選舉研究
第二期（83/11），pp.221-250

Selection Bias in Vote Choice Models

Shing-Yuan Sheng*

Introduction

The determinants of vote choice are of obvious interest to political scientists. A great number of vote choice models try to explain what factors determine people's vote choice. When estimating the models, most researchers include only the respondents who vote and give a clear answer in the survey question. People who do not vote, or do not give clear answers about their vote choices, are excluded when estimating the model. However, based on research findings about American voters, we have reason to suspect the representativeness of this voter subsample. It is quite possible that the voter subsample is biased toward higher educated, wealthier, and older people, and those interested in politics. Worse, those people are happened to be more likely to vote for the Republican party. Therefore, researchers who use only the voter subsample are likely to obtain inconsistent estimates.

The major purpose of this paper is to assess the sensitivities of vote choice models to selection bias resulting from excluding nonvoters. To do this, I will apply a bivariate probit selection model--developed by Dubin and Rivers (1989)-- to three major vote choice models. I rely upon data from the 1988 and 1990 American National Election Studies. This analytical strategy has two characteristics: First, I include data in both presidential election and mid-term election years, so that I can compare selection bias in different elections and in different years. Furthermore, because the turnout rates are quite different between presidential election years and mid-term election years, including both the 1 988 and 1990 data can help in comparing the susceptibilities to selection bias in different degrees of censoring cases. Second, I include three vote choice models, so that I can evaluate the sensitivities of different models to selection bias.

* Department of Political Science, National Chengchi University

In the following sections, I first describe the problem of selection bias in the vote choice model, and come up with an approach to deal with that bias. Second, I identify the causes which determine individuals' likelihood to vote or not. Then, I estimate three kinds of vote choice models with both uncorrected and corrected selection bias and assess their sensitivities to selection bias. Finally, research findings will show that all the three models are subject to selection bias. In addition, research findings show that the seriousness of selection bias is not necessarily related to turnout rate in the particular election year. Rather, it is more likely to be related to the specification of the model. Based on the findings, I suggest that researchers should deal with selection bias when estimating vote choice models.

Approach to Estimate the Selection Bias in Vote Choice Models

Social scientists are usually interested in a large population. In a vote choice model, for example, researchers are interested in vote choices of the whole voting age population. However, it is impossible for researchers to interview every eligible voter. Instead, they interview only a small part of the population, such as 2000 subjects in the population. Then, they infer the results obtained from the sample to the population. A representative sample is the crucial precondition for making effective inferences. However, in some situations, respondents act in a way that makes it impossible get measurements on certain variables. In the vote choice case, researchers cannot get data on vote choice of some respondents merely because these respondents do not vote. If the selection process is completely random; that is, voters and nonvoters have the same characteristics except for voting or not voting, then we do not worry too much about selection bias even when we exclude nonvoters. However, it is obvious that the selection process of voting or not voting is not random. Rather, much evidence has shown that people's choice of voting or not voting is determined by a series of factors, such as socioeconomic status, demographic backgrounds, political preferences and related political attitudes. Voters are skewed toward being more educated, wealthier, older, and more interested in politics (Campbell, et.al. 1960; Verba and Nie, 1972; Wolfinger and Rosenstone, 1980; Rosenstone and Hansen, 1993). Furthermore, the factors determining people's voter turnout also determine their vote choices. There is evidence that people with higher socioeconomic status are more likely to vote, and more likely to vote for the Republican party. Thus, voters and nonvoters may have different preferences on their choices. Therefore, researchers aiming at the whole population but using only the voter subsample (uncensored sample) when estimating the vote choice model may get inconsistent estimates.

One solution would be to ask nonvoters which candidate they would have voted for. However, this solution is not practical because researchers usually cannot control the data collection process (Dubin and Rivers, 1989). Meanwhile, even though researchers may be able to control the data collection process, they do not use this approach. This is because nonvoters may give quick answers with little thought about their vote choice. → 拿到data 先算过 不可行

On the other hand, weighting--a traditional way to deal with missing data--is not applicable to this case. Researchers usually weight according to some variables which is indicated important by research based on the voter subsample. However, if nonvoters are different from voters at the beginning, then weighting based on results obtained from voters may make the situation worse. → weight 不可行 反可能造成的問題.

As described above, neither asking nonvoters about their vote choices nor weighting can effectively deal with the selection bias problem emerging from a great part of the sample's nonvoting. Furthermore, the usual Heckman's two-stage approach (1979) which works for continuous dependent variable is not applicable to this particular case because the dependent variable in vote choice model is dichotomous. Therefore, I use Dubin and Rivers' bivariate normal selection model, which is developed from Heckman's model and deals with dichotomous dependent variable in outcome equation.

