party. That is, there is the number of partisan candidates for whom the estimates indicate the presence of each effect (primacy, latency, anti-primacy and anti-latency). This breakdown of the multivariate analysis gives us the ability to determine if there are any patterns in primacy or latency, for specific partisan candidates or for types of parties (for example, major relative to minor parties).

Table 3 Goes Here

In brief, no clear pattern emerges from Table 3. Democratic candidates seem slightly less likely to have anti-primacy effects in this election. AIP, Green, Natural Law and Libertarian candidates are likely to experience primacy and anti-latency effects. Peace and Freedom and Reform candidates were equally likely to experience all four effects Again, the conclusion from the presentation of the SUR results as summarized in Table 3 is that there is no clear pattern, by party of the candidate.

One caveat in this analysis is simply that two different elections are being examined and it is possible that different electorates (particularly more informed ones) may respond differently to candidate name orderings. However, including all of the statewide contests for these years does seem to demonstrate the lack of systematic evidence for an increase in a candidate's vote share due to one particular ballot placement. There is a great deal of variance across elections, with some contests and some elections having higher primacy or latency effects than others. These results are presented graphically for primacy and latency as well as for all coefficients to demonstrate that no clear pattern emerges. The graphs depict the coefficient along with the 95% confidence interval.

Examining the data with respect to all the coefficients observed provides a clear pattern of the distribution of ballot order effects. Note that a great deal of coefficients have 95% confidence intervals that fall completely below or above zero – this clearly depicts the argument for both a positive and a negative ballot order effect.

Figure 4 Goes Here

Looking for carefully at the coefficients for the effect of being placed first on the ballot, there is again a mixed pattern of both positive and negative effects where the 95% confidence interval is bounded away from zero. Note there are in fact more positive effects than negative effects.

Figure 5 Goes Here

Finally the graph is examined for the coefficients in the instance of being placed last on the ballot. There are a combination of positive and negative effects; note there are more negative effects than positive effects where the coefficient has a 95% confidence interval bounded away from zero.

Figure 6 Goes Here

The graphs above demonstrate the complicated picture which emerges in the case of the ballot order effect. While it does appear that more candidates benefit from being first, it is not clear at what point some candidates will benefit from being last nor lose vote shares from being first.

The coefficients from the matched estimates are presented next. Again, no clear pattern emerges with regard to primacy or latency. It is interesting to compare these coefficients to the coefficients above. Note that these estimates are slightly different than the earlier estimates in that voting technology is no longer included as a covariate.[33] Incorporating both the first- and last-coefficients provides a picture extremely similar to the unmatched data – a great deal of coefficients are bounded away from zero with their 95% confidence intervals.

Figure 7 Goes Here

[33]Initial analysis indicates that including voting technology would not change the analysis. For further details please contact Sinclair.

Note here that this graph appears to follow an identical trend to that in Figure 4. If the number of statistically significant coefficients (using the balanced dataset) was counted there is a similar trend. In Table 1 in the last row note the sum of significant coeficients from the balanced dataset. There are significantly more statistically significant coefficients. However, the trend is the same – there is no clear pattern which predicts any particular candidate or party is likely to benefit from being first or last. If anything, it became slightly more likely that candidates would in fact be negatively affected in the first or last position, as the number of statistically significant primacy coefficients increased by 7 while the number of anti-primacy coefficients increased by 8, and the number of statistically significant latency coefficients increased by 8 while the number of anti-latency coefficients increased by 9.

Next the analysis will break out the primacy and latency coefficients for the balanced sample to compare directly whether or not a candidate would prefer to be first or last.

Figure 8 Goes Here

Figure 9 Goes Here

7 Conclusion

In conclusion, the statistical study of the candidate ballot order effect has analyzed a wide variety of different electoral candidates and all of the candidates on the statewide ballot in the 1998 and 2002 California general elections. The analysis employed a multivariate statistical model that controlled for a number of other important variables that are likely to influence the outcomes of these contests (ideology, partisanship, voting technology, and demographic attributes of census tracts). A statistical methodology was utilized that is more appropriate for this type of electoral data than have previous studies. There is little systematic evidence that indicates that candidates are necessarily benefited in terms of their vote share by being listed first on the ballot. Rather, sometimes candidates appear to benefit by being first, other times being first actually decreases their vote shares. Sometimes candidates benefit by being last on the ballot, but sometimes they also do worse if they are last on the ballot.

