Exploiting Tom DeLay: A New Method for Estimating Incumbency Advantage *

Jasjeet S. Sekhon† and Rocío Titiunik‡

Associate Professor
Travers Dept. of Political Science
UC Berkeley

Ph. D. Candidate
Agricultural and Resource Economics
UC Berkeley

9/19/2008 (14:53)

*For valuable comments we thank Steve Ansolabehere, Bob Erikson, Shigeo Hirano, Luke Keele, Gary King, Walter Mebane, Jr., Rebecca Morton, Eric Schickler, Jonathan Wand, and participants of the Society of Political Methodology's Annual Summer Meeting, July 18–21, 2007. We thank Sam Davenport in the Texas Legislative Council, Nicole Boyle in the California Statewide Database, Gary Jacobson and Jonathan Wand for providing data. Sekhon also thanks the MIT Department of Political Science for hospitality during the summer of 2007. Matching software which implements the technology used in this paper can be downloading from `http://sekhon.berkeley.edu/matching/`. All errors are our responsibility.

† <sekhon@berkeley.edu>, http://sekhon.berkeley.edu/, Survey Research Center, 2538 Channing Way, UC Berkeley, 94720

‡ <rocio@are.berkeley.edu>, http://are.berkeley.edu/~rocio, Survey Research Center, 2538 Channing Way, UC Berkeley, 94720

Abstract

We propose a new method for estimating incumbency advantage which relies upon the successive implementation of multiple redistricting plans, and demonstrate that the manner in which previous work has used redistricting to identify the causal effect of incumbency results in biased estimates. Strikingly, we show that even if voters were redistricted at random, previous uses of redistricting as a research design would not yield unbiased estimates. Furthermore, even if the correct potential outcomes are used, the selection on observables assumption implicit in prior work is shown to be both theoretically implausible and to empirically fail a placebo test which our design passes.

We illustrate our method and the difficulties of previous methods using data from U.S. House elections in California and Texas. Contrary to the extant literature, we find that in these states there is no candidate specific personal vote—i.e., there is no personal incumbency advantage. We do, however, find a significant incumbent party effect. The absence of a candidate specific incumbency advantage is consistent with theoretical work which argues that existing positive estimates of incumbency advantage are plagued by selection problems.

1 Introduction

We propose a new method for estimating incumbency advantage which relies upon the successive implementation of multiple redistricting plans. We also demonstrate that the manner in which previous work has used redistricting to identify the causal effect of incumbency results in biased estimates because the wrong counterfactuals (i.e., potential outcomes) are used. This leads to the the surprising result that the current way in which scholars use redistricting to estimate incumbency leads to bias even if voters are assumed to be redistricted at random. Of course, in reality voters are not randomly moved during redistricting, and a selection on observables assumption must be made. Unfortunately, the selection on observables assumption implicit in prior work is theoretically implausible because it fails to condition on crucial covariates, and it empirically fails a placebo test which our design passes.

An extensive literature exists on whether the incumbency status of legislators in the United States affects their electoral outcomes. Indeed, it is one of the most studied topics in electoral politics.[1] While the exact magnitude of the estimated effect of incumbency varies across studies, there is widespread scholarly agreement on at least two issues: (i) being an incumbent has a positive effect on electoral outcomes, i.e. there is an *advantage* to incumbency, and (ii) this advantage was moderate during the first half of the 20th century (about 2 percent in terms of vote shares) and began to grow substantially in the mid-1960s (e.g., Erikson 1971; Ansolabehere and Snyder 2002; King and Gelman 1991).[2] Beyond this general agreement, however, the sources of the observed incumbency advantage and the causes of its growth remain highly contested. Some authors have emphasized the importance of direct officeholder resources such as name recognition, access to federal programs and pork, access

[1]Studies of the effects of incumbency for legislative offices include Alford and Brady (1989), Ansolabehere, Brady, and Fiorina (1988), Ansolabehere, Snyder, and Stewart (2000), Breaux (1990), Born (1979), Cox and Katz (1996), Cox and Morgenstern (1993), Erikson (1971), Erikson (1972), Ferejohn (1977), Fiorina (1977), Gelman and King (1990), Jacobson (1987), Jewell and Breaux (1988), King (1991), Krashinsky and Milne (1993), Krehbiel and Wright (1983), Mayhew (1974), Nelson (1979), Payne (1980).

[2]For a review on the debate about the causes of the increase in the incumbency advantage see Cox and Katz (1996) and Krehbiel and Wright (1983).

to technologies of public position-taking (Mayhew 1974), and opportunities to perform better constituency service (Fiorina 1977, 1989; Fenno 1978). Others have emphasized partisan dealignment, suggesting that incumbency per se may become a cue in deciding how to vote when partisan ties weaken (Erikson 1972; Nelson 1979; Burnham 1974; Ferejohn 1977). And other scholars have emphasized the ability of incumbents to scare-off high-quality challengers (Cox and Katz 1996; Levitt and Wolfram 1997; Jacobson and Kernell 1983).

All of these results notwithstanding, the literature faces formidable methodological challenges. Erikson (1971) was the first to recognize that traditional measures of incumbency advantage such as sophomore surge and retirement slump could be severely biased, and Gelman and King (1990) provided a formal analysis of these difficulties. The authors also proposed a method which estimates incumbency advantage under the assumption that candidates' decisions to run for election are exogenous to the votes they expect to obtain. However, if politicians make strategic entry and exit decisions, their proposed method does not provide a reliable solution to the problem of estimating the causal effect of incumbency.

Given these fundamental methodological difficulties, it is notable that much of the literature has accepted the premise of incumbency advantage for more than thirty years. An exception is Cox and Katz (2002) who argue that the observed advantage of incumbency is a spurious effect generated by the strategic entry of incumbents and challengers.[3] If incumbents' expectations of their electoral fortunes play an important role in their decisions to seek reelection and if incumbents' vote shares are partly based on party and not only on personal appeal, then a party's vote share will be larger when there is an incumbent of that party running and smaller when there is an open seat. Cox and Katz compare the average vote loss suffered by a party when its incumbent vacates the seat for voluntary reasons to the party's average vote loss when its incumbent vacates the seat for involuntary reasons and find the former to be larger than the latter, providing evidence that strategic entry is a severe source of bias.[4]

[3]Also see Ashworth and Bueno de Mesquita (2007) and Zaller (1998).

[4]Cox and Katz also emphasize the importance of the strategic entry decisions of challengers, and show that

As a way to partly avoid this selection bias, Ansolabehere, Snyder, and Stewart (2000) use the variation brought about by decennial redistricting plans to identify the causal effect of the personal appeal that the incumbent has given her history with her constituents. They sometimes refer to this personal appeal as the personal vote and other times as the benefits of "homestyle" or as "direct office holder benefits." Ansolabehere et al. exploit the fact that after redistricting most incumbents face districts that contain a combination of old and new territory, and hence face a combination of old and new voters. They analyze U.S. House elections at the county level from 1872 to 1990, and compare an incumbent's vote share in the new part of the district with her vote share in the old part of the district. Desposato and Petrocik (2003) employ the Ansolabehere et al. design to estimate the personal vote in California for the U.S. House and State Legislature elections using block-level data. Both studies find an average incumbency advantage of approximately 4 to 6 percent.[5] Carson, Engstrom, and Roberts (2007) use the design to estimate the personal vote in late-nineteenth-century House elections (1872–1900). They estimate the personal vote to be about 2.5% and note that during this time-period, nearly all of the incumbency advantage can be attributed to the personal vote.

Comparing the voting behavior of old voters and new voters within electoral races is intuitively appealing, as this approach holds constant many factors of the electoral environment that are likely to affect the electoral success of the incumbent. For example, since old voters and new voters face the same candidates, observed differences between their voting behavior cannot be attributed to the varying quality of challengers. However, while using redistricting as an empirical strategy to identify the incumbency advantage is promising, its correct implementation requires careful consideration of the manipulation involved in redistricting.

As we discuss in detail in Section 2, the manner in which previous work has used redistricting to identify the causal effect of incumbency leads to bias because the correct potential

strong challengers have been avoiding incumbents in the post-1966 period. In this period, strong challengers are more likely to enter foreseeably open seats contents than unforeseeably open seats contests.

[5]For Ansolabehere et al. (2000) this figure corresponds to the 1972 − 1988 period. Consistent with previous literature, their estimate of the incumbency advantage is smaller for earlier periods.

outcomes are not used. By carefully analyzing the redistricting manipulation, we reach the striking result that if even voters were assumed to be redistricted randomly, this design would result in biased estimates. Naturally, randomization is a simplifying assumption that does not hold true in reality, and therefore a selection on observables assumption must be made. But we also show that the selection on observables assumption implicit in previous uses of redistricting is theoretically implausible, and it empirically fails a placebo test even when the correct potential outcomes are used.

We propose a research design that uses the correct potential outcomes and makes a selection on observables assumption which passes the placebo test. This design relies on the succesive implementation of multiple redistricting plans. We estimate it with data from congressional elections in Texas, where two different redistricting plans were succesively implemented in 2002 and 2004. When multiple redistricting plans are not available, we propose the "second-best" design, which is similar in spirit to the design proposed by Ansolabehere et al. (2000). We estimate this second design using data from congressional elections in California and Texas. In all cases, we estimate the effects of incumbency using Genetic Matching (Sekhon forthcoming) to achieve covariate balance.

In addition to posing methodological difficulties, using redistricting as an identification strategy leads to some conceptual ambiguities. In particular, we argue that even when new voters do not have the same history with the incumbent as old voters they may nonetheless respond to the incumbent's record of constituency service, this is, to the incumbent's personal appeal, if they manage to gain some knowledge about it. We propose to interpret this "old voters, new voters design" as estimating how quickly new voters learn about the type of their new incumbent and not the personal vote, since some crucial components of the personal vote are left out by construction of the design. As a consequence, we challenge the notion implicit in previous work that observing similar incumbent vote rates from old and new voters should be interpreted as evidence that constituency service is not important.

Contrary to previous work, we do not find an incumbency effect when voters are moved from one incumbent to another and the party of the incumbent remains the same. New voters seem to quickly learn the type of their new incumbent when they are moved from one incumbent to another incumbent of the same party. That is, voters quickly learn the type of their new incumbent well enough that they do not vote differently than old voters. We estimate a zero effect even though the standard Ansolabehere et al. (2000) "old voters, new voters" design (which fails the placebo test) estimates a highly significant positive incumbency advantage of about 5.8% in our data.