To describe Dubin and Rivers' approach, I will start from Heckman's approach. Heckman lays out a two-stage approach to modeling a censored sample. The first step is to modeling the selection equation; in other words, to identify the causes (X_{2i}) which affect selection process. The dependent variable (Y_{2i}) in the selection equation is 1 if the sample is uncensored (observed), 0 if the sample is censored (unobserved). Since the dependent variable is dichotomous, Heckman suggests a probit analysis to estimate the selection equation. Then, he generates an Inverse Mill's Ratio (IMR) of the predicted values from the selection equation. That is,

$$IMR= \phi(\beta_2'X_2)/\ \Phi(\beta_2'X_2)$$

Where ϕ is the normal probability density function and Φ is the cumulative normal probability function. In fact, the IMR is the probability that an individual is selected into the observed sample.

The second step is to include the IMR as an additional regressor in the outcome equation. Heckman suggests a linear regression to estimate the outcome equation. That is,

$$Y_{1i}=\beta_1'X_1+\rho IMR(\beta_2'X_2)$$

The coefficient on the IMR (ρ) is an estimate of the correlation between the error

大小

terms of the selection equation and the outcome equation. The magnitude of the coefficient shows the size of the selection bias.

Obviously, if the dependent variable of the outcome equation is dichotomous, then Heckman's approach is not applicable. That is because when the dependent variable is dichotomous, the error terms will not be normally distributed. Then, the basic assumption of a joint normal distribution of the error terms will be violated, and hence the estimates will be inconsistent (Achen, 1986). To deal with this problem, Dubin and Rivers developed a solution for modeling a dichotomous dependent variable in outcome equation. Like Heckman, Dubin and Rivers modeling selection equation and outcome equation. However, unlike Heckman's two-stage estimation, they estimate the selection equation and outcome equation simultaneously by way of the Maximum Likelihood Estimation. (Please see detailed description of Dubin and Rivers' approach in Appendix A.) The two-stage estimation in the vote choice model implies that an individual decides to vote first and then decides for whom to vote. However, if an individual has a candidate preference, this may push him/her to vote. Therefore, vote and vote choice may simultaneously affect each other. Therefore, Dubin and Rivers' one-stage approach is better than two-stage estimation for vote choice model.

Besides Dubin and Rivers, Achen also develops one way to deal with dichotomous dependent variable in outcome equation. Like Heckman, Achen advocates a two-stage approach. Specifically, modeling the selection equation first, and put the expected value of the error term as an additional regressor when estimating the outcome equation. However, unlike Heckman, Achen suggests a linear probability model to estimate the selection equation, and recode the expected value to a 0-1 bound, that is λ. Then, he adds λ as an additional regressor to the outcome equation, and estimates the outcome equation by way of a nonlinear least squares. In Achen's approach, one needs to adjust the expected value to be a 0-1 bound. This may cause some distortion 扭曲 of the data. Also, one needs to adjust the error terms obtained from the second stage. On the other hand, Dubin and Rivers' approach can estimate the outcome and selection equation simultaneously and use the information efficiently. Therefore, Dubin and Rivers' approach is better than Achen's for the case of a dichotomous dependent variable.

Selection Mechanism of Voter Turnout

As described above, the first step in dealing with the selection bias is to identify causes affecting the selection process. In the vote choice case, the selection mechanism is a series of factors which determine individuals' voter turnout. A great deal of research has built a theoretical and empirical basis for the determinants of voter turnout. This helps me estimate the selection mechanism of voter turnout. Here, I do not repeat those theoretical arguments, but report the research findings related to my research.

A great deal of research indicates that individuals' sociological status (especially education) substantially affect their voting. The more educated and the wealthier people are more easily afford to vote (Campbell, et al. 1960; Verba and Nie, 1972; Wolfinger and Rosenstone, 1980; Rosenstone and Hansen, 1993). Age is also a substantial factor of voting. The younger people are less likely to vote than the older (Campbell, et al. 1960; Wolfinger and Rosenstone, 1980).

The effect of race on voting is not so clear. Some earlier research argue that blacks are less likely to vote (Campbell, et al. 1960). However, certain later research argues that blacks' voter turnout is not less than that of whites, when controlling for education and income (Wolfinger and Rosenstone, 1980). To test this proposition, I still put the race variable into the model.

In addition, research evidence has shown married people are more likely to vote than others (Campbell, 1960; Wolfinger and Rosenstone, 1980). This is because people are easily affected by interpersonal influence. A spouse has a substantial influence on voting, especially for people with lower education and lower political interests (Wolfinger and Rosenstone, 1980). In addition, there is evidence that new residents are less likely to vote than others. That is because people need to register first to be able to vote. New residents need time to adapt to new environments and to register. Therefore, new residents are less likely to vote (Square, Wolfinger and Glass, 1987).