These results differ from those found in the rest of the literature based upon four principle reasons. First and foremost, the ballot order problem is examined in the context of using a balanced dataset to ensure that effect observed is in fact due to ballot order and not simply an effect due to one of the covariates.Second, the effects on candidates in multiple ballot positions were examined. The findings here are surprising – that sometimes it was the case that candidates actually lost votes by being located first or last. It seems likely that there is a complicated process which determines the focal point of each voter. Ballot layout and voting technology are two possible factors. Third, a statistical procedure was used which accounted for the possible relationship between vote shares between candidates within a census tract. Finally, factors are accounted for which may determine different vote shares (the control variables).

The results also demonstrated that, regardless of the direction of the ballot order effect, the impact of being first or last on the ballot is generally of very small magnitude. Admittedly there is little change when the pattern of effects is examined using the balanced sample. However, without accounting for the lack of balance, it would have been impossible to know that this would be true. It is comforting that while these effects do exist, they appear with no particular pattern for no particular party and for no particular race. Furthermore, they are of fairly small magnitude. The largest positive increase a candidate received for being first was the Lt. Governor's race, with an increased primacy effect of 2.59% (with a 95% confidence interval ranging from 1.86 to 3.27). Thus, there is little reason to believe, that once the analysis controls for partisanship, ideology, and demographic factors, that ballot order effects (no matter their direction) are potentially large enough to influence anything but a very small fraction of races that are exceedingly close.

In brief, there is no evidence in the presentation that the candidate order effects – be they for being first or last, or be they positive or negative – seem more substantial in any type of California statewide contest, or for any party. The ballot order effects that are significant in this analysis appear to be uniform across contests and parties. That is, the analysis does not indicate that ballot order effects (of any direction, or for either being first or last) are not more likely for minor party candidates and are not more likely in less salient races.

Future research will need to consider when the ballot order effect is likely to be most salient. The Koppel and Steen and Imai and Ho papers suggest a stronger ballot order effect in *primary* elections. Thus, primacy may be more salient in primacy elections because (1) the partisan voting cue is absent; (2) voters are less likely to have formed opinions about candidates in primacy elections; or (3) some other factors. Ballot order effects could also vary by region, education level, and the general salience of the election.

The analysis has demonstrated that the ”ballot order effect” is not a single effect that may be characterized by candidate position, by contest or by party. Furthermore, our estimates indicate that when it exists, the ballot order effect is extremely small. Incorporating those two facts together results in the positive conclusion that California's current rotation and randomization system is, for the most part, not distributing additional vote share to any particular candidate or party in any systematic process. The only possible improvement would be additional rotations to eliminate any possible effects. It seems possible that even with a small effect, the additional votes procured by being in the ”remainder” could give a candidate a boost in vote shares. However, given that additional printing costs would be incurred for additional rotations, further research needs to explore other years and other types of elections.

8 Tables and Figures

Table 1: Summary Results from SUR Analysis by Contest, Original Data

	Primacy	Latency	Anti-Primacy	Anti-Latency	# of Candidates
Governor ’98	1	0	3	1	7
U.S. Senate ’98	1	2	2	2	7
Lt. Governor ’98	2	3	2	3	7
Sec. of State ’98	1	1	5	5	7
Controller ’98	1	2	1	1	7
Treasurer ’98	2	2	2	3	6
Attorney General ’98	1	1	2	3	5
Insurance Commissioner ’98	2	2	3	3	6
Governor ’02	4	1	0	2	6
Insurance Commissioner ’02	5	3	0	1	6
Lt. Governor ’02	5	1	1	2	7
Attorney General ’02	4	1	0	2	5
Sec. of State ’02	5	3	1	3	7
Controller ’02	3	1	0	2	5
Treasurer ’02	5	3	0	1	6
Total	42	26	22	34	94

Table 2: Summary Results from SUR Analysis by Contest, Balanced Data

	Primacy	Latency	Anti-Primacy	Anti-Latency	# of Candidates
Governor ’98	1	0	3	2	7
U.S. Senate ’98	2	2	3	2	7
Lt. Governor ’98	2	2	2	3	7
Sec. of State ’98	1	1	5	4	7
Controller ’98	2	3	2	0	7
Treasurer ’98	1	2	2	3	6
Attorney General ’98	0	1	2	2	5
Insurance Commissioner ’98	3	2	2	3	6
Governor ’02	4	1	0	2	6
Insurance Commissioner ’02	5	3	0	1	6
Lt. Governor ’02	5	1	1	2	7
Attorney General ’02	4	2	0	2	5
Sec. of State ’02	5	3	1	3	7
Controller ’02	2	2	0	2	5
Treasurer ’02	2	3	0	1	6
BalancedTotal	49	34	30	45	94
Original Total	42	26	22	34	94

Table 3: Number of Party Candidates By Name Order Effect, Summary Results from SUR Analysis

Original Data					
	Primacy	Latency	Anti-Primacy	Anti-Latency	# of Party Candidates Across Races
Democrat	9	6	4	6	15
AIP	10	6	3	6	15
Green	7	6	1	2	9
Liber.	7	1	5	10	14
Natural Law	5	4	4	5	12
PAF	2	3	3	4	8
Reform	2	2	2	2	6
Balanced Data					
	Primacy	Latency	Anti-Primacy	Anti-Latency	# of Party Candidates Across Races
Democrat	7	7	4	7	15
AIP	10	7	4	7	15
Green	7	7	1	1	9
Liber.	8	1	3	7	14
Natural Law	5	2	5	5	12
PAF	2	2	4	3	8
Reform	3	2	2	2	6

Figure 1: Example, Balanced and UnBalanced Data

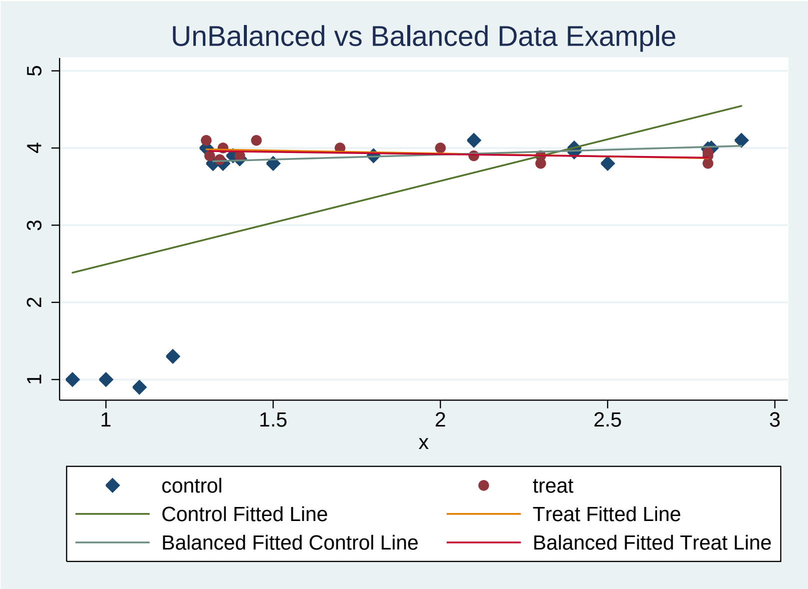

Figure 2: Unbalanced Percent Decline-to-State

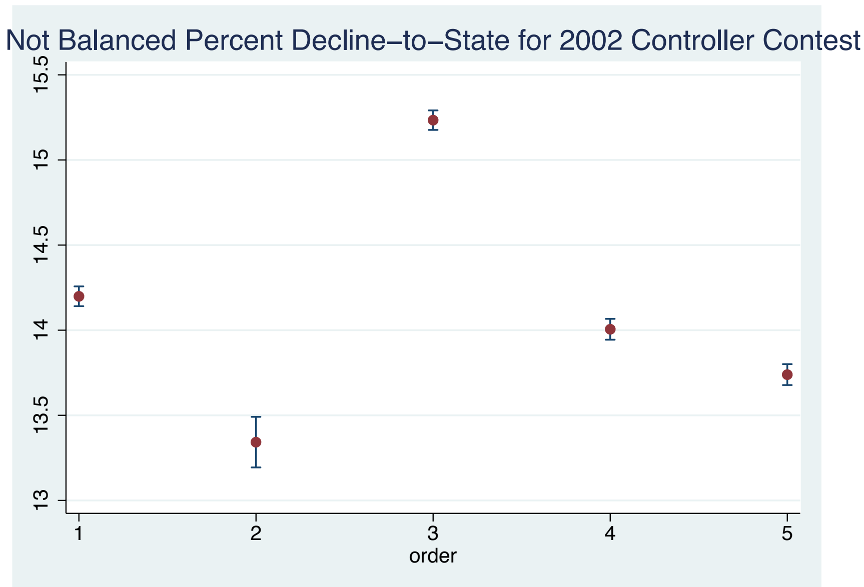

Figure 3: Balanced Percent Decline-to-State

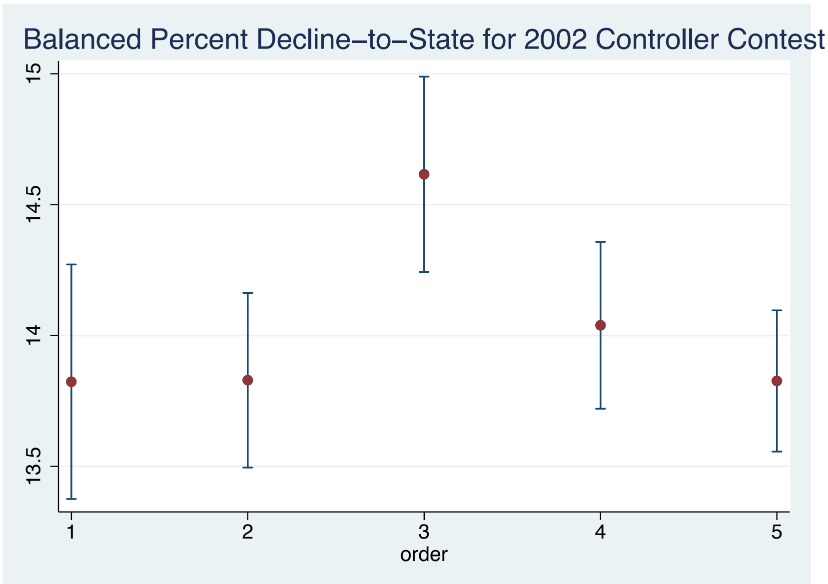

Figure 4: All Coefficients

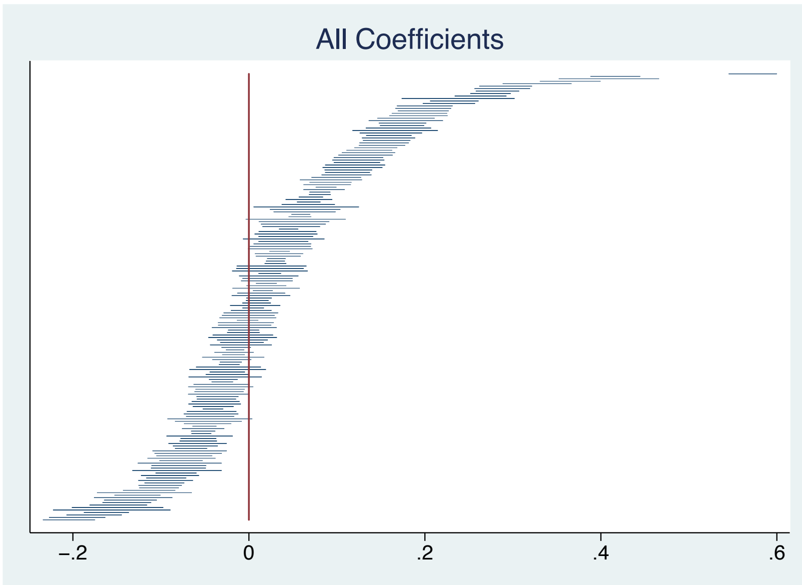

Figure 5: Primacy Coefficients

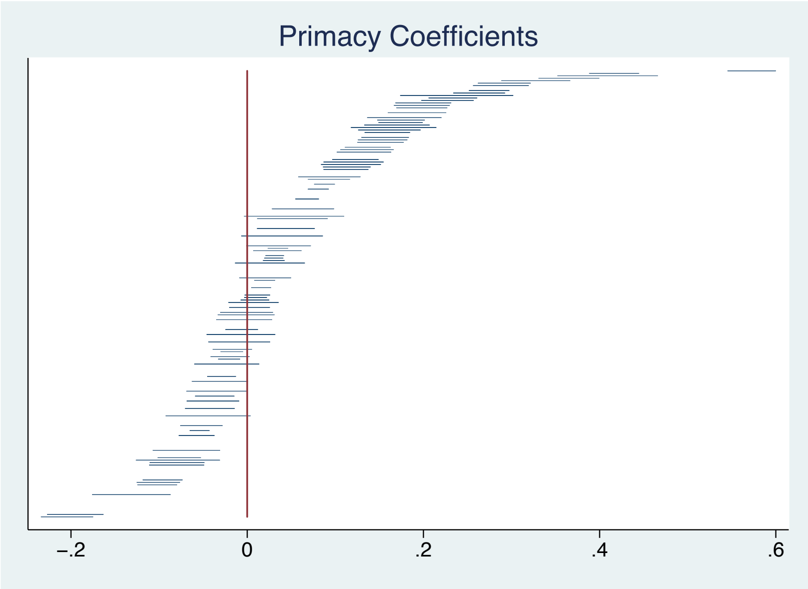

Figure 6: Latency Coefficients

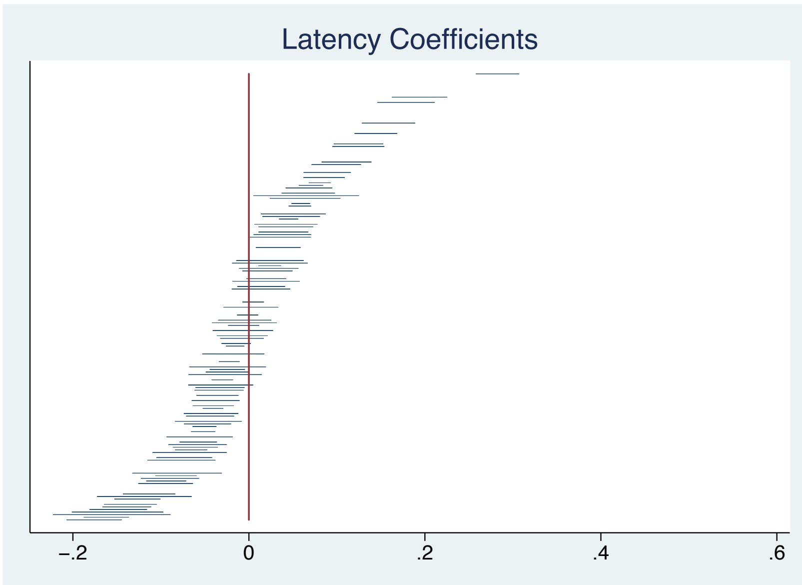

Figure 7: All Coefficients Using Balanced Data

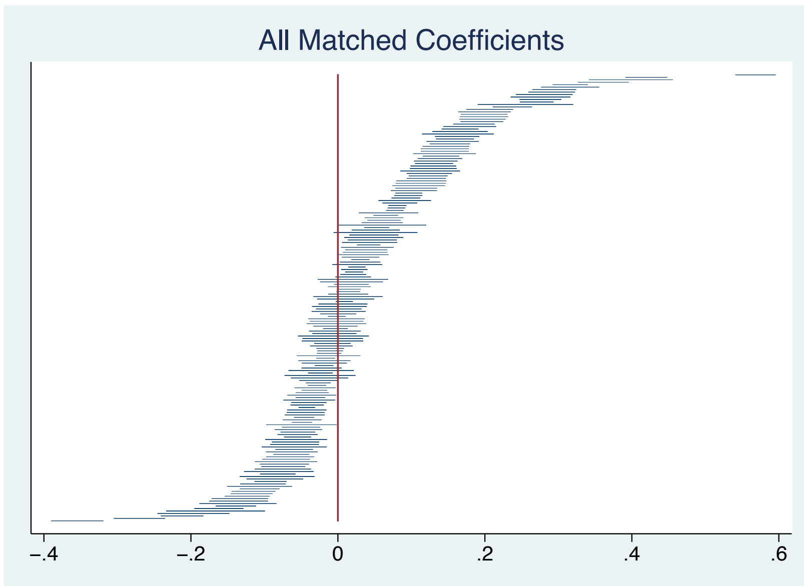

Figure 8: Primacy Coefficients Using Balanced Data

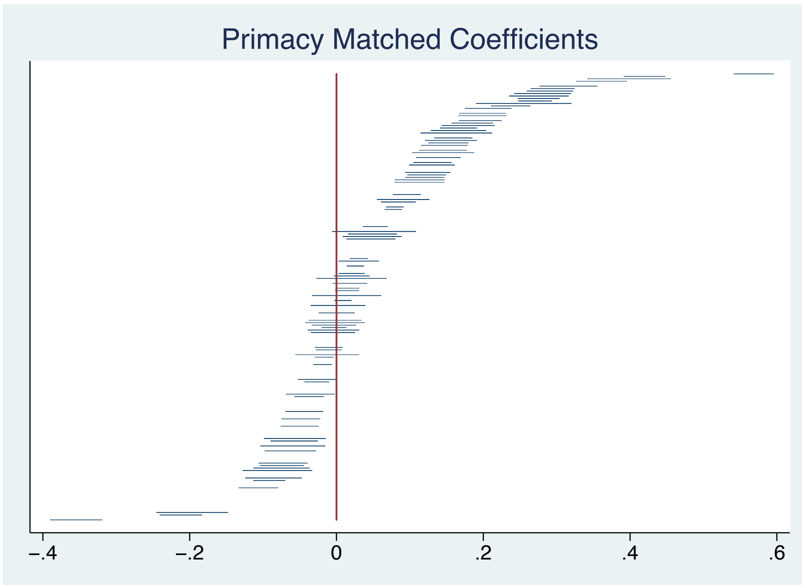

Figure 9: Latency Coefficients Using Balanced Data