We do, however, find a significant incumbency effect when voters are moved from one incumbent to another and the party of the incumbent changes. When the party label of the incumbent does change, new voters are less likely to support their new incumbent than old voters, possibly because they underestimate the constituency benefits provided by the new incumbent.

The paper is organized as follows. In the next section we discuss our research design and in Section 3 we discuss how we interpret the redistricting estimand. Section 4 describes the data used in our empirical application, with some additional details provided in Appendices A and B. In Section 5 we outline our estimation method, Genetic Matching. Section 6 presents the empirical results, and Section 7 concludes.

2 Research Design

In this section, we examine in detail the conditions that must hold for the the variation introduced by congressional redistricting to identify the effect of incumbency status on electoral outcomes. Redistricting induces variation in at least two dimensions: a time dimension, as voters vote both before and after redistricting, and a cross-sectional dimension, as some voters are moved to a different district while others stay in the district they originally belonged to. We are interested in learning about the incumbency advantage by comparing the behavior of voters who are moved to a new district (*new voters*) to the behavior of voters whose district remains unchanged across elections (*old voters*).

As first recognized by Ansolabehere et al. (2000), the attractiveness of redistricting as a research design relies upon the fact that old voters and new voters face the same electoral environment in the elections following redistricting. Indeed, this design holds constant many factors that have long been considered sources of bias, such as candidate quality and race-specific cues. However, a closer look at old voters and new voters reveals that redistricting is far from producing as clean a conterfactual group as has been assumed by previous work. In order to obtain valid estimates of the incumbency advantage from redistricting, a very careful analysis is needed.

Figure 1 illustrates some of the problems. This figure shows the empirical Quantile-Quantile (QQ) plots of the vote share received by the incumbent U.S. House member in 2000, comparing electoral units which were to be redistricted to a different incumbent in the 2002 election to units which were to remain with the same incumbent. Figure 1(a) shows the QQ plot for California, while Figure 1(b) shows the QQ plot for Texas. In both states, the empirical quantiles of the 2000 incumbent's vote share of the units which are to remain with the same incumbent in 2002 are everywhere larger than the empirical quantiles of the units whose incumbent is to change in 2002. This shows that those units which are to be redistricted vote for their *old* incumbent at a systematically *lower* rate than units whose incumbent will not change. In other words, there is a bias in which units are moved, with units with a lower incumbent's vote share in 2000 being more likely to be moved to a different incumbent in 2002.

If part of this tendency of new voters to vote for their old incumbent at a lower rate persists in the future, then comparing old voters and new voters will be biased towards finding that old voters vote for the incumbent at a higher rate, this is, that there is a significant incumbency advantage. However, it could be argued that this initial difference between both types of voters is generated by their having different partisan attachments. To account for this possibility, figures 1(c) and 1(d) show the same QQ plots than figures

1(a) and 1(b), respectively, but this time the QQ plots are produced after units whose incumbent does not change are matched to units whose incumbent does change on their Democratic share of the two-party presidential vote, which is the measure of "normal vote" used by Ansolabehere et al. (2000). As can be seen in figures 1(c) and 1(d), restricting the comparison to units whose normal vote is on average identical does not eliminate the bias. Even when new voters vote for the Democratic presidential candidate at the same rate than old voters, they still vote for their old incumbent at a lower rate. Thus, a simple comparison of old voters and new voters conditional on presidential vote is not enough to produce valid estimates of the incumbency advantage.

We now show these methodological complications in a more rigourous way. First, we show that although new voters are naturally defined as the voters whose district changes between one election and another, there is an ambiguity in the way in which old voters are defined, as these could be either the electorate of the district to which new voters are moved (henceforth *new neighbors*), or the electorate of the district to which new voters belonged before redistricting occurred (henceforth *old neighbors*). Second, independently of how old voters are defined, we show that only under strong assumptions does the difference in the behavior of old voters and new voters identify the effect of incumbency. We propose different research designs to address these issues.

To illustrate the first point, we consider the following thought experiment. We imagine that just before election t a redistricting plan randomly redraws the boundaries of an arbitrary district (referred to as district A), in such a way that some voters that used to be in this district are randomly chosen and moved to a new district (referred to as district B). From the point of view of district B's incumbent, in the first election after redistricting (referred to as election t) voters that come from district A are *new voters* and voters that were originally in B are *old voters*. In principle, it seems natural to compare how differently these two groups vote for the incumbent and attribute the difference to an incumbency effect (or more precisely, to a personal vote), since both types of voters face not only the the same incumbent but also the same challenger, the same campaign, the same cues, etc.[6]

Moreover, the assumption of random redistricting seems to make this comparison even more attractive. Since randomization (if successful) ensures exchangeability between treatment and control units, we may be tempted to claim that in this hypothetical case B's old voters are guaranteed to be valid counterfactuals for B's new voters. But a crucial feature of this experiment prevents this claim from being true: while this randomization guarantees that voters who stay in A (A's old voters) and voters who leave A (B's new voters) are exchangeable, randomization says nothing about the exchangeability of B's new voters and B's old voters. In the absence of redistricting, B's new voters would have been in a different district than B's old voters and therefore nothing ensures that B's old voters are a good counterfactual for what would have happened to the new voters in the absence of redistricting, precisely because in the absence of redistricting both groups of voters would not have been in the same district at all.

In other words, the fact that B's new voters are originally in a different district than B's old voters implies that both types of voters have different histories –this is, at election $t-1$, B's old and new voters may have faced incumbents who belonged to different parties, or candidates who were of different qualities, or campaigns that were managed in different ways, etc. Since these factors are likely to affect how new voters react to their new incumbent, in order to obtain meaningful estimates of the incumbency advantage one needs a design that balances these covariates between treated and control groups. The crucial point is that the randomization we are considering does not guarantee balance in the covariates related to the history of new voters and new neighbors and hence, without further assumptions, it is not appropriate to estimate the incumbency effect.

Formally, let T_i be equal to 1 if precinct i is moved from one district to another just before election t and equal to 0 if precinct i is not moved to a different district before election t,

[6] Although this thought experiment places constraints on which precincts may move, the results are general. That is, the conclusions are the same if we assume that every precinct in every district in the state has a positive probability of moving to any other district. However, the notation and discussion becomes unwieldy.

and let D_i be equal to 1 if precinct i has new voters in its district at election t and equal to 0 if precinct i has no new voters in its district at election t. Let $Y_0(i,t)$ be the outcome attained by precinct i if $T_i = 0$ and $D_i = 0$ (the precinct is not moved and does not have new neighbors, i.e., these are voters who stay in A after redistricting), let $Y_1(i,t)$ be the outcome attained by precinct i if $T_i = 0$ and $D_i = 1$ (the precinct is not moved and has new neighbors, i.e., these are voters who are in B before and after redistricting), and let $Y_2(i,t)$ be the outcome attained by precinct i if $T_i = 1$ and $D_i = 1$ (the precinct is moved and has new neighbors, i.e., these are voters who are moved from A to B).[7] Of course, the fundamental problem of causal inference if that for every precinct we observe only one of its three potential outcomes. This is, we only observe the *realized* outcome, defined as

$$Y(i,t) = Y_0(i,t) \cdot (1 - T_i) \cdot (1 - D_i) + Y_1(i,t) \cdot (1 - T_i) \cdot D_i + Y_2(i,t) \cdot T_i \cdot D_i \tag{1}$$

This implies that we cannot compute individual treatment effects and hence we must concentrate on estimating average effects. As is common with observational studies, we focus on the average treatment effect on the treated (ATT). Given the set-up of our hypothetical experiment, the ATT can be defined in two different ways:

$$ATT_0 \equiv E[Y_2(i,t) - Y_0(i,t) \mid T_i = 1, D_i = 1] \tag{2}$$

$$ATT_1 \equiv E[Y_2(i,t) - Y_1(i,t) \mid T_i = 1, D_i = 1] \tag{3}$$

It can be shown that the following condition is sufficient for ATT_0 to be identified:[8]

$$E[Y_0(i,t) \mid T_i = 1, D_i = 1] = E[Y_0(i,t) \mid T_i = 0, D_i = 0] \tag{4}$$

[7]The potential outcome when $T_i = 1$ and $D_i = 0$ is not defined because it is not possible to be moved from one district to another and not to have new neighbors.

[8]For a formal treatment of these and related assumptions, see, for example, Heckman, Ichimura, and Todd (1997).

Similarly, it can be shown that the following condition identifies ATT_1:

$$E\left[Y_1\left(i,t\right)\mid T_i=1,D_i=1\right]=E\left[Y_1\left(i,t\right)\mid T_i=0,D_i=1\right] \tag{5}$$

In words, Assumption (4) says that voters who stay in A and voters who are moved from A to B would have attained the same average outcomes if they hadn't been moved and if they had not received new neighbors in their districts. Assumption (5), on the other hand, states that voters who are originally in B and voters who are moved from A to B would have attained the same average outcomes if A's voters would not have been moved and B's voters would not have received new neighbors.

This makes clear that randomization does not imply that B's old voters are a valid counterfactual for B's new voters: while randomization, if successful, ensures that Assumption (4) be satisfied (and hence that the average treatment effect defined by Equation (2) be identified), randomization does *not* imply Assumption (5). In other words, randomization ensures exchangeability between the set of voters for which $(1-T_i)\cdot(1-D_i)=1$ (i.e., voters who stay in A after redistricting) and the set of voters for which $T_i\cdot D_i=1$ (i.e., voters who are redistricted from A to B), but not between the latter set of voters and the set of voters for which $(1-T_i)\cdot D_i=1$ (i.e., voters who are originally in B).

Indeed, a close examination of Assumption (5) reveals that it is a rather peculiar requirement, since in the absence of redistricting voters in A would have been in a different district than voters in B. The assumption that they would have attained the same average outcomes is a very strong one precisely because in the absence of redistricting these voters would have been in completely different populations.

Of course, that randomization does not guarantee that Assumption (5) be satisfied does not mean that this assumption *could not* be satisfied, but the crucial point that we wish to convey here is that there is nothing in the redistricting process itself, even if randomly assigned, that would make it natural to assume that new neighbors and new voters are exchangeable. Henceforth, we will refer to the design that uses old neighbors as counterfactuals as the "best old-neighbors design" and the design that uses new neighbors as counterfactuals as the "second-best design".

2.1 Making the most of old and new neighbors

We have shown that under this experiment the group guaranteed to be a valid counterfactual for the new voters is not the new neighbors (i.e., the electorate in the new district to which new voters are moved), but rather the old neighbors (i.e, the voters that are left behind in the new voters' original district). However, the question arises of whether the best old-neighbors design is appropriate to estimate the incumbency advantage. On the one hand, using old neighbors ensures that both new and old voters are from the same district and hence from the same population at baseline. But this design also introduces important sources of heterogeneity, since it compares voters who at election $t-1$ are in the same district (and hence face the same electoral environment) but who at election t are in different districts (and hence face a different incumbent, a different challenger, a different campaign, etc.). In principle, one could restrict the universe of the comparison to reduce this heterogeneity (for example, one could restrict the old and new district to have the same incumbent's party and the same challenger's quality). However, there is a crucial difficulty in adopting this approach, as in order to induce homogeneity one would have to condition on characteristics of the environment *after* redistricting, and since these characteristics are likely to have been affected by redistricting itself one runs the risk of introducing post-treatment bias. We therefore conclude that the best old-neighbors design is not appropriate to estimate the effect of incumbency status on electoral outcomes. [9]

We have yet to establish a design that is both valid and appropriate for estimating the incumbency advantage. In the next subsection we propose what we consider to be the best

[9]Note, however, that this design could be used to estimate how voters react to a change in the race or ethnicity of their incumbent, since in this case one wishes to consider the different electoral environments which incumbents of different races or ehtnicities bring about.

design to estimate the incumbency advantage using redistricting. But before turning to this design, we consider additional methodological issues that arise if one decides to implement the second best design despite its difficulties.

Since Assumption (5) is not valid even with random assignment, we define a weaker version of this assumption:

$$E\left[Y_1\left(i,t\right) \mid T_i = 1, D_i = 1, X\right] = E\left[Y_1\left(i,t\right) \mid T_i = 0, D_i = 1, X\right], \tag{6}$$

where X is a vector of observable characteristics. Assumption (6) can be shown to identify ATT_1 conditional on X and is considerably weaker than Assumption (5). Thus, if one were still interested in using B's original voters as counterfactuals despite the methodological difficulties, one could attempt to find the subpopulation of B's old voters who are most similar to the new voters on some set X of observable characteristics and use these as counterfactuals, under the assumption that once the joint distribution of X is equated among new voters and new neighbors, their average potential outcomes would have been identical in the absence of redistricting. But note that Assumption 6 defines a selection on observables assumption which is not guaranteed to hold even under random assignment!

To complicate things further, if Assumption 6 were true this approach would still not necessarily result in unbiased estimates, because the distribution of X between B's old and new voters is not guaranteed to be equal *even if conditional on X both groups of voters would have attained the same average outcomes in the absence of redistricting.*[10] The reason is that the support of the distribution of X among B's new voters may be different from the support of the distribution of X among B's old voters, a concern that becomes all the more relevant given that B's old and new voters were originally in different districts. In sum, the fact that new voters and new neighbors are never in the same population at baseline may imply that both groups are different by construction, and hence that unbiased estimates may

[10]See Heckman, Ichimura, Smith, and Todd (1998) for a formal proof that the lack of common support introduces bias.

not be achieved even if a strong identifying condition is assumed to hold.

Indeed, as mentioned above, the second-best design introduces a lack of common support by construction on covariates that are related to the previous history in the district. For example, new voters may have been moved from a Hispanic to a white incumbent, from a Democratic to a Republican incumbent, from a female to a male incumbent, or from a moderate to an extreme incumbent, while old neighbors by definition would face no such variation in the characteristics of their incumbent (assuming the incumbent runs in both elections). Since different previous histories will likely affect new voters' behavior differently, having balance on these history-related covariates is crucial to identify the causal effect of incumbency. Hence, the second-best design must be modified so that balance on these covariates is achieved.

One possible way of modifying the design is to narrow the set of movements between districts to include only homogeneous changes and hence reduce the imbalance in history-related covariates. For example, one could analyze only voters who are moved from a district represented by a white Democratic incumbent to a district that is also represented by a white Democratic incumbent to eliminate any party and race effects from the observed difference between old and new voters. This is valid strategy, although one obvious disadvantage is that in principle one could keep refining it almost without limit.[11] As can be seen, using new neighbors as counterfactuals poses important methodological challenges.

To summarize, so far we have identified two different designs, the second best design and the best old-neighbors design. The second best design, which compares voters whose district changes to their new neighbors after redistricting, not only requires strong assumptions but also cannot be directly used for estimating the incumbency advantage due to its inherent heterogeneity. In order to reduce this heterogeneity, one must restrict the universe of analysis to districts whose electoral environment was somehow homogeneous *before* redistricting. On

[11]In this case, for example, the movement from one white Democratic incumbent to another could be restricted further to consider only white Democratic incumbents with the same ideology—i.e only moderate Democrats or only extreme Democrats.

the other hand, the best old-neighbors design, which compares voters whose district changes to their old neighbors, requires much weaker assumptions and is directly justified by the redistricting manipulation. This design is not appropriate for estimating the personal vote, although it may be appropriate for other estimands.

2.2 Consecutive redistricting: the best design

We propose a different design. We consider a modification of the thought experiment introduced above, and imagine that after some voters are randomly moved from district A to B (and after election t takes place), another random redistricting plan is implemented right before election $t+1$ so that some voters who were in district A until after election t are randomly chosen and moved to district B. At $t+1$, there are three types of voters in district B: voters who always belonged to B (henceforth *original voters*), voters who became part of district B just before election t (henceforth *early new voters*), and the voters who became part of district B just before election $t+1$ (henceforth *late new voters*). In this case, the most natural way to estimate the causal effect of incumbency is to compare early new voters to late new voters, as not only do they both face the same electoral environment at election $t+1$, but they also have the same electoral environment up to election $t-1$, which implies that their histories are the same except for the fact that early new voters are moved to the new district one election earlier than late new voters. We call this the "best design", as it is free from the complications that arise in the two alternatives considered above.

To formally establish the parameter identified by the best design, let $W_{i,t+1}$ be equal to one if precinct i is moved from district A to district B at election $t+1$, and $W_{i,t+1}$ be equal to zero if precinct i is moved from A to B at election t and remains in B at election $t+1$. In other words, $W_{i,t+1}$ is a new-voter treatment indicator, where new voter is defined as voting in B for the first time at election $t+1$. Letting $Y_0\left(i,t+1\right)$ denote the outcome of i at election $t+1$ if $W_{i,t+1}=0$ and $Y_1\left(i,t+1\right)$ denote the outcome of i at election $t+1$ if

$W_{i,t+1} = 1$, we define the parameter of interest (ATT_B, where B refers to "best design") as

$$ATT_B \equiv E\left[Y_1(i, t+1) - Y_0(i, t+1) \mid W_{i,t+1} = 1\right] \tag{7}$$

which is identified under

$$E\left[Y_0(i, t+1) \mid W_{i,t+1} = 1\right] = E\left[Y_0(i, t+1) \mid W_{i,t+1} = 0\right] \tag{8}$$

In words, ATT_B is identified if late new voters and early new voters would have attained the same average outcomes if they both had been in the new district for exactly two elections. Below, we will show that randomization under this design together with an assumption of stationarity guarantees that Assumption (8) holds.

Since we assumed that both groups of voters are in the same district at election $t-1$, and that just before election t the set of voters for which $W_{i,t+1} = 1$ is randomly chosen and moved to district B, we have

$$E\left[Y_0(i, t-1) \mid W_{i,t+1} = 1\right] = E\left[Y_0(i, t-1) \mid W_{i,t+1} = 0\right] \tag{9}$$

This is, randomization guarantees that both groups of voters have the same pre-treatment average outcomes. But Assumption (9) does not imply Assumption (8), hence we need to add an assumption to the best design in order to obtain exchangeability at election $t+1$. We make the following additional assumption:

$$E\left[Y_0(i, t+1) - Y_0(i, t-1) \mid W_{i,t+1} = 1\right] = E\left[Y_0(i, t+1) - Y_0(i, t-1) \mid W_{i,t+1} = 0\right] \tag{10}$$

Assumption (10) together with Assumption (9) imply Assumption (8). In other words, if late new voters are randomly chosen *and* early new voters and late new voters would have followed the same path between election $t-1$ and election $t+1$ if they both had spent election t and election $t+1$ in the new district, ATT_B is identified.[12] As before, since in practice district boundaries are not randomly modified, in order to achieve identification of the parameters of interest in best design we must make the assumption that, *conditional on certain observable characteristics*, late new voters are exchangeable with early new voters. This is undoubtedly a strong assumption, but is plausible considering that we use the same data which participants in the redistricting battles fed into their computer programs to design their various redistricting plans.

Furthermore, the best design proposed here allows us to implement a crucial placebo experiment to test the validity of the identification strategy, because we observe the behavior of precincts which will be redistricted before they are redistricted. As such, the placebo test examines precincts which will be redistricted (or not) at election $t+1$ but which are in the same district in elections t, $t-1$, $t-2$, etc. Calling those to be redistricted in election $t+1$ "treated" and those who will not be redistricted in election $t+1$ "controls", we can arbitrarily denote $t-1$ to be the baseline year, and our placebo test is that in t there should be no significant difference between the outcomes of our treated and control groups, once we condition on an appropriate set of observable characteristics. In Section 6, we show that past presidential vote, which is the sole conditioning variable used by Ansolabehere et al. (2000) to satisfy selection on observables, is not sufficient to satisfy this placebo test. But a rich set of covariates which includes votes in past state-wide and House elections as well as past registration and turnout does satisfy the placebo test.

Before presenting the estimated effects of incumbency on electoral outcomes in California and Texas using the different designs described above, we discuss the interpretation of the redistricting estimand.

[12] For example, Assumption (10) would rule out a situation in which early new voters become more motivated after election $t-1$ and late new voters become more disengaged after election $t-1$. In this case, even if late new voters were moved to the new district at election t instead of at election $t+1$, we would still observe a difference between, say, the turnout rates of both groups.

3 The Redistricting Estimand: Personal Vote or Learning?

Ansolabehere et al. (2000) proposed to use redistricting as an identification strategy for the *personal vote*, this is, for the electoral advantage that an incumbent acquires by providing casework, bringing federal resources to the district, taking positions that match voters' tastes, etc. The authors distinguished the personal incumbency advantage from the incumbency advantage that stems from candidate quality and from incumbency as a cue that comes to replace weakening party ties.

Although it is clear that using redistricting as a research design to identify the effect of incumbency does not speak to the last two sources of incumbency advantage (since old and new voters face both the same candidates and the same cues), it is less clear that this identification strategy does capture all aspects of the personal vote. The comparison between new voters and old voters in a given district will only capture those elements of the personal vote that stem from the personal relationship that the incumbent has established with her constituents over the years, but will miss the electoral advantage that stems from the resources associated with being an incumbent, since these resources can, at least in principle, be targeted to both new and old voters. For example, the incumbent can exploit her "reputation" only among old voters, but she can deploy resources such as franking privileges, campaign funds, etc., to both old voters and new voters alike. Hence, by construction, this design cannot provide information about how the incumbent's vote share is affected by this type of resources.

It follows that if old voters and new voters were found to vote for the incumbent at the same rate, this should not be interpreted as evidence that there is no personal incumbency advantage, but rather as evidence that the incumbent's history with her constituents does not translate into a significant electoral advantage. This would be expected if information about the incumbent's past record of constituency service was available to new voters. Indeed, in a world where information about the incumbent were disseminated instantaneously and without frictions, there would be little reason for new voters and old voters to vote differently.

Thus, we offer a more precise interpretation of what exactly is being estimated when redistricting is used as an identification strategy for the incumbency advantage. We argue that this research design estimates how quickly new voters learn about the *type* of their new incumbent, this is, about how good of a job their new incumbent does at bringing pork, providing casework, and all of the other components of what is often called the personal vote. Under this interpretation, a zero effect would mean "instantaneous learning", and a non-zero effect would mean that new voters need more than one election cycle to gain full knowledge of their incumbent.

4 Empirical Application: California and Texas

We implement the best and second best designs in Texas and the second best design in California. In both cases, we analyze congressional elections between 1998 and 2006. In order to implement the best design, we take advantage of the fact that congressional districts in Texas were redrawn after the reapportionment that followed the 2000 census, and they were redrawn again before the 2004 elections in a highly controversial mid-decade plan that was engineered by former Republican House Majority Leader Tom DeLay (Bickerstaff 2007).[13]

These two consecutive congressional redistricting plans implemented in Texas in 2002 and 2004 give us the unique opportunity of implementing the best design to estimate the incumbency effect. We define *late new voters* as voters who were in a given district in the 2000 and the 2002 elections and in a different district in the 2004 election, and *early new voters* as voters who were in the same district as late new voters in the 2000 election but in the 2002 and 2004 elections were in the district to which late new voters are moved in 2004.[14] As mentioned above, this guarantees that both types of voters face the same electoral environment in the 2000 and the 2004 elections and hence is the most natural design to use

[13] See Appendix B for a detailed description of Texas redistricting plans.

[14] This is, if we call the original district "A" and the new district "B", in the 2000 election both early and late new voters are in A, in the 2002 election early new voters are in B and late new voters are still in A, and in the 2004 election both early and late new voters are in B.

redistricting as an identification strategy for the incumbency effect.

For Texas, data on electoral returns were collected from the Texas Legislative Council (TXLC) at the Voting Tabulation District (VTD) level.[15] VTDs are census blocks grouped to approximate voting precincts as closely as possible, providing a link between census data and electoral data.[16] Since there is a one-to-one mapping between VTDs and 2000 census blocks, we are able to track the electoral returns of the same geographical unit over time. Election returns reported by the TXLC include not only congressional elections, but other statewide and national elections such as state house, state senate, U.S. Senate, and presidential elections. Data files also include total voter registration, Hispanic voter registration estimated by surname match, voter turnout, and candidate information including the candidates' names and party affiliation, identification of black and Hispanic candidates, and identification of incumbency status.

For California, data on electoral returns were collected from the Statewide Database (SWDB) at the 2000 census block level.[17] As in Texas, by using 2000 census blocks as the unit of analysis we are able to track the electoral returns of the same geographical unit over time. Electoral returns include congressional, state house, state senate, U.S. senate, presidential, and other elections. The data also include registration and turnout for different age groups and party affiliations. The roster of congressional candidate and incumbents was obtained directly from the California Secretary of State, and data on race and ethnicity were obtained from the *Hispanic Americans in Congress* website, maintained by the Library of Congress, and the CRS Report for Congress 2008.

We added DW-Nominate scores and data on challengers' quality[18] to both the California

[15] Data were obtained from the TXLC's ftp website (`ftp://ftpgis1.tlc.state.tx.us`, with updates as of May 02, 2007, and from special requests from the TXLC.

[16] for details about how VTDs are constructed, data sources, and other issues regarding data construction, see Texas Legislative Council (2000, 2001).

[17] Data for 1998 and 2000 were directly obtained at the block level, while data for 2002 through 2006 were obtained at the precinct level and converted to 2000 census block level using convertion files provided by the Statewide Database.

[18] DW-Nominate scores were obtained from the Voteview Website (`http://voteview.com/dwnomin.htm`, with updates as of April 10, 2007), and challenger quality data were kindly provided by Gary C. Jacobson.

and Texas datasets. We also merged data from the 2000 census at the VTD-level for Texas and at the census block level for California. For Texas, census data from Summary File 1 was easily obtained at the VTD level by aggregating census blocks up to the VTD level; for California, we merged block-level data directly since our unit of analysis is the census block. Census data from Summary File 3 was converted to the VTD-level for Texas and to the block-level for California, although the assignment of Summary File 3 variables to blocks and VTDs is only approximate given that the smallest geographical unit for which Summary File 3 variables are reported is the block-group level, and there is no unique mapping between block-groups and VTDs. Summary File 1 variables include total population, population by age, white population, black population and Hispanic population. Summary File 3 variables include population by language spoken at home, population by employment status, population by place of birth, and population by highest education level achieved.

Every VTD and block in each final dataset was assigned to the congressional district it belonged to in each general election between 1998 and 2006, according to the congressional district plan that was effective at the time of each election in each state. Our final Texas dataset contains $8,040$ VTDs and our final dataset for California contains $284,040$ census blocks. We restrict these samples further in our analysis by keeping only VTDs and blocks which belong to closed seats in 2002, so that there is an incumbent running in 2002.[19]

5 Genetic Matching

We estimate the effects of incumbency using Genetic Matching (GenMatch), a nonparametric matching method proposed by Sekhon (forthcoming,2006), Sekhon and Grieve (2007) and Diamond and Sekhon (2005), which algorithmically maximizes the balance of observed covariates between treated and control groups. GenMatch is a generalization of propensity score and Mahalanobis distance matching, and it has been used by a variety of researchers

[19] See Appendix A for details.

(e.g., Brady and Hui 2006; Gilligan and Sergenti 2006; Gordon and Huber 2007; Herron and Wand forthcoming; Morgan and Harding 2006; Lenz and Ladd 2006; Park 2006; Raessler and Rubin 2005). The method uses a genetic algorithm (Mebane and Sekhon 1998; Sekhon and Mebane 1998) to optimize the balance of observed covariates as much as possible given the data, and does not depend on knowing or estimating the propensity score (though the method is improved when a propensity score is incorporated).

The idea underlying the GenMatch algorithm is that if neither the propensity score nor Mahalanobis distance is optimal for achieving balance in a given dataset, one should be able to search over the space of distance metrics and find something better. One way of generalizing the Mahalanobis metric is to include an additional weight matrix:

$$d(X_i, X_j) = \left\{ (X_i - X_j)' \left(S^{-1/2}\right)' W S^{-1/2} (X_i - X_j) \right\}^{\frac{1}{2}}$$

where W is a $k \times k$ positive definite weight matrix and $S^{1/2}$ is the Cholesky decomposition of S which is the variance-covariance matrix of X.[20]

GenMatch is an affinely invariant matching algorithm that uses the distance measure $d()$, in which all elements of W are zero except down the main diagonal, which consists of k parameters that must be chosen. Note that if each of these k parameters are set equal to 1, $d()$ is the same as Mahalanobis distance.[21]

This leaves the problem of how to choose the free elements of W. Many loss criteria recommend themselves. By default, cumulative probability distribution functions of a variety of standardized statistics are used as balance metrics and are optimized without limit. The default standardized statistics are paired t-tests and nonparametric KS tests. Sekhon (2006) shows that this loss functions work well in practice.

These statistics are not used to conduct formal hypothesis tests, because no measure of

[20]The Cholesky decomposition is parameterized such that $S = LL'$, $S^{1/2} = L$. In other words, L is a lower triangular matrix with positive diagonal elements.

[21]The choice of setting the non-diagonal elements of W to zero is made for reasons of computational power alone. The optimization problem grows exponentially with the number of free parameters. It is important that the problem be parameterized so as to limit the number of parameters which must be estimated.

balance is a monotonic function of bias in the estimand of interest and because we wish to maximize balance without limit. Descriptive measures of discrepancy generally ignore key information related to bias which is captured by probability distribution functions of standardized test statistics.[22] And these metrics, unlike those based on optimized distribution functions, perform poorly in a series of Monte Carlo sampling experiments just as one would expect given their properties. For details see Sekhon (2006).

By default, GenMatch attempts to minimize a measure of the maximum observed discrepancy between the matched treated and control covariates at every iteration of optimization. For a given set of matches resulting from a given W, the loss is defined as the minimum p-value observed across a series of standardized statistics. Conceptually, the algorithm attempts to minimize the largest observed covariate discrepancy at every step, which is accomplished by maximizing the smallest p-value at each step.[23] Because GenMatch is minimizing the maximum discrepancy observed at each step, it is minimizing the infinity norm. This property holds even when, because of the distribution of X, the Equal Percent Bias Reduction (EPBR) property (Rubin 1976a,b; Rubin and Thomas 1992) does not hold. GenMatch is able to retain good properties even when EPBR does not hold because a set of constraints can be imposed by the loss function optimized by the genetic algorithm.

The optimization problem described above is difficult and irregular, and the genetic algorithm developed by Mebane and Sekhon (1998) is used to conduct the optimization. Details of the algorithm are provided in Sekhon and Mebane (1998).

GenMatch has been shown to have better properties than the usual alternative matching methods both when the EPBR property holds and when it does not (Sekhon 2006; Diamond and Sekhon 2005). Although GenMatch can be combined with various matching methods such as optimal matching or greedy matching, in this paper we use 1-to-1 matching with

[22]For example, using several descriptive metrics, one is unable to recover reliably the experimental benchmark in a testbed dataset for matching estimators (Dehejia and Wahba 1999)

[23]More precisely lexical optimization will be done: all of the balance statistics will be sorted from the most discrepant to the least and weights will be picked which minimize the maximum discrepancy. If multiple sets of weights result in the same maximum discrepancy, then the second largest discrepancy is examined to choose the best weights. The processes continues iteratively until ties are broken.

replacement.

6 Results

6.1 Placebo Test

As discussed in our research design section, because redistricting involves the the nonrandom assignment of blocks and VTDs to Congressional districts, a selection on observables assumption must be made in order to make progress. Fortunately, a placebo test is available to check this assumption.

Our placebo test uses data from Texas, where the multiple redistricting allow us to implement the best design. The test examines VTDs which will be redistricted (or not) in 2004 but which are in the same district in 1998, 2000 and 2002. We assume that those to be redistricted for the 2004 election are treatment and those who will remain are control, and arbitrarily denote 2000 to be the baseline year. As explained in Section 2, our placebo test is that in 2002 there should be no significant difference between our treated and control groups in both vote intention and turnout.

Our dataset allows us to draw on a rich set of covariates based on electoral returns, registration files, and census data. Table 1 provides the covariates we use to perform the matching for our placebo test. Like past work, we use past presidential vote returns, but we also use data from statewide offices, registration figures, past turnout numbers, and the past vote for the Democratic Party's House candidate. Note that because both the treated and control units in this placebo test are drawn from the same Congressional district (as they are in our "best design"), we by definition match on the party of the incumbent, the historical quality of challengers, and other aspects of past races at the local, statewide and national level as experienced by the VTDs we are matching.

The variables listed in Table 1 were chosen on a priori theoretical grounds because we believed them to be theoretically important. This is the set of variables that we used in the first specification of the placebo test, and no further modification of the set was necessary to pass the test. For completeness, we then also matched on other variables drawn from the census (such as the percentage of the voting eligible population which is Hispanic, black, white, native born, and who speaks English), but these extra covariates were not necessary to reliably pass the placebo test.

The balance statistics in Table 1 show the excellent balance that Genetic Matching was able to find post matching. The mean differences between treatment and control groups, the maximum differences in the empirical QQ-plots, and the significance of the differences greatly shrank post matching in every case. The smallest bootstrap Kolmogrov-Smirnov p-value post matching is 0.235 and the second smallest is 0.601 while all of the pre-matching p-values are significant at 0.00.

Table 2 presents results for both incumbent vote and turnout, and shows that the estimates for the matched data are all statistically indistinguishable from zero. Post Genetic Matching point estimates and confidence intervals are from Hodges-Lehmann interval estimation.[24] The substantive results are unchanged if either bivariate overdispersed GLM estimation[25] or if Abadie-Imbens standard errors (Abadie and Imbens 2006) are used instead.

Figure 2 plots the QQ-plot for the incumbent vote in 2002 between treatment and control groups. The figure visually presents the results in Table 2. It is clear that the result for incumbent vote is zero, as it should be in this placebo test.

Past presidential vote, which is the sole conditioning variable used in previous work to

[24]Rosenbaum (2002) provides details, and (Hill and Reiter 2006) provide a simulation study comparing the performance of Hodges-Lehmann intervals relative to other methods of interval estimation for treatment effects using matching.

[25]This quasi-likelihood model defines the observed votes of a given party in a given VTD as the number of positive responses obtained out of the total number of trials, where the latter is defined as the total number of votes in the VTD. Typically, votes within a VTD will be clustered and hence the variance of the vote counts will be larger than the variance implied by the binomial model. This phenomenon is known as overdispersion, and is common with aggregate vote data because VTD-level variables fail to capture traits that affect voters' choices and vary across voters within each VTD. Overdispersed models correct the variance-covariance matrix to account for the extra variability in the data. See McCullagh and Nelder (1989) for a general treatment of overdispersed models for binary data, and Mebane and Sekhon (2004) for a discussion of overdispersion in aggregate election data.

satisfy the selection on observables assumption, is not sufficient to satisfy this placebo test. We have conducted placebo tests using means and medians of a number of past presidential elections. We present detailed results for using the 2000 presidential vote as the estimate of the normal vote. Figure 3(a) presents the balance on Presidential vote in 2000 matching on only this variable. As can be seen, balance is excellent. Figure 3(b) presents the QQ-plot for the estimand in question, House vote in 2002. Unlike the case for our rich set of covariates, there is a significant treatment effect. Table 3 shows the results for the formal placebo tests.

Notwithstanding the excellent balance displayed in Figure 3, it may be argued that balance is not good enough on past Presidential vote. The QQ-plot for Presidential vote in Figure 3 corresponds to a mean difference of only 0.00165, a maximum difference of 0.021 and a bootstrap KS p-value of 0.958. However, if a tight caliper is used, even better balance can be obtained. Using this caliper, balance on presidential vote improves to a mean difference of 0.000156, a maximum difference of 0.000362 and bootstrapped KS p-value of 1 (results not shown). But as the Data 2 results in Table 3 show, even when balance of Presidential vote is improved this much, balance on the House vote proportion in 2002 is still not good, as there is still a significant effect for incumbent vote in 2002 (p-value of 0.00001).

6.2 Incumbency Advantage When Party Remains The Same

Table 4 presents results from the standard Ansolabehere et al. (2000) "Old Voters, New voter" design for the treatment effect of the incumbent changing, but the party remaining the same. This is the design which fails the placebo test above because it only conditions on past presidential vote. Using this standard design, there is a 5.8% incumbency advantage with a p-value of 0.000. This point estimate and the confidence interval of 4.7% to 7.0% is similar to the average national estimate provided by Ansolabehere et al. (2000) in the modern era. And it is somewhat larger than Desposato and Petrocik's (2003) estimate from California. The standard design applied to Texas estimates a positive and highly significant incumbency advantage of the same magnitude as found in the existing literature.

Table 5 presents the results from our "best-design" for the treatment effect of the incumbent changing, but the party remaining the same. As we discussed before, this estimates how quickly new voters are able to learn about the type of their new incumbent.

The effect is estimated to be statistically indistinguishable from zero for both 2004 and 2006. And incumbency appears to have no effect on turnout. Note that our point estimates are also extremely small. For example, for 2004, the estimated vote proportion is 0.00637 and for 2006 it is 0.00843.

Table 6 presents our incumbency advantage estimates from the "second-best" design using the same covariates we used in our "best-design", with the addition of variables which attempt to measure details of the House election at baseline in 2000 and in 1998. In particularly, both of Jacobson's challenger quality measures in 2000 and in 1998 are added, and the party of the incumbent is held constant. As we extensively discussed in the research design section, this design is not as compelling as the previous one. Nonetheless, it allows us to use much more data (434 observations to estimate a single parameter).

As is made clear in Table 6, all of our estimates of the effect of incumbency are extremely small and all are *substantively* and *statistically* indistinguishable from zero. The largest absolute value of the point estimate is 0.009 (turnout in '06). This is of course insignificant, but even if it were significant, it would not be a substantively meaningful effect. As discussed in Section 3, we do not interpret these results as evidence that there is no such a thing as a personal incumbency advantage. Rather, we interpret them as evidence that when the party of the incumbent remains the same before and after redistricting, voters are able to learn very quickly about their new incumbent's record.

6.3 Incumbency Advantage When Party Changes

Table 7 presents the results for the incumbency advantage using our preferred design, but when the party of the incumbent changes. Here we find that there is a significant and large incumbency effect. Early new-voter VTDs vote for the incumbent party at a much higher rate than late new-voter VTDs: 12.6%. But by the time that late new-voter VTDs have been in the Congressional district for a term, this effect drops to about 3.9%. Voter turnout is not significantly effected by this incumbency effect.

7 Discussion

The use of so called natural experiments to estimate causal effects has recently become popular in economics and other social sciences. Although natural experiments offer significant advantages, they do not possess key benefits of actual experiments and hence require careful theoretical and statistical work to make valid inferences. First, there exists the obvious problem that natural experiments do not have random assignment with a known probability distribution and that selection on observables is a strong assumption which is difficult to justify. As in the current example, rarely can natural experiments be used without significant covariate adjustment.

A less often noted but crucial way in which natural experiments and actual experiments differ is that with the latter researchers a priori design the study so that randomization will ensure the identification of the causal effect of interest. As we have seen, the ideal experiment which one would construct to estimate the personal vote is different from the experiment which one would construct to estimate the effect of candidate ethnicity. And the differences are not simply limited to the obviously different treatments in the two cases. The different ideal experiments imply different identifying assumptions and hence different experimental designs ("best design" vs the "best old-neighbors design").

Of course, a benefit of natural experiments, which should not be underestimated, is that with such studies there is a (non-random) manipulation and as such there is hope that a conditional exchangeability assumption can be satisfied. It is easier to determine what is post- and what is pre-treatment than with the observational designs more commonly used, certainly easier than with cross-section data which lacks any manipulation such as the usual public opinion dataset.

We find that when the party of the incumbent remains the same being a new voter has no effect on the incumbent vote share, even though the standard "old voters, new voters" design estimates a highly significant positive effect of about 5.8% in our data. Since the standard design uses the incorrect potential outcomes and because it fails our placebo test, our results show that there is a methodological bias in existing published estimates of the incumbency advantage (e.g., Ansolabehere et al. 2000; Carson et al. 2007; Desposato and Petrocik 2003).

Our results are consistent with theoretical arguments that existing positive estimates of incumbency advantage are plagued by selection problems (Cox and Katz 2002; Zaller 1998). And our finding of a significant effect when the party of the incumbent changes is consistent with the results of Lee (forthcoming) who finds a significant party incumbency effect by the use of a regression discontinuity design. Voters appear to have a preference for remaining with the same party they had before.

New voters quickly learn the type of their new incumbent when voters are moved from one incumbent to another, and the party of the incumbent does not change. That is, voters quickly learn how good their new incumbent is at, for example, providing constituency services. But when the party label of the incumbent does change, new voters are less likely to support their new incumbent than old voters possibly because they underestimate the constituency benefits provided by the new incumbent. How well incumbents perform constituency service is almost certainly of the utmost importance. But new voters learn enough about the quality of this service so that they vote like old voters when the party label does not change. These results are consistent with findings that voters rely on partisan cues to help them learn about politicians and that partisan cues influence how quickly voters learn.

In the literature, finding a positive incumbency effect is much less common in other countries than the United States. Some work even documents a negative incumbency effect in legislative and executive offices in developing countries (Linden 2004; Uppal 2005). Our finding of a positive effect when the party of the incumbent remains the same suggests that partisanship plays an important part in the incumbency advantage—indeed the key part given that we do not find an effect when the party remains the same. And as has long been noted, the U.S. is exceptional in the important role which party identification plays in voting behavior (e.g., Budge, Crewe, and Farlie 1976).

In the future we plan to use our research design to examine how voters respond to other characteristics of incumbents such as their voting records or ideal points. The design can also be used to estimate how incumbents adapt to changes in their constituents. For example, do incumbents adjust their roll call voting behavior?

A Final sample sizes

Texas' territory is divided into 8,634 2004-VTDs and 675,062 2000-census blocks. Since even unpopulated areas were assigned a census block in 2000, 207 of these 8,634 VTDs have zero population and hence zero election returns. Of the remaining VTDs, some need to be discarded due to the phenomenon of "multiple congressional districting" which occurs when a VTD reports election returns for more than one congressional district in a given election. We exclude from the analysis all VTDs for which multiple redistricting occurs once or more in the entire period under analysis. This guarantees that all the VTDs that we keep correspond to areas that unequivocally belonged to a single congressional district in every congressional election between 2000 and 2004. There are 8,040 VTDs with positive population that satisfy this condition, which is our final sample size for Texas.

California's territory is divided into 533,163 2000-census blocks, of which 344,356 have positive population. Since 46,843 of these blocks have population of less than 10, many blocks have either zero registration or zero votes cast in some or all of the years under analysis. We restrict our sample to those blocks which have both positive registration and positive number of votes cast in the congressional election for all the years between 1998 and

2006. The final sample size is $284,040$ blocks.

B Texas and California redistricting plans in the 2000s

Texas implemented six different congressional district plans between 1990 and 2006.[26] After the reapportionment that followed the 1990 census, the districts enabled by the old $C001$ plan were redrawn. The 1992 elections were held under the new districts enacted by plan $C657$, which remained in effect until the 1996 primaries. In August 1996, 13 of Texas' 30 congressional districts were redrawn. The new plan, $C746$, was used in the 1996 general election and it remained in effect during the 1998 and 2000 elections.

In 2001, after the reapportionment following the 2000 census that created two new congressional seats, the Texas Legislature was in charge of redrawing the senate, house, congressional, and State Board of Education districts during the regular session of the 77th Legislature. But the plans failed to be considered by the full Senate and the full House, and the legislature adjourned without enacting new districts. A number of congressional proposals were submitted to state and federal courts. Finally, on November 14, 2001, the U.S. District Court issued an order adopting new congressional districts (Plan $C1151$) for the 2002 elections.

But plan $1151C$ was only in effect for the 2002 elections. In 2003, Republican majority leader Tom Delay led an effort to enact a new congressional district plan, with the objective of maximizing the number of Texas' Republicans elected to Congress in the 2004 and subsequent elections. After a legislative battle that included Democratic lawmakers massively fleeing to New Mexico and Oklahoma to avoid quorum, the new plan (Plan 1374) was passed in October, 2003. The 2004 primaries and general election were held under this new plan. Congressional districts were redrawn one more time in 2006.

[26] See the Texas Legislative Council's Redistricting website `http://www.tlc.state.tx.us/redist`,Texas Legislative Council (2000), and Texas Legislative Council (2001) for details about Texas' redistricting plans during the 1990s and 2000s.

In California, there was only one redistricting plan implemented in the 2000s. The districts in effect during the 1990s were redrawn by the 2001 redistricting plan, which was enacted in two separate bills in September 2001. Bills AB632 established Senate and Congressional districts, and bill SB802 established Assembly and Board of Equalization districts.

References

Abadie, Alberto and Guido Imbens. 2006. "Large Sample Properties of Matching Estimators for Average Treatment Effects." *Econometrica* 74: 235–267.

Alford, John R. and David W. Brady. 1989. "Personal and Partisan Advantage in US Congressional Elections." In Lawrence C. Dodd and Bruce I. Oppenheimer, editors, *Congress Reconsidered 4th edition* Washington, D.C.: CQ Press.

Ansolabehere, Stephen, David W. Brady, and Morris P. Fiorina. 1988. "The Vanishing Marginals and Electoral Responsiveness." *British Journal of Political Science* 22 (1): 21–38.

Ansolabehere, Stephen and James Snyder. 2002. "The Incumbency Advantage in U.S. Elections: An Analysis of State and Federal Offices, 1942-2000." *Election Law Journal* 1 (3): 315–338.

Ansolabehere, Stephen, James M. Snyder, and Charles Stewart. 2000. "Old Voters, New Voters, and the Personal Vote: Using Redistricting to Measure the Incumbency Advantage." *American Journal of Political Science* 44 (1): 17–34.

Ashworth, Scott and Ethan Bueno de Mesquita. 2007. "Electoral Selection, Strategic Challenger Entry, and the Incumbency Advantage." Working Paper.

Bickerstaff, Steve. 2007. *Lines in the Sand*. Austin, Texas: University of Texas Press.

Born, Richard. 1979. “Generational Replacement and the Growth of Incumbent Reelection Margins in the US House.” *American Political Science Review* 73 (3): 811–817.

Brady, Henry and Iris Hui. 2006. “Is it Worth Going the Extra Mile to Improve Causal Inference?”.” Paper presented at the 23rd Annual Summer Meeting of the Society of Political Methodology.

Breaux, David. 1990. “Specifying the Impact of Incumbency on State Legislative Elections.” *American Politics Quarterly* 18 (3): 270–286.

Budge, Ian, Ivor Crewe, and Dennis J. Farlie. 1976. *Party Identification and Beyond*. New York: John Wiley and Sons Ltd.

Burnham, Walter D. 1974. “Communications.” *American Political Science Review* 68 (1): 207–213.

Carson, Jamie L., Erik J. Engstrom, and Jason M. Roberts. 2007. “Candidate Quality, the Personal Vote, and the Incumbency Advantage in Congress.” *American Political Science Review* 101 (2): 289–301.

Cox, Gary W. and Jonathan N. Katz. 1996. “Why Did the Incumbency Advantage in U.S. House Elections Grow.” *American Journal of Political Science* 40 (2): 478–497.

Cox, Gary W. and Jonathan N. Katz. 2002. *Elbridge Gerry’s Salamander: The Electoral Consequences of the Reapportionment Revolution*. New York: Cambridge University Press.

Cox, Gary W. and Scott Morgenstern. 1993. “The Increasing Advantage of Incumbency in the U.S. States.” *Legislative Studies Quarterly* 18 (4): 495–514.

Dehejia, Rajeev and Sadek Wahba. 1999. “Causal Effects in Non-Experimental Studies: Re-Evaluating the Evaluation of Training Programs.” *Journal of the American Statistical Association* 94 (448): 1053–1062.

Desposato, Scott W. and John R. Petrocik. 2003. "The Variable Incumbency Advantage: New Voters, Redistricting, and the Personal Vote." *American Journal of Political Science* 47 (1): 18–32.

Diamond, Alexis and Jasjeet S. Sekhon. 2005. "Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies." Working Paper.

Erikson, Robert S. 1971. "The advantage of Incumbency in Congressional Elections." *Polity* 3 (3): 395–405.

Erikson, Robert S. 1972. "Malapportionment, Gerrymandering, and Party Fortunes in Congressional Elections." *American Political Science Review* 65 (4): 1234–1245.

Fenno, Richard F. Jr. 1978. *Home Style: House Members in Their Districts*. Boston: Little, Brown and Company.

Ferejohn, John A. 1977. "On the Decline of Competition in Congressional Elections." *American Political Science Review* 71 (1): 166–176.

Fiorina, Morris P. 1977. "The Case of the Vanishing Marginals: The Bureaucracy Did It." *American Political Science Review* 71 (1): 177–181.

Fiorina, Morris P. 1989. *Congress: Keystone of the Washington Establishment*. New Haven: Yale University Press.

Gelman, Andrew and Gary King. 1990. "Estimating Incumbency Advantage without Bias." *American Journal of Political Science* 34 (4): 1142–1164.

Gilligan, Michael J. and Ernest J. Sergenti. 2006. "Evaluating UN Peacekeeping with Matching to Improve Causal Inference." `http://sekhon.berkeley.edu/causalinf/papers/gilligan_sergenti_06.pdf`.

Gordon, Sandy and Greg Huber. 2007. "The Effect of Electoral Competitiveness on Incumbent Behavior." *Quarterly Journal of Political Science* 2: 107–138.

Heckman, James J., Hidehiko Ichimura, Jeffrey Smith, and Petra Todd. 1998. "Characterizing Selection Bias Using Experimental Data." *Econometrica* 66 (5): 1017–1098.

Heckman, James J., Hidehiko Ichimura, and Petra Todd. 1997. "Matching as an Econometric Evaluation Estimator." *Review of Economic Studies* 65 (2): 261–294.

Herron, Michael C. and Jonathan Wand. forthcoming. "Assessing Partisan Bias in Voting Technology: The Case of the 2004 New Hampshire Recount." *Electoral Studies*.

Hill, Jennifer and Jerome P. Reiter. 2006. "Interval Estimation for Treatment Effects Using Propensity Score Matching." *Statistics in Medicine* 25: 2230–2256.

Jacobson, Gary C. 1987. "The Marginals Never Vanished: Incumbency and Competition in Elections to the U.S. House of Representatives." *American Journal of Political Science* 31 (1): 126–141.

Jacobson, Gary C. and Samuel Kernell. 1983. *Strategy and Choice in Congressional Elections*. New Haven: Yale University Press.

Jewell, Malcolm E. and David Breaux. 1988. "The Effect of Incumbency on State Legislative Elections." *Legislative Studies Quarterly* 13 (4): 495–514.

King, Gary. 1991. "Constituency Service and Incumbency Advantage." *British Journal of Political Science* 21 (1): 19–28.

King, Gary and Andrew Gelman. 1991. "Systemic Consequences of Incumbency Advantage in U.S. House Elections." *American Journal of Political Science* 35 (1): 110–138.

Krashinsky, Michael and William J. Milne. 1993. "The Effects of Incumbency in U. S. Congressional Elections, 1950-1988." *Legislative Studies Quarterly* 18 (3): 321–344.

Krehbiel, Keith and John R. Wright. 1983. "The Incumbency Effect in Congressional Elections: A Test of Two Explanations." *American Journal of Political Science* 27 (1): 140–157.

Lee, David S. forthcoming. "Randomized Experiments from Non-random Selection in U.S. House Elections." *Journal of Econometrics*.

Lenz, Gabriel S. and Jonathan McDonald Ladd. 2006. "Exploiting a Rare Shift in Communication Flows: Media Effects in the 1997 British Election." `http://sekhon.berkeley.edu/causalinf/papers/LaddLenzBritish.pdf`.

Levitt, Steven D. and Catherine D. Wolfram. 1997. "Decomposing the Sources of Incumbency Advantage in the U.S. House." *Legislative Studies Quarterly* 22 (1): 45–60.

Linden, L. 2004. "Are incumbents really advantaged? The Preference for Non-Incumbents in Indian National Elections." Working Paper.

Mayhew, David R. 1974. "Congressional Elections: The Case of the Vanishing Marginals." *Polity* 6 (3): 295–317.

McCullagh, Peter and John A. Nelder. 1989. *Generalized Linear Models*. New York: Chapman & Hall.

Mebane, Walter R. Jr. and Jasjeet S. Sekhon. 1998. "GENetic Optimization Using Derivatives (GENOUD)." Software Package. `http://sekhon.berkeley.edu/rgenoud/`.

Mebane, Walter R. Jr. and Jasjeet S. Sekhon. 2004. "Robust Estimation and Outlier Detection for Overdispersed Multinomial Models of Count Data." *American Journal of Political Science* 48 (2): 391–410.

Morgan, Stephen L. and David J. Harding. 2006. "Matching Estimators of Causal Effects: Prospects and Pitfalls in Theory and Practice." *Sociological Methods & Research* 35 (1): 3–60.

Nelson, Candice J. 1979. "The Effect of Incumbency on Voting in Congressional Elections." *Political Science Quarterly* 93: 665–678.

Park, Johng Hee. 2006. "Causal Effect of Information on Voting Behavior from a Natural Experiment: An Analysis of Candidate Blacklisting Campaign in 2000 South Korean National Assembly Election." Working paper.

Payne, James L. 1980. "The Personal Electoral Advantage of House Incumbents, 1936-1976." *American Politics Research* 8 (4): 465–482.

Raessler, S. and D. B. Rubin. 2005. "Complications when using nonrandomized job training data to draw causal inferences." *Proceedings of the International Statistical Institute*.

Rosenbaum, Paul R. 2002. *Observational Studies*. New York: Springer-Verlag 2nd edition.

Rubin, Donald B. 1976a. "Multivariate Matching Methods That are Equal Percent Bias Reducing, I: Some Examples." *Biometrics* 32 (1): 109–120.

Rubin, Donald B. 1976b. "Multivariate Matching Methods That are Equal Percent Bias Reducing, II: Maximums on Bias Reduction for Fixed Sample Sizes." *Biometrics* 32 (1): 121–132.

Rubin, Donald B. and Neal Thomas. 1992. "Affinely Invariant Matching Methods with Ellipsoidal Distributions." *Annals of Statistics* 20 (2): 1079–1093.

Sekhon, Jasjeet S. 2006. "Alternative Balance Metrics for Bias Reduction in Matching Methods for Causal Inference." Working Paper.
URL `http://sekhon.berkeley.edu/papers/SekhonBalanceMetrics.pdf`

Sekhon, Jasjeet S. forthcoming. "Multivariate and Propensity Score Matching Software with Automated Balance Optimization: The Matching package for R." *Journal of Statistical Software*.

Sekhon, Jasjeet S. and Richard Grieve. 2007. "Bias Adjustment in the Economic Evaluation of Observational Data: the use of a non-parametric approach (Genetic Matching)." Paper presented at the International Health Economics Conference, Copenhagen, July 2007.

Sekhon, Jasjeet Singh and Walter R. Mebane, Jr. 1998. "Genetic Optimization Using Derivatives: Theory and Application to Nonlinear Models." *Political Analysis* 7: 189–203.

Texas Legislative Council, Research Division. 2000. *Guide to 2001 Redistricting*. Austin, Texas: Texas Legislative Council.

Texas Legislative Council, Research Division. 2001. *Data for 2001 Redistricting in Texas*. Austin, Texas: Texas Legislative Council.

Uppal, Yogesh. 2005. "The (Dis)advantaged Incumbents: Estimating Incumbency Effects in Indian State Legislatures." *Center for the Study of Democracy*. Symposium: Democracy and Its Development. Paper G05-06.

Word, David L. and R. Colby Perkins. 1996. "Building a Spanish Surname List for the 1990's: A New Approach to an Old Problem." *Population Division, U. S. Bureau of the Census*. Technical Working Paper No. 13, Washington D.C.

Zaller, John. 1998. "Politicians as Prize Fighters: Electoral Selection and Incumbency Advantage." In John Geer, editor, *Party Politics and Politicians* Baltimore: Johns Hopkins University Press.

Table 1: Balance for Placebo Test Covariates

	Before Matching			After Matching		
Variable	mean diff	D-statistic	KS-pvalue	mean diff	D-statistic	KS-pvalue
Dem Pres. vote share '00	.0447	.100	0.00	.00459	.0337	0.953
Dem House vote share '00	.159	.305	0.00	.00693	.0344	0.678
Dem House vote share '98	.127	.340	0.00	.00585	.0368	0.996
Dem Senate vote share '00	.0426	.120	0.00	.00576	.0317	0.846
Dem Governor vote share '98	.0305	.0974	0.00	.00510	.0241	0.942
Dem Att. Gen. vote share '98	.0353	.141	0.00	.00683	.0358	0.868
Dem Comptroller vote share '98	.0304	.208	0.00	.00499	.0373	0.994
Voter turnout '00	.0331	.102	0.00	.00607	.0327	0.943
Voter turnout '98	.028	.199	0.00	.0111	.0378	0.235
Registration '00	.0308	.157	0.00	.00736	.0608	0.601

The mean difference are the the simple differences between treatment and control, the D-statistic is the largest difference in the empirical QQ-plot on the scale of the variable, and the KS-pvalue is from the bootstrapped Kolmogrov-Smirnov test.

Table 2: Results of Placebo Tests with All Key Covariates

	Estimate	95% CI		p.value
Incumbent vote '02	0.00245	−0.00488	0.00954	0.513
Turnout '02	0.00334	−0.00443	0.0112	0.412

Post Genetic Matching point estimates and confidence intervals are from Hodges-Lehmann Interval Estimation. The substantive results are unchanged if overdispersed GLM estimation or Abadie-Imbens SE are used instead. There are 474 observations.

Table 3: Results of Placebo Tests for Past Presidential Vote Only

	Estimate	95% CI		p.value
Data 1				
Incumbent vote '02	0.0237	0.0178	0.0294	0.000
Turnout '02	−0.000785	−0.00875	0.00718	0.846
Data 2				
Incumbent vote '02	0.0285	0.0160	0.0413	0.000
Turnout '02	0.00246	−0.0180	0.0218	0.831

Post Genetic Matching point estimates and confidence intervals are from Hodges-Lehmann Interval Estimation. The substantive results are unchanged if overdispersed GLM estimation or Abadie-Imbens SE are used instead. The first dataset contains 2666 observations, and the second dataset contains 412 observations.

Table 4: Incumbency Advantage Same Party, Standard “Old Voters, New voter” Design Using only Past Presidential Vote

	Estimate	95% CI		p.value
Incumbent vote ’04	0.0579	0.0470	0.0700	0.000
Incumbent vote ’06	0.0104	0.00317	0.0176	0.005

Post Genetic Matching point estimates and confidence intervals are from Hodges-Lehmann Interval Estimation. The substantive results are unchanged if overdispersed GLM estimation or Abadie-Imbens SE are used instead. There are 412 observations.

Table 5: Incumbency Advantage Same Party, Best Design

	Estimate	95% CI		p.value
Incumbent vote ’04	0.00637	−0.00428	0.0177	0.254
Incumbent vote ’06	0.00843	−0.00938	0.0258	0.457

Post Genetic Matching point estimates and confidence intervals are from Hodges-Lehmann Interval Estimation. The substantive results are unchanged if overdispersed GLM estimation or Abadie-Imbens SE are used instead. There are 166 observations.

Table 6: Incumbency Advantage Same Party, Second Best Design

	Estimate	95% CI		p.value
Incumbent vote ’04	0.00214	−0.00807	0.0124	0.690
Incumbent vote ’06	0.00472	−0.00539	0.0149	0.378
Turnout ’04	−0.00361	−0.0195	0.0126	0.652
Turnout ’06	0.00944	−0.00397	0.0240	0.162

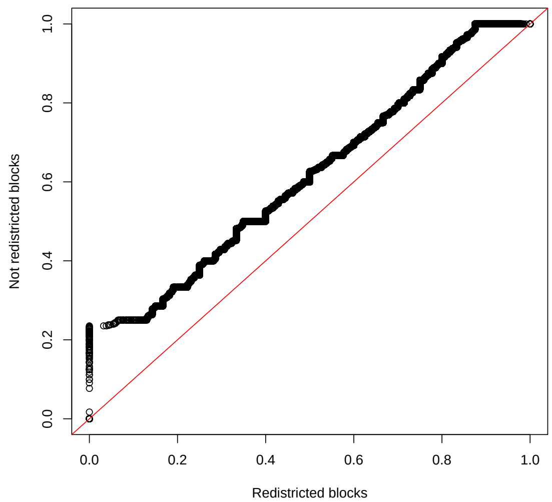
Post Genetic Matching point estimates and confidence intervals are from Hodges-Lehmann Interval Estimation. The substantive results are unchanged if overdispersed GLM estimation or Abadie-Imbens SE are used instead. There are 434 observations.

Table 7: Incumbency Advantage Different Party, Best Design

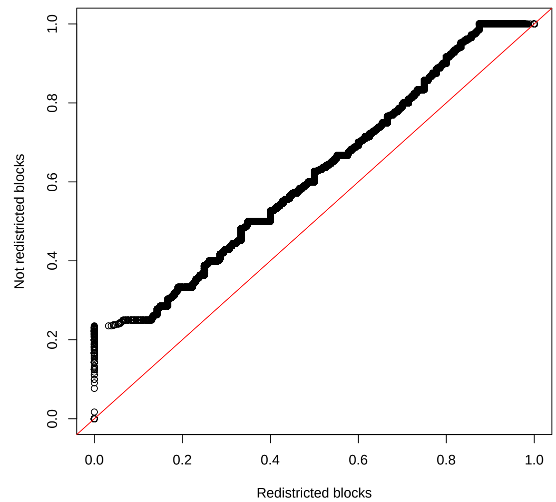
	Estimate	95% CI		p.value
Incumbent vote '04	0.119	0.0595	0.191	0.0000
Incumbent vote '06	0.0389	0.00973	0.0692	0.0106
Turnout '04	-0.00405	-0.0482	0.0352	0.903
Turnout '06	-0.0261	-0.0772	0.0212	0.299

Post Genetic Matching point estimates and confidence intervals are from Hodges-Lehmann Interval Estimation. The substantive results are unchanged if overdispersed GLM estimation or Abadie-Imbens SE are used instead. There are 70 observations.

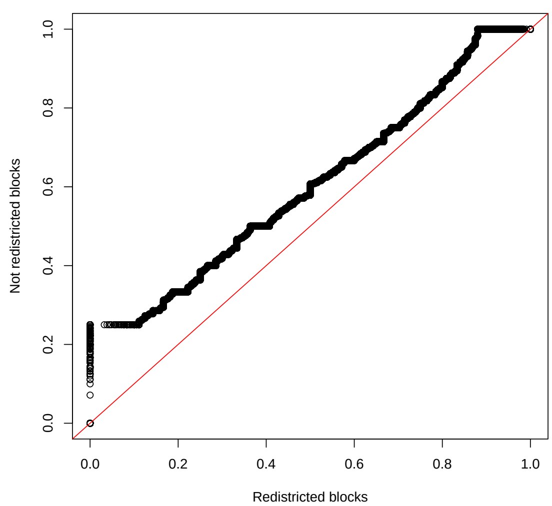
Figure 1: QQ Plots for 2000 Vote for Incumbent House Member



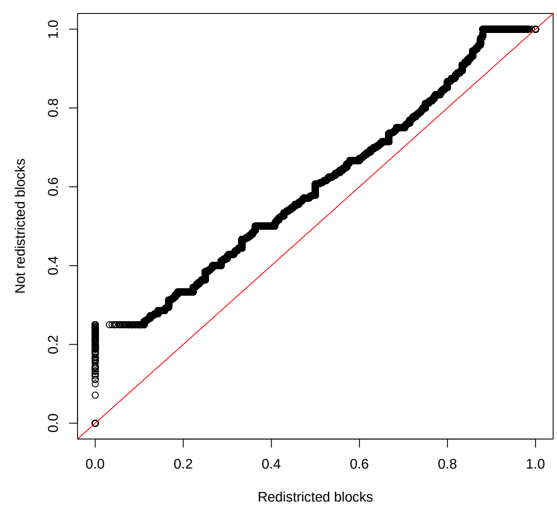
(a) California, unconditional



(b) Texas, unconditional



(c) California, matched on 2000 presidential vote



(d) Texas, matched on 2000 presidential vote

Figure 2: QQ Plot for Placebo Test Conditioning on All Key Covariates

2002 Vote for the Incumbent House Member

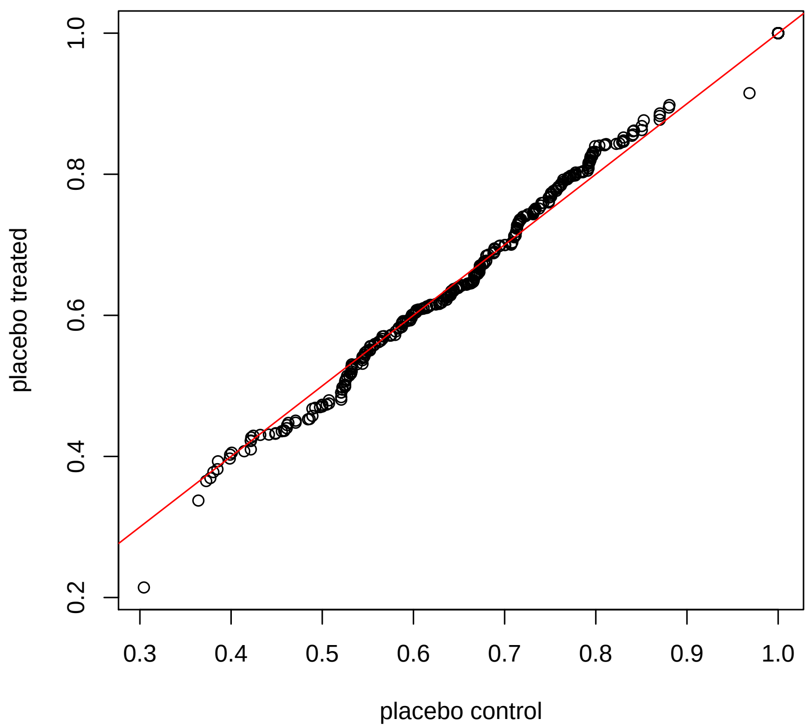
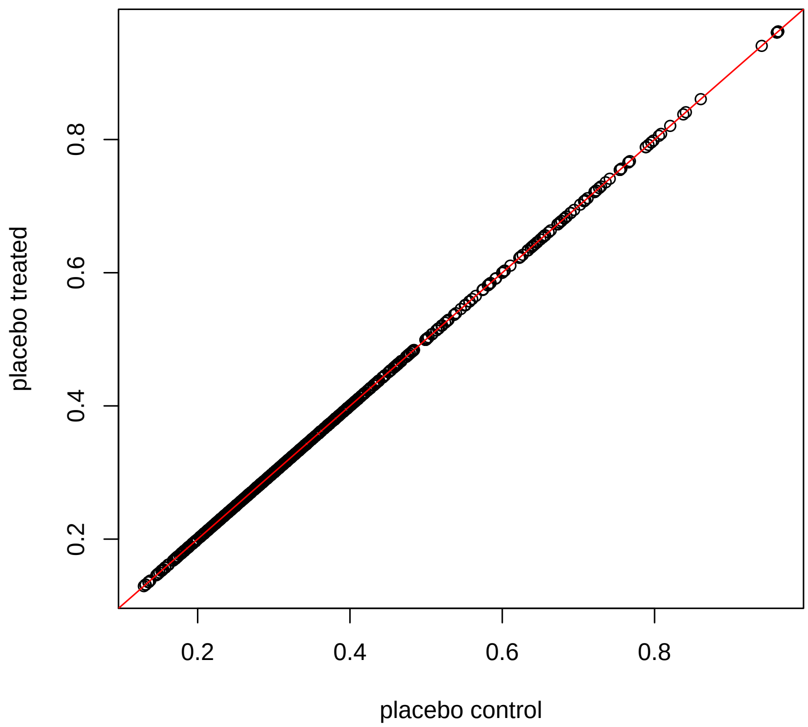
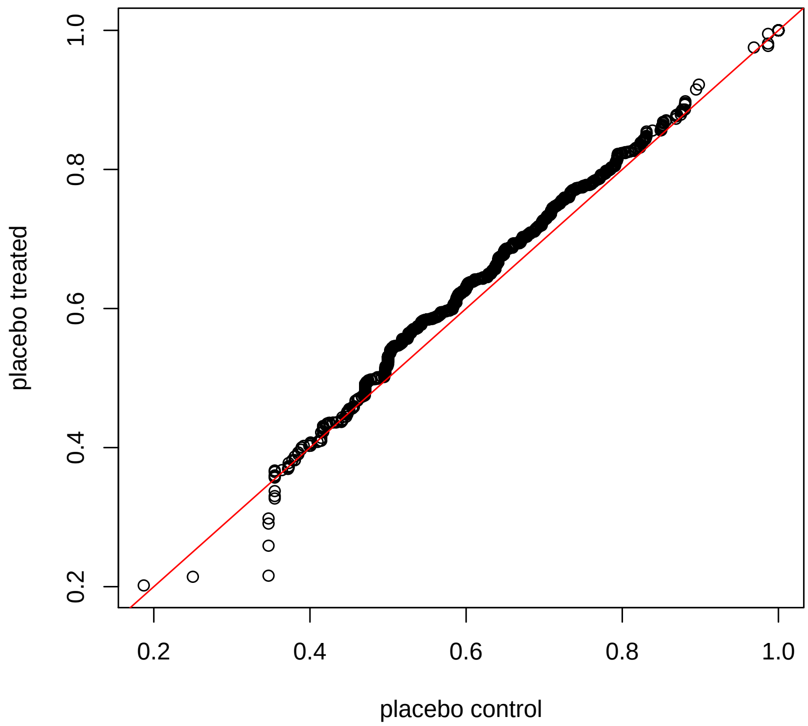


Figure 3: QQ Plots for Placebo Test Conditioning Only On Presidential Vote



(a) 2000 Presidential Vote (Baseline)



(b) 2002 Vote for the Incumbent House Member (Placebo Test)