People's political interests also affect their likelihood of voting. This relation is held even when controlling for related background characteristics. People who are more interested in politics are more likely to be interested in campaign, care about the election outcomes, and expose themselves to the mass media to get related information. All of these factors may cause them to be exposed to the mobilization of the mass media, parties, and candidates. These people are then more likely to vote (Campbell et al, 1960; Rosenstone and Hansen, 1993). In addition, much research has shown that people with higher sense of political efficacy will participate more in politics because they feel that they are able to have impacts on government and that government will respond to them (Campbell et. al., 1960; Finkel, 1985; Gant and Luttbeg, 1991; Rosenstone and Hansen, 1993).

Furthermore, people's political orientations, such as strength of partisanship and affect for candidates, also affect their likelihood of voting. Stronger partisanship and stronger likes or dislikes about a candidate will drive people to be more concerned about the election outcomes, and hence more likely to vote (Rosenstone and Hansen, 1993).

Having identified the causes affecting the selection mechanism, the next step is to estimate the selection equation. The dependent variable of selection equation is dichotomous, with 1 if voting, and 0 if not voting.[1] Therefore, I estimate the selection equation by way of a probit analysis. Table 1 shows the probit estimates of the selection equation.

Table 1 shows that most variables have expected relationships. Although some of the magnitudes of the effects differ across election years, these differences are reasonable. Specifically, the socioeconomic status characteristics --education and income, play more important roles in the presidential election year than in the congressional election year. On the other hand, interests in campaign, affect for candidate, and new resident play more important roles in the mid-term election year. These differences reflect Americans' different levels of interests in the general and in the mid-term elections. Because people are less interested in the mid-term elections, special forces are important in mobilizing people to vote. Therefore, if people are more interested in elections, or they have special likes or dislikes about a candidate, they are more likely to vote. Meanwhile, because people are less interested in mid-term elections, they are less likely to register if they move to a new place. But, if they are really interested in the elections, they may register

Table 1: Probit Estimates of Selection Equation

	總統 1988			1990			期中.
	Model I	Model II	✓ Model III	Model I	Model II	Model III	✓ Model IV
Education	1.02 (.15)	1.02 (.15)	.98 (.15)	.76 (.14)	.73 (.14)	.74 (.14)	.73 (.14)
Income	.79 (.17)	.81 (.17)	.92 (.15)	.46 (.15)	.46 (.15)	.46 (.15)	.50 (.14)
Age	1.25 (.20)	1.27 (.20)	1.29 (.20)	1.26 (.17)	1.28 (.17)	1.29 (.17)	1.28 (.17)
New Resident	-.28 (.13)	-.27 (.13)	-.28 (.13)	-.71 (.15)	-.71 (.15)	-.71 (.15)	-.71 (.15)
Care about Elections	.25 (.08)	.25 (.08)	.25 (.08)	.40 (.08)	.39 (.07)	.39 (.07)	.39 (.07)
Interest in Campaign	.78 (.13)	.77 (.13)	.78 (.13)	1.00 (.12)	.99 (.11)	.98 (.11)	.98 (.11)
Newspaper Usage	.42 (.08)	.42 (.08)	.42 (.08)	.40 (.08)	.40 (.08)	.40 (.08)	.40 (.08)
Strength of Partisanship	.67 (.13)	.66 (.12)	.67 (.12)	.41 (.11)	.41 (.11)	.39 (.11)	.39 (.11)
Affect for Candidate	.42 (.18)	.43 (.18)	.45 (.18)	.77 (.27)	.77 (.27)	.79 (.27)	.79 (.27)
Political Efficacy	.45 (.13)	.45 (.13)	.44 (.13)	-.12 (.12)	——	——	——
Race (Black)	-.08 (.12)	——	——	-.11 (.10)	-.10 (.10)	——	——
Married	.13 (.09)	.14 (.09)	——	.03 (.08)	.04 (.08)	.04 (.08)	——
Constant	-2.22 (.16)	-2.25 (.16)	-2.21 (.15)	-2.30 (.13)	-2.33 (.13)	-2.35 (.13)	-2.34 (.13)
N	1,595	1,595	1,595	1,745	1,750	1,750	1,750
Correct Predicted (%)	80.25	80.19	79.62	74.61	74.29	74.29	74.46

Note: 1. All variables are scaled from 0-1 interval. Measurement of the variables are described in Appendix B.
2. The dependent variable is voter turnout, with 1 if voting and 0 if not voting.
3. The entries are probit coefficients (with standard errors in parentheses.)

and vote the same as old residents. Therefore, the variable of new resident is more important in the mid-term election year than in a presidential year.

As shown in Table 1, race (black) and marriage status (married) have expected direction though they have large standard errors compared to their coefficients both in 1988 and 1990. The relatively small coefficient on the race variable does not contrast to recent research (Wolfinger and Rosenstone, 1980). That is, blacks' voter turnout is no less than that of whites when controlling for education, income, and relative political attitudes.

The standard error of political efficacy in 1990 is large relative to its coefficient. Worse, the direction of the coefficient is negative, which contrasts to theoretical expectations. Comparing Model I and II in 1990, we may find that the estimates of other variables are the same with or without political efficacy. