## CALTECH / MIT
## VOTING TECHNOLOGY PROJECT

**A multi-disciplinary, collaborative project of**
**the California Institute of Technology – Pasadena, California 91125 and**
**the Massachusetts Institute of Technology – Cambridge, Massachusetts 02139**

# WHO OVERVOTES, WHO UNDERVOTES, USING PUNCHCARDS? Evidence from Los Angeles County

**Betsy Sinclair**
**Caltech**

**R. Michael Alvarez**
**Caltech**

**Key words:** ***demographics, undervotes, overvotes, punchcards, residual votes, Los Angeles County***

**VTP WORKING PAPER #7**
**September 2003**

# Who Overvotes, Who Undervotes, Using Punchcards? Evidence from Los Angeles County

D.E. "Betsy" Sinclair†

R. Michael Alvarez‡

September 9, 2003

† Graduate Student, Division of Humanities and Social Sciences, California Institute of Technology (betsy@hss.caltech.edu); ‡ Professor of Political Science, Division of Humanities and Social Sciences, California Institute of Technology (rma@hss.caltech.edu). We thank Conny McCormack (Los Angeles County Registrar Recorder), and Brian Ikenaga and Vern Cowles (of the Los Angeles County Registrar Recorder Office) for their assistance with this project. Some of the data we use in our analysis came from the California Statewide Database, made available by the Institute of Governmental Studies at the University of California, Berkeley. We also thank Catherine Wilson for her insights. Karen

Kerbs and Mary King Sikora provided administrative help. Earlier versions of this research were presented at the 2002 Annual Meetings of the Midwest Political Science Association, and at Stanford University's Department of Political Science's American Empirical Seminar; we thank participants in both forums for their comments and critiques. This research was supported by the Caltech/MIT Voting Technology Project with a research grant from the Carnegie Corporation; however the conclusions reached here reflect the views of the authors and not necessarily the Caltech/MIT Voting Technology Project. This research was also supported by grants to Alvarez from the IBM Corporation through the University Matching Grants Program, and from the Haynes Foundation.

# Abstract

In this study we examine over- and undervotes from the November 2000 General Election in Los Angeles County. Los Angeles County is the nation's largest election jurisdiction, and it used a punchcard voting system in that election. We use precincts as our unit of analysis and merge the 2000 election data with census data and voter registration data; our dataset allows us to examine all of the countywide races in 2000 (including candidate and ballot measures). We use a multivariate statistical analysis employing negative binomial regression to test hypotheses regarding the relationship between precincts' political and demographic characteristics and over- and undervotes. We demonstrate that both over- and undervotes vary systematically across precincts in Los Angeles County, a finding that we argue has important implications for the representation of political interests.

# 1 Introduction

In every election ballots for some races are not counted. These votes, commonly called the "residual vote" or "rolloff rate", result from three different problems: overvoted ballots (ballots where there are more votes cast for a particular race than allowed), undervoted ballots (ballots where there is no indication of a vote for a race), and spoiled or uncounted ballots (ballots that are not counted at all for various reasons). As a result of the 2000 presidential election, there is now increased awareness of both overvoting and undervoting, and increased concern that both might be affected by the mechanism used to cast votes (called the "voting system"), by differences in voter sophistication, familiarity with the election process, or other voter attributes.

Clearly, some voting systems generate a higher residual vote rate than other systems, a fact noted in many studies (Ansolabehere 2001; Alvarez, Sinclair and Wilson 2003; Knack and Kropf 2003; Posner 2001; Brady et al. 2001; Montgomery 1985; Shocket, Heighberger and Brown (1992); GAO 2001; U.S. House Committee on Government Reform 2001). Punch card voting systems generally have the highest residual vote rates in recent elections, averaging 2.5% in presidential races from 1988-2000, followed by electronic voting systems, optical scan systems, lever machines, and paper ballots (Caltech/MIT 2001).[1] Furthermore, when studies have examined overvoting and undervoting rates for different voting systems, stark differences are found; some voting systems, like electronic and lever machines, can block overvoting — also, when certain voting systems are used (like precinct-based optical scanning and newer "touchscreen" voting machines), voters can check for overvotes and undervotes before their ballots are counted.[2]

However, we are not concerned simply with the existence of residual votes, or which voting systems produce higher rates of overvotes or undervotes. Instead, our concern is whether overvotes and undervotes occur randomly within a population of voters who use the same voting system, here the "Votomatic" punchcard voting system.[3] Thus, rather than examine residual voting rates *across different voting systems*, our approach here is to examine them *across precincts* in one election jurisdiction to see if they are randomly distributed over the population of voters, *holding the type of voting system constant.*

This is a critical question for students of elections. While there is a vast literature on voting behavior, too large to easily summarize here, this literature has largely ignored what happens when a citizen enters the voting booth and casts a ballot. Instead, research in this area usually assumes that a voter's preferences are directly and accurately translated into ballot choices, and hence are converted certainly into election outcomes. If error is introducted into the voting process, either by the mechanical or electronic device that records and counts ballots, or by differences in how certain voters interact with the voting device, then election returns may not accurately reflect voter preferences. Errors and uncounted votes also make it possible that in a close election, the election outcome might be affected by imperfections in the voting process. This, combined with the possibility that uncounted or uncountable votes might be more likely to come from certain types of voters, implies that the voting process itself might affect the quality of political representation.

The most pressing question for our study, then, is whether certain groups of voters are more likely to cast residual votes, in jurisdictions using punchcard voting systems. Recent studies have found that residual votes tend to be concentrated in areas with more nonwhite, poor, and poorly educated voters or residents (GAO 2001; U.S. House Committee on Government Reform 2001; Alvarez, Sinclair and Wilson 2003; Herron and Sekhon 2003; Knack and Kropf 2003; Posner 2001; Quirk et al. 2002; Fraser 1985; Tomz and Van Houweling 2003). If this is true, the interests of these types of voters may not be representated accurately in the political process.

Further evidence indicates that "roll-off" is greater for nonwhites and lower education voters (Walker 1966; Darcy and Schneider 1989; Bullock III and Dunn 1996) and that overvoting is greater in precincts with relatively large numbers of blacks, Hispanics, and registered Democrats in the Florida 2000 general election (Herron and Sekhon 2003). But some studies (especially Montgomery 1985 among the early studies, and more recently Ansolabehere 2001 and GAO 2001) find little support for this argument; other studies find that nonwhite voters had higher rates of uncounted votes in counties using punchcard voting systems (Alvarez, Sinclair and Wilson 2003; Tomz and Van Houweling 2003). Last, a number of studies have shown that some voting systems have historically decreased turnout and increased residual vote (White 1960; Mather 1964; Mather 1986; Nichols and Strizek 1995; Thomas 1968).

The possibility that nonwhite voters might both be more likely to use voting systems (like punchcards) that have higher rates of uncounted votes, and the possibility that nonwhite voters have higher rates of uncounted votes than white voters when they both use the same punchcard voting system, has led to a series of lawsuits claiming that these differences constitute violations of both the Voting Rights Act and the Fourteenth Amendment to the United States Constitution. Such cases have been filed recently in Illinois, Georgia, Florida, and California; some of these cases (for example in California) have resulted in the prohibition of some punchcard voting systems.[4] Furthermore, the recent federal election reform bill (the "Help America Vote Act") contains a number of provisions that may lead to the elimination of many current punchcard voting systems in the United States.

Our study uses data from Los Angeles County, California. We examine over- and undervote rates for a series of election contests in the 2000 election. We integrate into our database precinct-by-precinct information on voter registration, race and ethnicity, gender, and the use of non-English ballots. We test hypotheses regarding partisanship, race and ethnicity, and language use, on over- and undervoting. Early research by Steifbold (1965) provides an analytic framework for generating hypotheses about the types of voting behavior that can lead to over- or undervoted ballots (which Steifbold called "invalid" ballots). The first type of voting behavior leading to over- and undervoted ballots Steifbold (1965) called ballots cat by "apathetic invalids", "who respond(s) to the general electoral mobilization because it is the thing to do, but who remains actually indifferent to the political system. Whether out of ignorance, negligence or cross-pressures," they over- or undervote (page 406). The other group Steifbold called the "highly politicized invalid", "who knows exactly how he would vote if he could find the party corresponding to his ideas, but who, not finding that party, deliberately invalidates his ballot as a political act" (page 406).

Based on the available precinct-level data, we expect that over- and undervotes occur more often amongst voters with lower incomes and lower education levels. In Los Angeles County, these tend to be precincts that are heavily non-white, thus we anticipate seeing higher rates of over- and undervoting in heavily non-white precincts. Also, we expect that women may have higher under- and overvoting rates than men, as Steifbold found. Regarding partisanship, we anticipate first that partisans use their identifications as a voting cue, or as information searching mechanisms, and thus partisans should be less likely to over- or undervote. Last, we expect that minor party voters might be more disaffected with the political system and thereby more likely to cast over- and undervotes as protest votes.[5]

In the next two sections of this paper we discuss our rationale for this case study, and particulars of the data. Thereafter, we present our hypotheses in more detail. We test our hypotheses with a multivariate analysis of the precinct-by-precinct data. We conclude with a discussion of the implications of our work for the literature on elections and voting.

## 2 Punchcards in Los Angeles County

In this paper we use data from the largest and most complex election jurisdiction in the United States: Los Angeles County, California. Los Angeles County has a population of over 9.8 million residents, and more than 4 million registered voters. There are over 500 different political districts in Los Angeles County, roughly 5000 voting precincts, and ballots are provided for seven different language groups: Chinese, Korean, Japanese, Spanish, Tagalog and Vietnamese. In the 2000 presidential election, almost 2.8 million voters participated in the election, 68% of registered voters, more votes than were cast statewide in 41 of 50 states. A full 20% of the ballots cast came from absentee voters (543,143).

In addition to being the largest and most complex electoral jurisdiction in the United States, Los Angeles County also used the "Votomatic" punchcard voting system in 2000, and has used this same voting system for over three decades.[6] There are two basic types of punchcard voting systems; those that have many pre-scored, small rectangles on the punchcard that are referenced by numbers, and those that have fewer and larger pre-scored circles on which the actual ballot information is printed. "Votomatic", one example of the former type of punchcard voting system, was used in the 2000 presidential election in 5 California counties; while Datavote, used in 21 of California's counties in the 2000 election, is an example of the latter punchcard voting system.[7]

Thus, the use of Los Angeles County as a case for analysis is important for a number of reasons. First, with millions of votes cast, and almost 5,000 precincts, in a racially, economically, and socially diverse county, we have important variance on a number of independent variables. Second, as voters participating in the 2000 election all used the "Votomatic" punchcard system, we can hold the voting system constant in this analysis.

Of course, as this is a study of only one election jurisdiction in California, there are limitations to the generalizability of the results we report below. Los Angeles County is a complex election jurisdiction, and is more diverse than all but a small handful of American counties. Furthermore, elections in Los Angeles County are conducted within the parameters established by California election law, and as election laws differ across states this also potentially impacts our ability to draw generalizations to other counties in other states. With these caveats in mind, we proceed with our analysis.

## 3 Overvotes and Undervotes

We obtained our election data from the Los Angeles County Registrar Recorder, who provided a precinct-by-precinct breakdown of over- and undervotes, as well as all voter registration information. We then obtained our ethnic and racial, as well as gender, demographics from the California Statewide Database, which reorganizes census and voter registration data into consolidated voting precincts primarily for use in redistricting efforts.[8] It is important to note that our race and ethnicity data come from the 2000 U.S. Census, and thus represent estimates of race and ethnicity for the entire population in each precinct, not for the voting population; at this time the necessary data for estimating the racial and ethnic profiles of each precinct in Los Angeles County are not available. However, the gender data we use are taken from the voter registration rolls, and thus are estimates of the gender breakdown of the registered voter population in each precinct.[9] In order to integrate the over- and undervote data we obtained from the Los Angeles County Registrar Recorder, with the available census and registration data, we focus on consolidated precincts, a subset of Los Angeles County's total voting population in this election (votes cast in consolidated precincts represent 2,231,83 of the total ballots cast, including 12,089 presidential overvotes and 49,327 presidential undervotes). Consolidated precincts are a geographic organization for precincts that does not include absentee voters from the precinct unless the entire precinct is voting via mail ballot, thus as demographic information is not available for absentee voters in our database we do not include absentee votes in our multivariate analysis.[10]

We begin by presenting overvote and undervote summary statistics for all of the major candidate races on the Los Angeles County ballot in the 2000 election for only consolidated precincts: President, U.S. Senate, Los Angeles County District Attorney, and Los Angeles County Assessor.[11] The first two candidate races were statewide; the latter two were in Los Angeles County only. There were also nine ballot measures in this election, one county (A), and eight state (32-39). County Measure A would have increased the County's Board of Supervisors from five to nine members (it failed with

64% of those voting opposing it). Of the eight statewide measures, four generated some controversy: Proposition 34, that imposed new statewide campaign contribution and spending requirements; Proposition 36, a measure that changed the nature of penalties for drug offenses; Proposition 38, which would have instituted a school voucher system; and Proposition 39, which lowered the necessary vote threshold to 55% for passage of school bonds. Propositions 34, 36 and 39 all passed, while 38 failed, in both the state and Los Angeles County.

**Table 1 Goes Here**

Beginning with the four candidate races (ordered in Table 1 as they appeared on the Los Angeles County ballot), we see that the average number of overvotes per precinct is 2.4, while the number of undervotes in this race averages 9.8 per precinct. Moving to the next race on the ballot, the U.S. Senate race involving incumbent Dianne Feinstein against Republican challenger Tom Campbell (and five other minor party candidates), we see the overvote increasing slightly to an average of 2.8, but the undervote average increases dramatically to 22.3. The next race on the ballot, a heavily publicized and closely contested District Attorney race between incumbent Gil Carcetti (who lost the O.J. Simpson trial) and Steve Cooley, demonstrates a dramatically lower overvote average of 0.4, but a much higher overvote rate than either the Presidential or Senate races (85.4 average). The last candidate race in our analysis, the County Assessor's race, involved sixteen candidates: in this sixteen-candidate race we found that the average number of overvotes per precinct was 4.3, while the undervote was 101.8.

When we look at the percentages of over- and undervotes relative to all votes cast in each race, we see the same basic pattern as just discussed. Generally, with the only exception being the District Attorney's overvote rate, as we go down the ballot both over- and undervotes increase in the candidate races. For overvotes, the rate increases almost two-fold, but for undervotes, the rate increases almost ten-fold.

However, the over- and undervote statistics reported in Table 1 for the ballot measures are quite different from those for the candidate races. The first three ballot measures (A, 32 and 33) have relatively low average overvotes (just over 0.5 average per precinct or an average of about 0.12 of all ballots cast), but the overvote rate increases for Proposition 34 and stays somewhat higher than for the first three, for the remaining ballot measures. The same, though, is not true for the undervote rates. The lowest average undervote is for Proposition 38 (27.2 average undervote per precinct), while the greatest is for Proposition 37 (65.5 average per precinct). The measures with the lowest average undervotes are generally those that were most salient in this election: Measures 36, 38 and 39.

### 3.1 Explaining Over- and Undervotes

One important question — which we do not attempt to resolve in this paper — is whether over- and undervotes are unintentional or intentional acts. There is very little research directly on this question, other than Steifbold (1965) and Knack and Kroft (2001). In fact, we attempt to test for both intentional and unintentional over- and undervotes in our multivariate analysis, using Steifbold's basic categorization of "invalid" ballots. As we discussed earlier, Steifbold finds two types of "invalid" voters: those who are poorly informed and poorly connected to the political system, and those who are well informed, but ambivalent about their ballot choices. Of course, poorly informed voters might cast over- or undervoted ballots because they are making mistakes, are unfamiliar with the voting system, or as a true expression of their imperfect information; our data cannot easily differentiate intentional from unintentional over- or undervoting. However, well informed, but ambivalent, voters are likely to be casting protest over- or undervotes, thus we infer that if we see "invalid" votes from such voters, they are most likely intentional.

This past research implies that there might exist a level of voter confusion leading to over- and undervotes. We hypothesize that over- and undervotes are related to voter sophistication and connectedness with the political system, which include factors such as income and education, race and ethnicity, primary language, and gender. Thus, we suggest that the residual vote does not arise at random from the population and may influence the outcome of elections.

In order to understand the relationship between the demographic and political attributes of precincts and over- and undervotes in Los Angeles County, we use a form of multivariate analysis — negative binomial regression. We do not use linear regression because, although that is the standard statistical tool for multivariate data analysis, the dependent variables we use (overvotes and undervotes) have distributions that do not meet the fundamental assumptions of linear regression (King 1998, Section 5.9). Nor do we employ any of the common ecological inference techniques like Goodman's regression or King's EI model (Achen and Shively 1995; King 1997; Tomz and Van Houweling 2003); given that we have three different hypotheses about over- and undervoting we wish to test simultaneously (each of which involves multiple independent variables), it is computationally difficult if not impossible to use the existing ecological inference techniques.[12] We discuss the negative binomial model in more detail below.

Our multivariate negative binomial models, however, only include variables measuring a limited set of demographic and political characteristics of each precinct. The variables we have in our database include the partisan registration for the precinct (the fraction of registered voters that are Democratic, Republican, nonpartisan, or minor party registrants); the fraction of the precinct's population that is Hispanic, White, Black, or Asian, Hawaiian-Pacific Islander; the fraction of the registered voters who are female and male, and the number of ballots that were cast in languages other than English. Of course, we would prefer to also include a measure for the number of each candidates in the race, and other contextual variables (like education and income) that could be important in determining over- and undervotes. However, it is impossible to include such contextual measures in our multivariate models in this particular setting; testing for the simultaneous effects of context relative to demographic and political attributes of voters, as well as their possible interactions, we leave to future research.[13]

## 3.2 Multivariate Model Specification

Our hypotheses center on analyses of overvotes and undervotes as dependent variables in multivariate statistical models. Overvotes and undervotes, at the precinct level in Los Angeles County, have two important properties that make simple statistical analysis using common regression techniques problematic. The first is that we are using dependent variables that are counts; that is, we specify our dependent variables so that they measure the number of overvotes, or the number of undervotes, in each voting precinct.[14] As counts, both the over- and undervote measures have a lower bound of zero, and thus they do not have a continuous distribution. Secondly, the distribution of over- and undervotes both have a long tail to the right. Thus, both the lower bound and the rightward skewness of the over- and undervote distributions indicates that simple ordinary least squares regression is inappropriate.

Instead, we first utilized a Poisson regression model, appropriate for data of the sort we have. However, when we estimated Poisson regression models for over- and undervotes, for each of the four election contests under examination in this paper (a total of eight models), we found significant evidence in each case for *overdispersion* which is a violation of a fundamental assumption of the Poisson regression model. Thus, we used a negative binomial regression model, which allows for situations where there is overdispersion in a Poisson-type model.[15]

Formally, a Poisson model is specified as $y_i \sim Poisson(\mu_i)$ where $\mu_i = exp(x_i\beta)$; in our example $y_i$ is undervotes or overvotes. The negative binomial model follows a Poisson-like process in its estimation, where $y_i \sim Poisson(\mu_i*)$, with $\mu_i$ including another variable so that $e_i^u$ follows a gamma distribution with mean 1 and variance $\alpha$. That is, $\mu_i* = exp(x_i\beta + u_i)$ and $e_i^u \sim gamma(\frac{1}{\alpha}, \frac{1}{\alpha})$, where $\alpha$ is the overdispersion parameter. The original Poisson model corresponds to a situation where $\alpha = 0$.

To understand the coefficients in this regression, it is important to remember that the dependent variable is count data and that the model is highly nonlinear. While in an independent variable will indicate some change in the number of overvotes or undervotes in Los Angeles County, due to the nonlinearity of the model we transform the estimated coefficients to make them more readily understandable. Below we present the coefficients as "incidence risk ratios" (IRR), or as $e^{\beta}$, rather than report the untransformed coefficients that are more difficult to interprete. The IRR coefficients can be read much like the usual multiple regression coefficient, as the effect of a 1 unit change in an independent variable on the dependent variable. Later we also transform the estimated results into counterfactual estimates that present these results in another understandable manner.

While the dependent variable, overvotes or undervotes, is count data, the independent variables are percentages of the precinct's population. For example, the variable "Hispanic" is the voting age population which self-identified as Hispanic in the 2000 Census as a percentage of the voting age population in that precinct. This is true for all racial and gender variables. For the party registration variables, we consider the number registered for a party out of the total number of registered voters in the precinct. The non-English ballot variable measures the number of non-English ballots requested in the precinct out of the total number of ballots requested. Thus, all of the independent variables are percents, measuring the level of that variable in the precinct, with the exception of one variable.

Because the dependent variables are counts, it is possible that the relative sizes of the precincts would impact the number of overvotes more then the demographic variables. As a consequence, we include a measure of the differences in size between precincts as an independent variable. Since turnout would capture more than size, we create a variable called "size" which measures the total number of ballots cast in that precinct as a percentage of the total number of ballots cast in Los Angeles County and is scaled to fit appropriately with the other independent variables.[16]

## 3.3 Negative Binomial Model Results

We present the overvote negative binomial results in Table 2 and the undervote results in Table 3. Each table is organized so that the four candidate races are given in columns, with the first column for a particular race giving the IRR estimate and the standard error, while the second column provides the z-score, assessing the statistical significance of each estimate. Note that all of the IRR estimates are positive by definition, so the direction of the estimated effect (that is, whether it increases or decreases overvotes or undervotes) can be seen by examination of the sign of the z-score.

**Tables 2 and 3 Go Here**

In the overvote results (Table 2) the race variables — Hispanic, Black, and Asian-Hawaiian-Pacific Islander — all have positive and significant estimated effects for all contests. This means that a percentage change in these populations in each precinct will yield additional overvotes. To understand these coefficients, again it is important to note that they are denoted as $e^b$ rather then $b$. Also note that the demographic variables, except size, are all modeling percentages in a precinct. Thus, a 1% change in the population will produce an increase in the number of overvotes, which can be calculated by taking the natural log of the coefficient, multiplying that variable by 0.01, and then raising the coefficient $b$ to $e^b$ and multiplying by the number of precincts.

In this election, we have 5,038 voting precincts. This means that, for example, a 1% change in the Hispanic population in Los Angeles County increases the number of overvotes by approximately 217. To put this estimate into context, recall that there were 12,089 presidential overvotes in this election; thus, a 1% increase in the Hispanic population of LA County increases the presidential overvote by 1.7% (we know that this change would increase the overvotes because the z-score is positive).

The percent female variable is also positive and significant for all contests in the overvoting results, implying that precincts with more women have more overvotes. Next, Democratic registration and Republican registration are both negative and significant for all the contests. Also, in the overvote results we find that the measure for the fraction of nonpartisan voters in the precinct is also negative and significant for all contests except the Senate race, where it is insignificant. Thus when we put the partisan registration results together, and recall that the excluded category in this analysis is the number of minor party registrants, these multivariate results imply that in precincts with higher rates of minor party registration have more overvotes. These could be intentional, protest-based, overvotes, to the extent that minor party registrants in Los Angeles County are like Steifbold's informed and active, but disaffected, minor party voters. Last, in the overvoting results reported in Table 2 the non-English ballot request variable is significant and positive for all contests except the District Attorney, where again it is insignificant. The picture these regression results present are that race, gender, party registration, and ballot language all have an impact on over- and undervotes.

For undervotes, presented in Table 3, the race variables are again all positive and significant for the President, Senate, and District Attorney contests but negative and significant for the Assessor contest. The Assessor's race is clearly an anomaly, both due to the extremely high undervote rate (22.98%), and to the large number of candidates contesting this election. The apparent inconsistencies between our results for the Assessor's race and the other contests raises the possibility that our specification could be incorrect for the Assessor's race.

The percent female variable is again positive and significant for all contests. Thus, in precincts that have more women, both the overvoting and undervoting rates are higher than in precincts that are more male. This is consistent with Steifbold's (1965) results, and is also consistent with research on the relative level of political information of men and women (Alvarez 1997; Delli Carpini and Keeter 2000). If women possess less political information than men, this could result in higher rates of over- and undervoting.

The coefficients for the percent registered Democrat and nonpartisan are positive and significant for all contests except the presidential contest, where they are not significant. This is a reversal from the overvote regression. The percent registered Republican coefficient is negative and significant for the presidential contest but positive and significant for the Senate, District Attorney, and Assessor contest. No trends emerge in the undervote results for party registration.

Last, the coefficient on the non-English ballot variable is positive and significant for the President, Senate, and District Attorney results, but negative and significant for the Assessor contest. This is consistent with the general pattern seen in the overvoting results above, where we found that in most of the candidate races examined, precincts with higher rates of non-English ballot useage had higher overvote rates. Thus, perhaps we are seeing that, controlling for race, gender and partisanship, precincts with higher rates of non-English ballot use have less informed voters, or perhaps that these precincts have more voters who have difficulty navigating the political process due to language barriers.

Thus, like the overvote regression data, we see important and consistent patterns in the undervote multivariate results, especially for the race, non-English ballot, and gender variables. The coefficients of the Hispanic, Black, and Asian-Hawaiian-Pacific Islander voting populations are almost entirely positive, with the exception of the undervotes for the Assessor contest. It is likely that in the Assessor contest, with so many candidate choices and with the contest occurring later on the ballot, that other factors such as "roll-off" or lack of knowledge about the contest are more influential in preventing voters from casting a vote. The female voting population is universally significant and positive. The impact of party registration is unclear, as the coefficients fluctuate from positive to negative. Finally, the impact of a non-English ballot is such that it is often positive and significant with the exception of the Assessor's contest. This implies that race, gender, and ballot language – specifically, females, Hispanics, Blacks, Asian-Hawaiian-Pacific Islanders, and those who request a non-English ballot – are consistent predictors of higher over- and undervote rates.

For comparison, an OLS regression analysis produces similar results.[17] Using OLS, with the percent over- or undervote as the dependent variable and identical independent variables (minus the size variable which is no longer necessary), the results are qualitatively similar to the negative binomial results reported in Tables 3 and 4. For example, for presidential overvotes, the only difference is that all the OLS coefficients are significant whereas the negative binomial coefficient for female is not significant. However, for reasons articulated above, we prefer the negative binomial model because the data is not linear. With clusters of data at zero, it makes more sense to look at a nonlinear model. The negative binomial model is specifically designed to examine event data and is well-suited for over- and undervotes. Although this model is a better fit for the data, it does produce coefficients which are more difficult to understand. As a consequence, we have provided a counterfactual analysis helps to demonstrate the effects the coefficients predict in a substantially meaningful manner.

