Exploiting Tom DeLay: A New Method for Estimating Incumbency Advantage and the Effect of Candidate Ethnicity on Turnout*

Jasjeet S. Sekhon† and Rocío Titiunik‡

Associate Professor
Travers Dept. of Political Science
UC Berkeley

Ph. D. Student
Agriculture and Resource Economics
UC Berkeley

9/29/2007 (02:01)

*For valuable comments we thank Steve Ansolabehere, Bob Erikson, Luke Keele, Gary King, Walter Mebane, Jr., Rebecca Morton, Shigeo Hirano, 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, 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 and for estimating how voters respond to other characteristics of candidates such as their ethnicity. Our method relies upon the unique multiple redistrictings which were done in Texas between 2002 and 2006. We demonstrate that the manner in which previous work has used redistricting to identify the causal effect of incumbency results in biased estimates because the correct potential outcomes are not used. Strikingly, even if voters were redistricted at random, previous uses of redistricting as a research design would not yield unbiased estimates. And 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.

Contrary to the extant literature, we find that in U.S. House elections 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.

We also estimate the effect of incumbent ethnicity on voter behavior and find that turnout decreases when voters are moved from a white to an Hispanic incumbent. Hispanic registration declines when voters are so moved, while non-Hispanic registration does not significantly change.
1 Introduction

We propose a new method for estimating incumbency advantage. We 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 surprising result that the current way in which scholars use redistricting to estimate incumbency leads to bias even if voters were 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 of what it fails to condition on, and the assumption empirically fails a placebo test which our design passes. We also exploit redistricting as an identification strategy to address important questions regarding the way in which voters respond to candidates’ race and ethnicity.

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. 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). 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


2For a review on the debate about the causes of the increase in the incumbency advantage see Cox and Katz (1996) and Krebsi 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) provide a formal analysis of these difficulties. The authors also propose 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. 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.

---

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. 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.

Just because new voters do not have the same history with the incumbent as old voters does not mean that they will not respond to an incumbent’s record of constituency service

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

\[ \text{For Ansolabehere et al. (2000) this figure corresponds to the 1972–1988 period. Consistent with previous literature, their estimate of the incumbency advantage is smaller for earlier periods.} \]
such as providing pork or public position-taking if new voters can somehow learn about these activities. One could argue that this old voters vs. new voters design actually estimates how quickly new voters learn about the type of their new incumbent. That is, how quickly they learn about how good of a job their new incumbent does at bringing home federal funds, performing casework and all of the other components of what is often called the personal vote.

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 outcomes are not used. Strikingly, even if voters were redistricted at random, previous uses of redistricting as a research design would result in biased estimates. 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, and the assumption empirically fails a placebo test even when the correct potential outcomes are used.

Our research design uses the correct potential outcomes and makes a selection on observables assumption which passes the placebo test. This design is made possible by the unique multiple redistrictings which were done in Texas between 2002 and 2006. We estimate the effects of incumbency using Genetic Matching (Sekhon forthcoming) to achieve covariate balance.

Contrary to the extant literature, we find that in U.S. House elections there is no candidate specific personal vote—i.e., there is no personal incumbency advantage. That is, 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. We estimate a zero personal incumbency advantage even though the standard Ansolabehere et al. (2000) “old voters, new voters” design (which fails the placebo test) estimates a highly significant positive personal 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. The absence of a candidate
specific incumbency advantage is consistent with theoretical arguments that existing positive
estimates of incumbency advantage are plagued by selection problems (Cox and Katz 2002;
Zaller 1998). And a significant estimate for the party incumbency advantage is consistent
with the results of Lee (forthcoming) who finds a significant party incumbency effect by the
use of a regression discontinuity design.

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 the type of their new incumbent well enough that they do not vote differently
than old voters. 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.

We also use redistricting as an identification strategy to address important questions
regarding the way in which voters respond to candidates’ race and ethnicity. Estimating the
effect on voter behavior of the race and ethnicity of incumbents is also plagued by selection
problems. For example, as we show, Hispanic precincts with high baseline turnout are more
likely to be moved during redistricting to Hispanic majority-minority districts than Hispanic
precincts with relatively low turnout. Therefore, any estimation strategy, such as that of
Barreto, Segura, and Woods (2004), which relies upon the correlation between turnout and
the ethnicity of the incumbent without taking into account this selection processes will result
in biased estimates—estimates which will be generally positively biased. Gay’s (2001) study
of the relationship between black Congressional representation and political participation
suffers from a similar selection problem. Black Congressional incumbents may be more
likely to be elected in districts where blacks have higher turnout than where blacks have lower
turnout. Our research design can be used to estimate these ethnic effects while accounting
for the selection process, and we offer some preliminary results.

The paper is organized as follows. In the next section we discuss our research design
and in Section 3, we briefly describe our data, which is more fully described in Appendix A. In Section 4, we outline our estimation method, Genetic Matching. Section 5 presents the results, and Section 6 concludes.

2 Research Design

We exploit 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).

Although this comparison seems intuitive, a careful analysis reveals that there are important methodological complications. First, 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, only under strong assumptions does the difference in the behavior of old voters and new voters identify the effect of incumbency. In this section, we propose different research designs to address these issues.

To illustrate the first point, we consider the following thought experiment. Imagine that just before election $t$ a redistricting plan randomly redraws the boundaries of an arbitrary district (call it 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 (call it district $B$). From the point of view of district $B$’s incumbent, at election $t$ (the first election after redistricting) 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 advertising campaign, the same cues, etc.

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 that 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. Put differently, 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 will 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

\footnote{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.}
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 $t$, 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 $B$).\footnote{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.} 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

$$Y(i, t) = Y_0(i, t) \cdot (1 - T_i) \cdot (1 - D_i) + Y_1(i, t) \cdot (1 - T_i) \cdot D_i + Y_2(i, t) \cdot T_i \cdot D_i$$  \hspace{1cm} (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 will concentrate 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:

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

$$ATT_1 \equiv E[Y_2(i, t) - Y_1(i, t) \mid T_i = 1, D_i = 1]$$  \hspace{1cm} (3)$$
It can be shown that the following condition is sufficient for $ATT_0$ to be identified:\footnote{For a formal treatment of these and related assumptions, see, for example, \citet*{Heckman:1997}.}

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

(4)

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

$$E[Y_1(i, t) \mid T_i = 1, D_i = 1] = E[Y_1(i, t) \mid T_i = 0, D_i = 1]$$

(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 old neighbors (i.e, the voters that are left behind in the new voters’ original district). But 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 strategy, 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.

Nonetheless, we can use this design to estimate how voters react to a change in the race
or ethnicity of their incumbent. Indeed, to estimate the effect of incumbent race on voter turnout it is most natural to compare new voters and old neighbors. Unlike in the case where we wish to estimate the personal vote for an incumbent, in this case we wish to consider the different electoral environment which, say, a Hispanic incumbent brings about rather than a white incumbent. Therefore, we estimate the effects of moving new voters from a white incumbent to another white incumbent, and we separately estimate the effects of moving new voters from a white incumbent to a Hispanic incumbent. Note that in this case we do want the electoral environment to be different for new voters and old voters after redistricting, as we want to capture precisely how new voters react to a change in the electoral environment that involves a change in the ethnicity of the incumbent. Thus, for the purposes of estimating the effect of incumbent race on turnout, the best old-neighbors design is indeed the design which is most closely related to what we wish to estimate.

We have, however, 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 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:

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

where \( X \) is a vector of observable characteristics. Assumption (6) can be shown to identify \( ATT_1 \) conditional on \( X \) and is considerable 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

\footnote{When we say “white incumbent” we refer to a non-Hispanic white incumbent.}
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. 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 not be achieved even if a strong identifying condition were 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 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.

\footnote{See [Heckman, Ichimura, Smith, and Todd, 1998] for a formal proof that the lack of common support introduces bias.}
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. As can be seen, using new neighbors as counterfactuals poses important methodological challenges.

2.2 Consecutive redistricting: the best design

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 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. Although this design is not appropriate for estimating the personal vote, it is appropriate for estimating the effect of incumbent’s race and ethnicity on turnout.