## 3.4 Counterfactual Analysis: Three Precincts

The picture emerging from these results is that certain groups of voters are less likely to have their preferences represented when they vote using the "Votomatic" punchcard voting system. So far we have focused our discussion on the basic direction of the estimated effect of each set of independent variables in the analysis and their statistical significance. In order to gauge the relative magnitude of the effects we estimate in the overvote and undervote models, we undertake a counterfactual analysis. In our counterfactual study, we pick three actual Los Angeles County precincts, depending on their relative rates of over- and undervoting: one with average rates, the others on the high or low end of of the over- and undervoting distributions. Then for each precinct, we perform two counterfactuals. First, we increase the value of each respective independent variables by 5%, and second, we increase the value of each variable by 10%. We produce a predicted number of presidential over- and undervotes for each precinct, for each independent variable, under each scenario. The average precinct is located in the city of Hawthorne (precinct 3178), which is in southern Los Angeles County, near Los Angeles International Airport. Precinct 3178 in Hawthorne had 2 overvotes and 10 undervotes in the 2000 presidential election. The precinct with high rates of over- and undervotes is from Los Angeles (precinct 5033, 16 presidential overvotes and 25 presidential undervotes), and the low precinct is in Glendora (precinct 193, located in the eastern part of Los Angeles County, in the foothills of the San Gabriel Mountains, and did not have any presidential over- or undervotes in the 2000 election). We present the results of this counterfactual analysis in Table 4.

**Table 4 Goes Here**

Table 4 is organized so that the results for the 5% increase scenario are provided in the top panel, and the results for the 10% increase are in the lower panel. Each precinct spans three columns, giving the profile of each precinct, and then the predicted presidential over- and undervote rates. We also provide in the top row of Table 4 the baseline prediction from our negative binomial models for each precinct's over- and undervote rates. When we consider these specific examples, it is clear that the effect of changing the precinct demographics by this magnitude does not have a startling effect on overvotes and undervotes. But when examining these counterfactual predictions, we must emphasize that as the counts of over- and undervotes are quite small to begin with — again, the Hawthorne precinct only had 2 overvotes and 10 undervotes — in the 2000 presidential election — we should not expect large changes based on modest 5% increases in our independent variables. Also, here are focused here on the presidential over- and undervotes; and both were generally low in the presidential contest. Rather, the goal of this counterfactual analysis is to give a realistic prediction for the impact of modest changes in the independent variables on over- and undervote rates.

In our Hawthorne precinct, our model predicts 2 overvotes and 8 undervotes in precinct 3178. Adding 5% to each demographic variable for the Hawthorne precinct serves to decrease the predicted overvotes to 1 for each partisan category. The same changes, though, for the Los Angeles precinct, increases overvotes to 9 in three instances (Hispanic, female, and non-English counterfactuals), and decreases the predicted overvote rate to 7 for the three partisan categories. We predict an increase in the undervote rate for the Los Angeles precinct by 1 undervote for the female counterfactual, and a decrease by 1 undervote in the Republican counterfactual. Last, 5% increases in independent variables in Glendora serve to increase the overvote rate in five instances (Black, Asian-Hawaiian-Pacific Islander, female, nonpartisan, and non-English); we also predict an increase in the undervote rate by 1 undervote in the female counterfactual.

The lower panel of Table 4 gives the results from the 10% counterfactual, and not surprisingly, we do find that the predicted over- and undervotes are greater than in the 5% counterfactual. Beginning with Hawthorne, the "average" precinct, increasing the respective percentages of Democratic, Republican, or nonpartisan voters again decreases the overvote prediction to 1 overvote; this same magnitude change for the percentage of female voters increases the undervote rate to 9. In the Los Angeles precinct we see larger estimated effects, as an increase of 10% leads to a predicted increase of 2 more overvotes in the Hispanic and non-English counterfactuals, an increase of 1 overvote in the Black, Asian-Hawaiian-Pacific Islander, and female counterfactuals; we see a decrease in the predicted overvote rate of 2 overvotes in the Democratic and Republican counterfactuals, and of 1 overvote in the nonpartisan counterfactual. In the Los Angeles precinct we see an increase of 6 predicted undervotes in the Hispanic counterfactual, 2 in the female counterfactual, and a decrease of one overvote in the Republican counterfactual. Last, in Glendora, a 10% increase leads to an increase to 1 overvote in five counterfactuals, and an identical increase in the undervote results for four of the counterfactuals.

Thus, this counterfactual analysis shows that each of the independent variables does impact both over- and undervotes. We generally see that race and gender are leading to higher over- and undervote predictions, while the partisanship estimates point to lower levels of over- and undervotes. But, the counterfactual analysis does paint a nuanced portrait of how over- and undervote dynamics differ across a heterogeneous election jurisdiction; we found the strongest impacts in the counterfactual analysis in the precinct with already high levels of over- and undervoting. Last, when we focus on specific precincts, where the numbers of over- and undervotes are low, we do not see large increases in either rate in either of our counterfactual scenarios. Yet when we place these estimates in the appropriate context, that Los Angeles County has approximately 5,000 precincts, it is clear that even relatively slight changes in precinct profiles can aggregate into relatively large numbers of over- or undervotes countywide.

## 4 Conclusion

Normatively, race and gender should have no significant role in the residual vote. The relatively higher levels of residual votes in Los Angeles County, attributable to the punch card system, magnify the effects of race and gender in voting. While our analysis does not universally demonstrate that race and gender have a role to play in every contest, it is true that in most contests, minorities and women have a greater propensity to over- or undervote. The implication is that the voices of these voting populations may be underrepresented in close elections as they are more likely to over- or undervote.

Future research should focus on the roles of income and education on residual vote. Further, it should examine other counties as well. While Los Angeles is a prime candidate for research as it includes a plethora of social and ethnic representatives, we do note that not all counties have similar demographic compositions. As a consequence, we have composed a similar analysis across counties in California looking at the impact of vote technology and demographic variables on the residual vote (Alvarez, Sinclair and Wilson 2003). This study provides a great deal more variance and permits us to include the impact of income on residual votes, but reaches the same qualitative conclusion about race and voting system use as we found in this paper. Nonwhites have higher residual vote rates, especially when they use punchcard voting systems like "Votomatic".

We hypothesize that voters are more likely to cast a residual vote when their education levels are lower, or they have less familiarity with English. In general, we believe that an overvote is a mistake made by the voter, perhaps due to a misunderstanding about the election process. Also, we think that some undervotes are also mistakes, but that undervotes are also generated by a lack of awareness about specific contests. While our results are consistent with these inferences about the underlying motivations behind both over- and undervotes, our behavioral inferences are indirect and clearly need further examination. The question of whether residual votes are intentional or unintentional awaits future research, perhaps experimental in nature.

Soon some punchcard systems (including the "Votomatic") will be decertified and no longer used in California (Alvarez, Sinclair and Wilson 2003). Also, one likely result from the recent passage of the "Help America Vote Act" should be the effective end of punchcard voting in federal elections in the United States. Whether the transition from punchcard voting to new systems will mark an era wherein the residual vote occurs at random in the population or not is unknown. We hope that the new systems will, at least, reduce the number of uncounted ballots and minimize the impact of race and gender on the residual vote.

# Notes

[1]The terminology "voting system" used here refers to the physical means by which voters indicate their preferences in an election context, and how those are then tabulated. Currently in the United States there are four major types of voting systems. Paper ballots are voting systems where the choices facing voters are printed on a sheet of paper, the voter indicates his or her preferences on the paper, and the ballots are then counted by hand. Lever machines are mechanical devices where voters indicate their choices by pulling selected levers, and the votes are tabulated mechanically via a counter. Punch card voting systems employ computer-readable punch cards, of two basic types: "Votomatic" punch card systems use a card that is inserted into a mechanical holder with a voting booklet, and the voter indicates his or her choices by punching certain holes out; and "Datavote" punch cards, where the choices are printed on the ballot, and voters indicate their preference by again punching out an appropriate location on the card. Last, there are direct recording electronic devices ("DREs") and touchscreen type computerized voting systems; the DREs are essentially computerized lever machines, while touchscreen systems are similar to bank automatic teller machines. For further discussion of the differences between these systems and their use in the United States, see Saltzman (1988) or Caltech/MIT (2001).

[2]For detailed discussion of how these specific voting systems have worked in recent elections, see Alvarez, Sinclair and Wilson's (2003) California's analysis or Quirk et al.'s (2002) study of Illinois.

[3]We define the "Votomatic" punchcard system below in more detail.

[4]Hasen (2001) cites that at least five cases had been filed by the time his paper was published: California (Common Cause v. Jones), Florida (NAACP v. Harris), Illinois (Black v. McGuffage) and Wirth v. ESS, and Georgia (Andrews v. Cox).

[5]Steifbold concludes his study by asserting that "Invalid votes are most conspicuous

among women, older citizens, village and farm dwellers, previous nonvoters, and supporters or former supporters of minor political parties ... They are based at once on the most apathetic segments of the population and ... on political radicals and disaffected persons in outlying regions who have not done well in economic life" (pages 405-406).

[6]In late 2003 and early 2004 Los Angeles County will transition from the Votomatic punchcard voting system to a very similar "Inkavote" system. "Inkavote" uses similar cards to Votomatic, but the difference is that instead of punching out "chads", voters will use a pen to fill small areas in with "Inkavote".

[7]In the 2000 election, 30 California counties used punchcard voting systems, 27 used optical scanning, and 1 used an electronic voting system. Of the 30 punchcard-using counties, 21 used Datavote, 5 used Votomatic, and 4 used Pollstar (similar to Votomatic). Of the optical scanning counties, 11 used centralized scanning and 16 used precinct-based scanning. See Alvarez, Sinclair and Wilson (2003) for further discussion.

[8]The California Statewide Database is run by the Institute for Governmental Studies at the University of California, Berkeley. For more information, see http://swdb.berkeley.edu.

[9]To produce the estimates of the gender profile of registered voters in the precinct the California Statewide Database uses the voter registration rolls, and estimates from the name of each voter their gender. It is possible to do the same type of estimation from the surname of each registered voter to identify some types of ethnic and racial identifications — especially for Latinos and Asians. But the use of surnames to identify Blacks is impossible, and thus we prefer to use the estimates of the Hispanic, Black and Asian populations of each precinct. The U.S. Census Bureau has not released data for lower levels of geographic aggregation in California that would be necessary to produce better estimates, like the citizen voting aged population for each racial and ethnic group.

[10]Comparing the over- and undervote rates of absentee and precinct voters is an important avenue of future research; preliminary analysis (available from the authors upon request) indicates that in this particular election absentee voters typically had lower over-

and undervote rates than precinct voters.

[11]We consider only these four races because they were the four that were on every Los Angeles County ballot. Of course, there were other candidate races on the ballot, for example, congressional, state legislative, and a multitude of local races. Unfortunately, comparing these across precincts in which the races or the dynamics of the races vary is well beyond the scope of this present analysis.

[12]Tomz and Van Houweling (2003) use a form of Goodman's regression in their analysis of Black and White residual vote rates in Louisiana and South Carolina. However, the demographic realities of Los Angeles County are more complicated, with significant populations of Hispanics and Asians, in addition to Blacks and Whites. Just this complexity makes any ecological inference technique more difficult to implement, as we no longer have a relatively simple two-by-two table of ecological behavior to estimate. The computational complexity only gets worse when we consider adding variables for gender and partisanship into the analysis.

[13]One approach, using this data, is to pool observations across the four contests and introduce a variable on the right-hand side of the negative binomial models that measures the number of candidates in each race, and interactions of that measure with other right-hand side variables to determine which voters might be more influenced by the number of candidates on the ballot. For this to work, we need to assume that the parameters are homogenous across the four contests; when we test for that assumption, we find that it does not hold (which will be demonstrated again below when we present the coefficient estimates for the models we estimated for each contest). With heterogeneous coefficients across contests, we have a fundamental indeterminancy in the model, as obviously the number of candidates in each contest is collinear with indicators for each contest. In future research we will revisit this question, perhaps by turning to data from legislative contests, by pooling across elections, or by using data from different levels of aggregation.

[14]The other way we could specify our dependent variables is as percentages; we could

divide the number of overvotes in each precinct by the total ballots cast. As a percentage, though, the dependent variables would still have properties that would render ordinary least squares regression invalid. Instead, an econometric technique like grouped logit might be appropriate if we were to use percentages. We prefer to specify our models in terms of counts, because later in our analysis this will allow for a simple estimation of the overall impact of changes in independent variables on the number of over- or undervotes cast throughout Los Angeles County.

[15]Technically the negative binomial model we use assumes that the data generating process is a gamma mixture of Poission variables, and by estimating a parameter for the shape of the gamma distribution, we are able to determine the amount of overdispersion in the data. We use the nbreg procedure in STATA to estimate the negative binomial models reported below.

[16]Using precinct turnout to control for the relative sizes of precincts is troublesome because of the fact that turnout itself is a function of variables like education and income, which are correlated with many of the independent variables in our model and which we cannot directly measure at the precinct level in Los Angeles County. We have rescaled the "size" variable by multiplying it by 10,000.

[17]The OLS regression results are available from the authors upon request.

# 5 References

Achen, Christopher H. and W. Phillips Shively. 1995. *Cross-Level Inference*. Chicago: The University of Chicago Press.

Alvarez, R. Michael. 1997. *Information and Elections*. Ann Arbor: The University of Michigan Press.

Alvarez, R. Michael, D.E. "Betsy" Sinclair, and Catherine Wilson. 2003. "Counting Ballots and the 2000 Election: What Went Wrong?" In Ann Crigler, Marion Just, and Edward McCaffery, editors, *Rethinking the Vote*, New York: Oxford University Press, 2003.

Brady, Henry E., Justin Buchler, Matt Jarvis, and John McNulty. 2001. "Counting All the Votes: The Performance of Voting Technology in the United States." Working paper, UC Data: Data Archive and Technical Assistance webpage. http://ucdata.berkeley.edu.

Bullock III, Charles S. and Richard E. Dunn. 1996. "Election Roll-Off: A Test of Three Explainations." *Urban Affairs Review*. Vol. 32, No. 1, 71-86.

Caltech/MIT Voting Technology Project. 2001. "Voting: What Is, What Could Be". Caltech/MIT.

Darcy, R. and Anne Schneider. 1989. "Confusing Ballots, Roll-off, and the Black Vote." *Western Political Quarterly*. 42: 347-364.

Delli Carpini, Michael X. and Scott Keeter. 2000. "Gender and Political Knowledge." In

Sue Tolleson-Rinehart and Jyl J. Josephson, *Gender and American Politics*, Armonk, New York: M.E. Sharpe.

Fraser, Jeannette Lynn. 1985. "The Effects of Voting Systems on Voter Participation: Punch Card Voting Systems in Ohio". Ph.D. manuscript, Ohio State University.

Hasen, Richard L. 2001. "Bush v. Gore and the Future of Equal Protection Law in Elections." *Florida State Law Review*, 29, 2, 377-406.

Herron, Michael C. and Jasjeet S. Sekhon. 2003. "Overvoting and Representation: An examination of overvoted presidential ballots in Broward and Miami-Date counties." *Electoral Studies*, forthcoming.

King, Gary. 1998. *Unifying Political Methodology: The Likelihood Theory of Statistical Inference*. Ann Arbor: The University of Michigan Press.

King, Gary. 1997. *A Solution to the Ecological Inference Problem*. Princeton: Princeton University Press.

Knack, Stephen and Martha Kropf. 2003. "Invalidated Ballots in the 1996 Presidential Election: A County-Level Analysis." *Journal of Politics*, forthcoming.

Knack, Stephen and Martha Kropf. 2001. "Roll Off at the Top of the Ballot: Intentional Undervoting in American Presidential Elections." Manuscript, University of Missouri at Kansas City.

Mather, George B. 1964. *Effects of the Use of Voting Machines on Total Votes Cast: Iowa - 1920-1960*. Institute of Public Affairs: The University of Iowa, Iowa City.

Mather, George B. 1986. *Lost Votes: Effects of Methods of Voting on Voter Participation Iowa 1920-1984*. Institute of Public Affairs Division of continuing Education: The University of Iowa, Iowa City.

Montgomery, Michael J. 1985. "Voting Systems and Disenfranchisement". *Election Politics* : 16-19.

Nichols, Stephen M. and Gregory A. Strizek. 1995. "Electronic Voting Machines and Ballot Roll-Off." *American Politics Quarterly*. Vol. 23, No. 3, 300-318.

Posner, Richard A. 2001. *Breaking the Deadlock: The 2000 Election, The Constitution, and the Courts*. Princeton, NJ: Princeton University Press.

Quirk, Paul J., James H. Kuklinski, and Philip Habel, 2002. "The Machinery of Democracy: Voting Systems and Ballot Miscounts in Illinois." University of Illinois, Institute of Government and Public Affairs.

Saltzman, Roy G. 1988. "Accuracy, Integrity, and Security in Computerized Vote-Tallying." Institute for Computer Sciences and Technology, National Bureau of Standards, NBS Special Publication 500-158.

Shocket, Peter A., Neil R. Heighberger, and Clyde Brown. 1992. "The Effect of Voting Technology on Voting Behavior in a Simulated Multi-Candidate City Council Election: A

Political Experiment of Ballot Transparency." *Western Political Quarterly*, 45(2), 521-37.

Stiefbold, Rodney P. 1965. "The Significance of Void Ballots in West German Elections." *American Political Science Review* 59,2: 391-407.

Thomas, Norman C. 1968. "Voting Machines and Voter Participation in Four Michigan Constitutional Revision Referenda". *Western Political Quarterly*. 21(3), 409-19.

Tomz, Michael and Robert P. Van Houweling. 2003. "How Does Voting Equipment Affect the Race Gap in Voided Ballots?" *American Journal of Political Science* 47,1: 46-60.

Walker, Jack L. 1966. "Ballot Forms and Voter Fatigue: An Analysis of the Office Block and Party Column Ballots." *Midwest Journal of Political Science*, Volume 10, Issue 4, 448-463.

White, John P. 1960. "Voting Machines and the 1958 Defeat of Constitutional Revision in Michigan." Institute of Public Administration at the University of Michigan, Papers in Public Administration No. 35.

Table 1: **Overvotes and Undervotes**

| Contest | Overvote Mean | Percent | Undervote Mean | Percent |
|---|---|---|---|---|
| President | 2.40 | 0.55 | 9.79 | 2.23 |
| Senate | 2.81 | 0.64 | 22.32 | 5.04 |
| District Attorney | 0.46 | 0.20 | 85.37 | 19.27 |
| Assessor | 4.27 | 0.96 | 101.81 | 22.98 |
| Measure A | 0.52 | 0.12 | 62.13 | 14.02 |
| Measure 32 | 0.52 | 0.12 | 48.62 | 10.98 |
| Measure 33 | 0.51 | 0.12 | 60.53 | 13.66 |
| Measure 34 | 0.79 | 0.18 | 57.57 | 13.00 |
| Measure 35 | 1.08 | 0.24 | 54.35 | 12.27 |
| Measure 36 | 1.16 | 0.26 | 41.46 | 9.36 |
| Measure 37 | 0.85 | 0.19 | 65.47 | 14.78 |
| Measure 38 | 0.79 | 0.18 | 27.15 | 6.13 |
| Measure 39 | 0.84 | 0.19 | 40.95 | 9.24 |

Table 2: **Negative Binomial Analysis, Overvotes**

| | President | | Senate | | District Attorney | | Assessor | |
|---|---|---|---|---|---|---|---|---|
| Variable | IRR | z | IRR | z | IRR | z | IRR | z |
| Size | 1.7980* | 22.15 | 1.794* | 23.85 | 1.7536* | 10.87 | 1.7046* | 28.18 |
| | (0.0476) | | (0.0440) | | (0.0906) | | (0.0323) | |
| Hispanic | 8.8022* | 30.03 | 5.6829* | 27.36 | 3.9941* | 10.27 | 4.9589* | 31.72 |
| | (0.6375) | | (0.3609) | | (0.5385) | | (0.2503) | |
| Black | 3.4308* | 11.15 | 1.8499* | 6.14 | 2.2425* | 3.91 | 3.1847* | 15.02 |
| | (0.3792) | | (0.1855) | | (0.4630) | | (0.2456) | |
| Asian/Hawaii | 2.9754* | 10.34 | 1.8729* | 6.43 | 2.0701* | 3.46 | 2.5403* | 12.50 |
| /Pacific | (0.3138) | | (0.1855) | | (0.4354) | | 0.1894 | |
| Female | 4.9126* | 4.52 | 3.884* | 4.22 | 5.6569* | 2.55 | 2.0248* | 2.81 |
| | (1.7306) | | (1.2498) | | (3.8378) | | (0.5081) | |
| Democratic | 0.0915* | -9.90 | 0.1921* | -7.51 | 0.0335 | -7.35 | 0.4439* | -4.77 |
| | (0.0221) | | (0.0422) | | (0.0155) | | (0.0755) | |
| Republican | 0.0331 | -15.06 | 0.0275* | -17.29 | 0.0072* | -11.29 | 0.2616* | -8.46 |
| | (0.0075) | | (0.0057) | | (0.0032) | | (0.0415) | |
| Nonpartisan | 0.1699* | -4.76 | 0.8871 | -0.36 | 0.0538* | -4.05 | 0.3507* | -3.94 |
| | (0.0632) | | (0.2978) | | (0.0388) | | (0.0933) | |
| Nonenglish | 6.4285* | 5.95 | 4.8819* | 5.39 | 0.6142 | -0.75 | 3.6802* | 5.71 |
| | (2.009) | | (1.436) | | (0.4015) | | (0.8403) | |
| $\alpha$ | 0.1475 | | 0.1366 | | 0.1444 | | 0.0757 | |
| | (0.0112) | | (0.0744) | | (0.0447) | | (0.0064) | |
| Wald | 12147.96 | | 16574.25 | | 1518.95 | | 40224.14 | |

Note: All variables were converted into percentages before running regressions. Also, * indicates a 95% confidence interval and ** indicates a 90% confidence interval.

Table 3: **Negative Binomial Analysis, Undervotes**

| | President | | Senate | | District Attorney | | Assessor | |
|---|---|---|---|---|---|---|---|---|
| Variable | IRR | z | IRR | z | IRR | z | IRR | z |
| Size | 1.8302* | 44.03 | 1.8745* | 64.64 | 1.8238* | 87.77 | 1.8648* | 112.03 |
| | (0.0251) | | (0.0182) | | (0.0125) | | (0.0104) | |
| Hispanic | 1.8421* | 17.65 | 1.4560* | 15.20 | 1.2871* | 14.38 | 0.4126* | -62.70 |
| | (0.0638) | | (0.0360) | | (0.0226) | | (0.0058) | |
| Black | 1.4321* | 6.58 | 1.2921 | 6.48 | 1.8166* | 21.67 | 0.5615 | -25.59 |
| | (0.0782) | | (0.0511) | | (0.0500) | | (0.0127) | |
| Asian/Hawaii | 1.6314* | 9.15 | 1.8170* | 16.32 | 1.2670* | 8.85 | 0.5615* | -26.96 |
| /Pacific | (0.0873) | | (0.0665) | | (0.0339) | | (0.0121) | |
| Female | 4.0382* | 7.67 | 5.3232* | 12.97 | 5.3178* | 18.93 | 2.2740* | 11.33 |
| | (0.7348) | | (0.6863) | | (0.4693) | | (0.1649) | |
| Democratic | 1.0266 | 0.22 | 1.2379* | 2.48 | 7.7696* | 34.70 | 24.1852* | 66.45 |
| | (0.1250) | | (0.1068) | | (0.04584) | | (1.1595) | |
| Republican | 0.7433* | -2.62 | 1.7336* | 6.92 | 6.6414* | 34.70 | 15.63* | 61.89 |
| | (0.0840) | | (0.1378) | | (0.3623) | | (0.6945) | |
| Nonpartisan | 1.0950 | 0.47 | 12.4180* | 19.01 | 29.15* | 36.35 | 263.07* | 74.19 |
| | (0.2093) | | (1.6453) | | (2.7052) | | (19.76) | |
| Nonenglish | 1.7654* | 3.24 | 1.8859* | 5.09 | 1.1077 | 1.15 | 0.8359* | -2.42 |
| | (0.3101) | | (0.2349) | | (0.0983) | | (0.0618) | |
| $\alpha$ | 0.0701 | | 0.0420 | | 0.0330 | | 0.0193 | |
| | (0.0035) | | (0.0018) | | (0.0010) | | (0.0006) | |
| Wald | 153152.34 | | 553298.36 | | 2182983.76 | | 3568426.97 | |

Note: All variables were converted into percentages before running regressions. Also, * indicates a 95% confidence interval and ** indicates a 90% confidence interval.

Table 4: **Precinct-Specific Counterfactuals, Presidential Contest**

| | Hawthorne: Average | | | Los Angeles: High | | | Glendora: Low | | |
|---|---|---|---|---|---|---|---|---|---|
| Community | Profile | Over | Under | Profile | Over | Under | Profile | Over | Under |
| Baseline | | 2 | 8 | | 8 | 13 | | 0 | 4 |
| 5% Increase Scenario | | | | | | | | | |
| Hispanic | 25.5 | 2 | 8 | 82.6 | 9 | 13 | 15.1 | 0 | 4 |
| Black | 9.2 | 2 | 8 | 1.5 | 8 | 13 | 2.6 | 1 | 4 |
| Asian/Hawaiian /Pacific | 19.6 | 2 | 8 | 13.8 | 8 | 13 | 8.6 | 1 | 4 |
| Female | 50.8 | 2 | 8 | 55.8 | 9 | 14 | 59.1 | 1 | 5 |
| Democratic | 49.2 | 1 | 8 | 68.1 | 7 | 13 | 28.8 | 0 | 4 |
| Republican | 33.0 | 1 | 8 | 14.7 | 7 | 12 | 53.0 | 0 | 4 |
| Nonpartisan | 13.5 | 1 | 8 | 15.0 | 7 | 13 | 13.1 | 1 | 4 |
| Nonenglish | 12.6 | 2 | 8 | 9.0 | 9 | 13 | 0.3 | 1 | 4 |
| 10% Increase Scenario | | | | | | | | | |
| Hispanic | 25.5 | 2 | 8 | 82.6 | 10 | 19 | 15.1 | 1 | 5 |
| Black | 9.2 | 2 | 8 | 1.5 | 9 | 13 | 2.6 | 1 | 4 |
| Asian/Hawaiian /Pacific | 19.6 | 2 | 8 | 13.8 | 9 | 13 | 8.6 | 1 | 5 |
| Female | 50.8 | 2 | 9 | 55.8 | 9 | 14 | 59.1 | 1 | 5 |
| Democratic | 49.2 | 1 | 8 | 68.1 | 6 | 13 | 28.8 | 0 | 4 |
| Republican | 33.0 | 1 | 8 | 14.7 | 6 | 12 | 53.0 | 0 | 4 |
| Nonpartisan | 13.5 | 1 | 8 | 15.0 | 7 | 13 | 13.1 | 0 | 4 |
| Nonenglish | 12.6 | 2 | 8 | 9.0 | 10 | 13 | 0.3 | 1 | 5 |