But there is still another design available. Consider a modification of the thought experiment introduced above. 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: the voters who always belonged to B (henceforth original voters), the voters who became

\[\text{In this case, for example, the movement from one white Democratic incumbent to another could be restricted further to consider only white Democratic incumbents with the same ideology—i.e only moderate Democrats or only extreme Democrats.}\]
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 $Y_0(i,t+1)$ denote the outcome of $i$ at election $t+1$ if $W_{i,t+1} = 0$ and $Y_1(i,t+1)$ 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

$$\text{ATT}_B \equiv E[Y_1(i,t+1) - Y_0(i,t+1)|W_{i,t+1} = 1]$$

which is identified under

$$E[Y_0(i,t+1)|W_{i,t+1} = 1] = E[Y_0(i,t+1)|W_{i,t+1} = 0]$$

In words, $\text{ATT}_B$ 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

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

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:

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

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.\footnote{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.}

Of course, to actually implement the best design one would need three consecutive elections held under three different redistricting plans, a situation that is hard to find in practice given that district boundaries are normally redrawn after each decennial census. But Texas has recently been an exception to this rule. As in the rest of the United States, in 2001 Texas congressional districts were redrawn after the reapportionment that followed the 2000 census, which implied that districts changed between the 2000 and the 2002 election. But congressional districts were changed again before the 2004 elections, in a highly controversial mid-decade plan that was engineered by former Republican House Majority Leader Tom DeLay \cite{Bickerstaff2007}. See appendix \ref{app:tx_redistricting} for a detailed description of Texas redistricting plans.
Thus, redistricting in Texas gives 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.\footnote{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.} 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 redistricting as an identification strategy for the incumbency effect.

In sum, we have identified three different designs to exploit redistricting as an identification strategy: the best design that compares early and late new voters when at least two redistricting plans are consecutively implemented, the second best design that compares new voters with their new neighbors in the new district, and the old-neighbors best design that compares new voters with their old neighbors in their original district. As mentioned above, we do not use the old-neighbor best design to estimate the incumbency advantage because the necessary modifications would introduce post-treatment bias. However, we do use this design to estimate how voters react to a change in the race or ethnicity of their incumbent.

We introduced the different designs by means of presenting different ideal experiments in which redistricting was supposed to be random, which assured the exchangeability of new voters and the different types of old voters considered. Since in practice district boundaries are not randomly modified, in order to achieve identification of the parameters of interest in the three analyzed designs we make the assumption that, *conditional on certain observable characteristics*, the group of redistricted voters that we define as the treatment group is exchangeable with the group of non-redistricted voters that we choose as the control group. This is undoubtedly a strong assumption, but one that becomes increasingly more plausible the larger the set of observable characteristics in the conditioning set. More importantly,
we use the same data which participants in the redistricting battles fed into their computer programs to design their various redistricting plans.

Finally, we are interested in how the best design and the new-neighbor second-best design can be compared. There are at least two differences that are worth mentioning. First, even if both designs yield unbiased estimates, they need not give the same estimate of the incumbency effect as the designs estimate different parameters. In particular, since in the best design early new voters spend only one additional election in the new district, this design captures the incumbency effect that can accrue in a single election cycle (which in this case lasts two years). In contrast, in the new-neighbor second-best design original voters in the new district will typically have spent several elections with the incumbent, and hence this design estimates the incumbency effect that accrues after a longer period of time. Second, the best design 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) 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. We can arbitrarily denote 2000 to be the baseline year, and our placebo test is that in 2002 there should be no significant different between our treated and control groups. 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.

After briefly describing the data and our estimation method, we present the estimated effects of incumbency on electoral outcomes using the different designs described above.
3 Data

We analyze congressional elections in Texas between 2000 and 2006. Data on electoral returns were collected from the Texas Legislative Council (TXLC) at the Voting Tabulation District (VTD) level. VTDs are census blocks grouped to approximate voting precincts as closely as possible, and therefore they provide a link between census data and electoral data (for details about how VTDs are constructed, data sources, and other issues regarding data construction, see Appendix A).

Since all election data corresponding to the 1998, 2000, 2002, 2004 and 2006 elections are in terms of 2004 VTDs, we are able to track the same geographical unit over time. Election returns contain 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. Election returns reported by the TXLC include absentee ballots. We also added DW-Nominate scores (Voteview Website, \url{http://voteview.com/dwnomin.htm}) and data on challengers’ quality\footnote{Challenger’s quality data was directly provided to us by Gary Jacobson.} to our dataset.

We constructed VTD-level census data and merged these data to the electoral data provided by the TXLC. Census data from Summary File 1 was easily obtained at the VTD level by aggregating census blocks up to the VTD level. Census data from Summary File 3 was also taken to the VTD level, although the assignment of Summary File 3 variables to 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.

Our final dataset contains 8,040 VTDs. Every VTD in this final sample was assigned
to the congressional district it belonged to in each general election between 2000 and 2006, according to the congressional district plan that was effective at the time of each election.

4 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 (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:

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

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 \(X\).\(^{15}\)

Note that if one has a good propensity score model, one should include it as one of the covariates. If this is done, both propensity score matching and Mahalanobis matching can

\(^{15}\)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.
be considered special limiting cases of Genetic Matching: if the propensity score contains all of the relevant information in a given sample the other variables will be given zero weight\textsuperscript{16} while GenMatch will converge to Mahalanobis distance if that proves to be the appropriate distance measure.

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\textsuperscript{17}.

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.\textsuperscript{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 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\textsuperscript{18}. 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

\textsuperscript{16} Technically, the other variables will be given weights just large enough to ensure that the weight matrix is positive definite.

\textsuperscript{17} 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.

\textsuperscript{18} For example, using several descriptive metrics, one is unable to recover reliably the experimental benchmark in a testbed dataset for matching estimators\textsuperscript{Dehejia and Wahba (1999)}. 
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.\textsuperscript{19} 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. Therefore, if an analyst is concerned that matching may increase the bias in some linear or nonlinear function of $X$ even if the means are reduced, GenMatch allows the analyst to put in the loss function all of the functions of $X$ which may be of concern.

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). Even when the EPBR property holds and the mapping from $X$ to $Y$ is linear, GenMatch has better efficiency—i.e., lower mean square error (MSE)—in finite samples. When the EPBR property does not hold as it generally does not, GenMatch retains appealing properties and the differences in performance between GenMatch and the other matching methods can become substantial both in terms of bias and MSE reduction. In short, at the expense of computer time, GenMatch dominates the other matching methods in terms of MSE when assumptions required for EPBR hold and, even more so, when they do not.

Although GenMatch can be combined with various matching methods such as optimal

\textsuperscript{19}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 process continues iteratively until ties are broken.
matching or greedy matching, in this paper we use 1-to-1 matching with replacement.

5 Results

5.1 Placebo Test

As discussed in our research design section, because redistricting involves the nonrandom assignment of 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 relies on the fact that we observe the behavior of VTDs which will be redistricted before redistricting occurs. As such, 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. We arbitrarily denote 2000 to be the baseline year, and 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. The substantive results are unchanged if either bivariate overdispersed GLM estimation or if Abadie-Imbens standard errors are used instead.

Figure 1 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

---

The substantive results are unchanged if either bivariate overdispersed GLM estimation or if Abadie-Imbens standard errors are used instead.

Figure 1 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

---

The substantive results are unchanged if either bivariate overdispersed GLM estimation or if Abadie-Imbens standard errors are used instead.

Figure 1 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
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. The top panel in Figure 2 presents the balance on Presidential vote in 2000 matching on only this variable. As can be seen, balance is excellent. The bottom panel 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 2, it may be argued that balance isn’t good enough on past Presidential vote. However, the QQ-plot for Presidential vote in Figure 2 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. Figure 3 presents these results and shows that although balance of Presidential vote is terrific, balance on the House vote proportion in 2002 is still not good. The balance on presidential vote has now improved to a mean difference of 0.000156, a maximum difference of 0.000362 and bootstrapped KS p-value of 1. But as the Data 2 results in Table 3 show, there is still a significant effect for incumbent vote in 2002 (p-value of 0.00001).

5.2 Personal Incumbency Advantage

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% personal 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 of the personal vote provide by Ansolabehere et al. (2000) in the modern era. And it is somewhat larger than Desposato and Petrocik’s
estimate from California. The standard design applied to Texas estimates a positive and highly significant personal 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. This estimates the personal incumbency advantage which is the candidate specific advantage that results from the benefits of office holding such as name recognition, opportunities to perform constituency service, public position taking, and federal resources such as the ability to provide pork.

The personal incumbency advantage 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 personal 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.

5.3 Incumbent Party Advantage

Table 7 presents the results for the incumbent party advantage using our preferred design. 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.

5.4 Estimating the Effect of Incumbent Ethnicity

Given the paucity of black members of Congress in Texas, we are only able to reliably estimate the ethnic effects of white and Hispanic members. Table 8 presents the results of being moved from a white incumbent to another white incumbent of the same party. And Table 9 presents the results for moving from a white incumbent to a Hispanic incumbent of the same party.

Both tables show that there is no significant effect on the vote proportion which the incumbent receives. And there is no effect on turnout when VTDs are moved from a white incumbent to another white incumbent. However, when VTDs are moved from a white incumbent to a Hispanic incumbent even of the same party, there is a large and significant drop in turnout: -4.1% in 2004. But this effect diminishes by 2006, and it is in fact no longer significant.

Given the aggregate nature of the data, it is impossible to say with any reliability whether this significant turnout decrease is due to the behavior of whites, Hispanic or some other group of voters. VTDs and precincts in Texas are heterogeneous enough that the method of bounds is not by itself informative to make ecological inferences. The subset of VTDs for which the method of bounds is informative is tiny. Texas is unlike Cook County, IL, where the method of bounds itself is highly informative for a significant minority of VTDs (Herron and Sekhon 2005). We are in the process of gathering individual voter files from Texas and using Spanish surname matches to try to get a better understanding of who exactly is turning out less.

We are able to examine how overall registration and ethnic specific registration changes
over time. We are able to breakdown the registration data by Hispanic and non-Hispanic registration. Table 10 presents these results. Hispanic registration is less in both 2004 and 2006 than it would have been if the incumbent had remained white. There is no significant change in non-Hispanic registration although the point estimates are negative.

6 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. Natural experiments require careful theoretical and statistical work to make valid inferences. There are the obvious problems 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

---

22 Hispanic voter registration is estimated by surname match between the State Voter Registration Master File and the Census list of Spanish surnames. See Appendix A for additional details.
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 there is no personal incumbency advantage even though the standard “old voters, new voters” design estimates a highly significant positive personal incumbency advantage 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 personal vote (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 incumbent party effect 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 incumbent party effect suggests that partisanship plays an important part in the incumbency advantage—indeed the key part given that we do not find a personal incumbency effect. 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 Data

Data on election returns and relationship between census blocks and congressional districts and were obtained from the Texas Legislative Council (TXLC)’s ftp website (ftp://ftpgis1.tlc.state.tx.us, with updates as of May 02, 2007) and from special requests from the TXLC. Demographic and socioeconomic data was obtained from the U.S. Census Bureau Summary File 1 and Summary File 3 datasets. 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.

All electoral data are collected from the TXLC ftp site is derived from precinct level returns provided by Texas counties and the Texas Office of the Secretary of State. For the purposes of redistricting, the TXLC allocates all precinct level election data to the most recent general election voting tabulation districts (VTD). While all U.S. Census Bureau data is reported by census geographical units, such as census tracts and blocks, election data in Texas is reported by voting precincts which may change between elections. VTDs are census blocks grouped to approximate voting precincts as closely as possible, and they were created by TXLC staff in a project sponsored by the U.S. Census Bureau to provide a link between
population and election data. They provide the closest approximation to each county voting precinct for which population data will be available.

Voting precinct maps are collected from all counties and merged into a statewide precinct map database in order to create a statewide map of VTDs. A computer program then selects the census block boundary nearest to the voting precinct boundary. Where the counties have drawn their precinct boundaries on visible features such as rivers, streets, etc., VTDs match county election precincts exactly. Where election precinct boundaries are drawn on nonvisible features, VTD boundaries are drawn along the census block boundaries that most closely follow the actual precinct boundaries (see Research Division of the Texas Legislative Council, 2000, 2001). In the vast majority of cases, either VTD boundaries are actual precinct boundaries or there is very little distance between the location of the precinct boundary and the location of the census block chosen as the VTD boundary. Only in a few cases a close approximation cannot be achieved. Once VTD boundaries have been determined, an algorithm is used to allocate the election data from precincts to census blocks, and these data are then added up to VTDs.

All election data corresponding to 1998, 2000, 2002, 2004 and 2006 elections are in terms of 2004 VTDs, which allow us to track the same geographical unit over time. Election returns contain total voter registration, Hispanic voter registration estimated by surname match between the State Voter Registration Master File and the 1990 Census List of Spanish Surnames voter turnout (reported as the sum of election day and early voting turnout), and candidate information for Democratic and Republican candidates (results from third-party, independent, or write-in candidates are not included unless the candidate received enough votes statewide to affect the outcome of an election). Candidate information includes the candidate’s name and party affiliation, identification of black and Hispanic candidates, and identification of incumbency status. Election returns reported by the TXLC include

23 “In most cases where the two [precincts and VTDs] do not match, the differences will be minor and will involve no population”, Texas Legislative Council (2000, p. 11).
absentee ballots. These are typically included in the early votes provided by the counties to the TXLC, and they are generally allocated to a VTD according to the location of the voter. In the cases where the county reports early votes as a county total, votes are allocated throughout the county according to an allocation algorithm.

We constructed VTD-level census data. Census data from Summary File 1 was easily obtained at the VTD level by aggregating census blocks up to the VTD level. Summary File 3 could not be as straightforwardly assigned to VTDs, as the smallest geographical unit for which Summary File 3 is reported is the block-group level, and there is no unique mapping between block-groups and VTDs. In order to obtain Summary File 3 at the VTD level, we replicated the observed block-groups percentages of each variable at the block level. For example, if 30% of the population of a given block-group speaks Spanish at home, it is assumed that 30% of the population of each block within that block-group speaks Spanish at home. This approximation obviously introduces some error, except in the case that a block-group is entirely contained within a VTD. Of the total 14,463 block-groups in which the state of Texas is divided, 6,643 belong to a single VTD.

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.

A.1 Final sample size

The territory of the state of Texas is divided into a total of 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. This occurs for one of two reasons: (i) if the data
are coming from a different year than 2004 the precincts for that election may be somewhat
different than they are for 2004, and therefore there could have been multiple precincts
covering the same area that the 2004 VTD now encompasses, (ii) counties occasionally have
“split precincts” and hence report multiple contests for a single precinct.

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 popula-
tion that satisfy this condition, which is our final sample size.

A.2 Mapping between VTDs and congressional districts

Every VTD in the final sample was assigned to the congressional district it belonged to
in each general election between 2000 and 2006, according to the congressional district plan
that was effective at the time of each election. This was done using files provided by the
TXLC mapping 2000 census blocks to congressional districts for each congressional district
plan enacted in this period. This mapping was then double checked with the name of the
district that was reported in the VTD-level election data; both sources agreed with almost
no exceptions.\textsuperscript{25}

B Texas redistricting plans in the 2000s

Texas implemented six different congressional district plans between 1990 and 2006.\textsuperscript{26}
After the reapportionment that followed the 1990 census, Texas gained three congressional
seats in the U.S. House of Representatives and reached a total of 30 seats. The district

\textsuperscript{25}In the few cases of disagreement, the district reported in the election data was the one assigned, as this
indicated the actual district for which the votes were cast.

\textsuperscript{26}See the Texas Legislative Council’s Redistricting website \url{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.
boundaries enacted by the old plan C001 were redrawn and the 1992 elections were held under the new districts enacted by plan C657. This plan was supposed to remain in effect until the following reapportionment in 2002, but a suit (Vera v. Richards, 861 F. Supp. 1304 (S.D.Tex. 1994), aff'd sub nom. Bush v. Vera, 517 U.S. 952, 116 S. Ct. 1941 (1996)) was filed in January 1994 in federal district court in Houston challenging Texas’ C657 plan as unconstitutionally racially gerrymandered. In September 1994, the court in Vera v. Richards issued an order allowing the congressional elections in November 1994 to be held using the C657 plan, but it directed the legislature to develop a new plan to correct invalid Districts 18, 29, and 30.

The 1994 congressional elections and the 1996 primaries were held under plan C657, but in August 6, 1996, federal district court (Vera v. Bush, 933 F. Supp. 1341 (S.D.Tex. 1996)) enters an order that redraws 13 of Texas’ 30 congressional districts according to the new C746 plan. The order also provides for a new primary, with candidates from all parties running against one another, to be held in the redrawn districts on the same day as the 1996 general election. The remaining 17 congressional districts are found to be legal and are unaffected by the new plan. 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.

However, 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 massive fleds by Democratic lawmakers to New Mexico and Oklahoma to break 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, after primary elections had already been held. The U.S. Supreme court found that Congressional District 23 violated Section 2 of the Voting Rights Act and on August 2006, the U.S. District Court for the Eastern District of Texas ordered Plan 1438C, which redrew congressional districts 15, 21, 23, 25, and 28. Elections for these five districts were held in a special election concurrent with the general election on November 7, 2006. The five districts from Plan 1438C were combined with the 27 unchanged districts in Plan 1374C to create a new statewide congressional plan, which was called Plan 1440C and was used in the 2006 general election.

References


Ansolabehere, Stephen and James Snyder. 2002. “The Incumbency Advantage in U.S. Elec-


Table 1: Balance for Placebo Test Covariates

<table>
<thead>
<tr>
<th>Variable</th>
<th>Before Matching</th>
<th>After Matching</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean diff</td>
<td>D-statistic</td>
</tr>
<tr>
<td>Dem Pres. vote share '00</td>
<td>.0447</td>
<td>.100</td>
</tr>
<tr>
<td>Dem House vote share '00</td>
<td>.159</td>
<td>.305</td>
</tr>
<tr>
<td>Dem House vote share '98</td>
<td>.127</td>
<td>.340</td>
</tr>
<tr>
<td>Dem Senate vote share '00</td>
<td>.0426</td>
<td>.120</td>
</tr>
<tr>
<td>Dem Governor vote share '98</td>
<td>.0305</td>
<td>.0974</td>
</tr>
<tr>
<td>Dem Att. Gen. vote share '98</td>
<td>.0353</td>
<td>.141</td>
</tr>
<tr>
<td>Dem Comptroller vote share '98</td>
<td>.0304</td>
<td>.208</td>
</tr>
<tr>
<td>Voter turnout '00</td>
<td>.0331</td>
<td>.102</td>
</tr>
<tr>
<td>Voter turnout '98</td>
<td>.028</td>
<td>.199</td>
</tr>
<tr>
<td>Registration '00</td>
<td>.0308</td>
<td>.157</td>
</tr>
</tbody>
</table>

The mean difference are 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

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>95% CI</th>
<th>p.value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incumbent vote '02</td>
<td>0.00245</td>
<td>-0.00488</td>
<td>0.00954</td>
</tr>
<tr>
<td>Turnout '02</td>
<td>0.00334</td>
<td>-0.00443</td>
<td>0.0112</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>95% CI</th>
<th>p.value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incumbent vote '02</td>
<td>0.0237</td>
<td>0.0178</td>
<td>0.0294</td>
</tr>
<tr>
<td>Turnout '02</td>
<td>-0.000785</td>
<td>-0.00875</td>
<td>0.00718</td>
</tr>
<tr>
<td>Data 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incumbent vote '02</td>
<td>0.0285</td>
<td>0.0160</td>
<td>0.0413</td>
</tr>
<tr>
<td>Turnout '02</td>
<td>0.00246</td>
<td>-0.0180</td>
<td>0.0218</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>95% CI</th>
<th>p.value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incumbent vote '04</td>
<td>0.0579</td>
<td>0.0470−0.0700</td>
<td>0.000</td>
</tr>
<tr>
<td>Incumbent vote '06</td>
<td>0.0104</td>
<td>0.00317−0.0176</td>
<td>0.005</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>95% CI</th>
<th>p.value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incumbent vote '04</td>
<td>0.00637</td>
<td>−0.00428−0.0177</td>
<td>0.254</td>
</tr>
<tr>
<td>Incumbent vote '06</td>
<td>0.00843</td>
<td>−0.00938−0.0258</td>
<td>0.457</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>95% CI</th>
<th>p.value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incumbent vote '04</td>
<td>0.00214</td>
<td>−0.00807−0.0124</td>
<td>0.690</td>
</tr>
<tr>
<td>Incumbent vote '06</td>
<td>0.00472</td>
<td>−0.00539−0.0149</td>
<td>0.378</td>
</tr>
<tr>
<td>Turnout '04</td>
<td>−0.00361</td>
<td>−0.0195−0.0126</td>
<td>0.652</td>
</tr>
<tr>
<td>Turnout '06</td>
<td>0.00944</td>
<td>−0.00397−0.0240</td>
<td>0.162</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>95% CI</th>
<th>p.value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incumbent vote '04</td>
<td>0.119</td>
<td>0.0595</td>
<td>0.191</td>
</tr>
<tr>
<td>Incumbent vote '06</td>
<td>0.0389</td>
<td>0.00973</td>
<td>0.0692</td>
</tr>
<tr>
<td>Turnout '04</td>
<td>−0.00405</td>
<td>-0.0482</td>
<td>0.0352</td>
</tr>
<tr>
<td>Turnout '06</td>
<td>−0.0261</td>
<td>−0.0772</td>
<td>0.0212</td>
</tr>
</tbody>
</table>

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.

Table 8: Estimating the Effect of Incumbent Race: White to White Incumbent

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>95% CI</th>
<th>p.value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incumbent vote '04</td>
<td>0.00555</td>
<td>−0.00503</td>
<td>0.0166</td>
</tr>
<tr>
<td>Incumbent vote '06</td>
<td>0.012</td>
<td>−0.000877</td>
<td>0.0267</td>
</tr>
<tr>
<td>Turnout '04</td>
<td>0.0491</td>
<td>−0.0169</td>
<td>0.114</td>
</tr>
<tr>
<td>Turnout '06</td>
<td>0.0129</td>
<td>−0.000877</td>
<td>0.0268</td>
</tr>
</tbody>
</table>

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 296 observations.
Table 9: Estimating the Effect of Incumbent Race: White to Hispanic Incumbent

<table>
<thead>
<tr>
<th>Estimate</th>
<th>95% CI</th>
<th>p.value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incumbent vote ’04</td>
<td>−0.00674</td>
<td>−0.0165</td>
</tr>
<tr>
<td>Incumbent vote ’06</td>
<td>0.0167</td>
<td>−0.0075</td>
</tr>
<tr>
<td>Turnout ’04</td>
<td>−0.0406</td>
<td>−0.0699</td>
</tr>
<tr>
<td>Turnout ’06</td>
<td>−0.0139</td>
<td>−0.0413</td>
</tr>
</tbody>
</table>

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 164 observations.

Table 10: Registration Changes: White to Hispanic Incumbent

<table>
<thead>
<tr>
<th>Estimate</th>
<th>95% CI</th>
<th>p.value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hispanic Registration ’04</td>
<td>−0.0197</td>
<td>−0.0338</td>
</tr>
<tr>
<td>Non-Hispanic Registration ’04</td>
<td>−0.00449</td>
<td>−0.0381</td>
</tr>
<tr>
<td>Hispanic Registration ’06</td>
<td>−0.0279</td>
<td>−0.0429</td>
</tr>
<tr>
<td>Non-Hispanic Registration ’06</td>
<td>−0.00233</td>
<td>−0.0393</td>
</tr>
</tbody>
</table>

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 164 observations.
Figure 1: QQ Plots for Placebo Test
Figure 2: QQ Plots for Presidential Vote only Placebo Test
2000 Presidential Vote (Baseline)

2002 Vote for the Incumbent House Member (Placebo Test)
Figure 3: QQ Plots for Presidential Vote only Placebo Test (tight caliper)
2000 Presidential Vote (Baseline)

2002 Vote for the Incumbent House Member (Placebo Test)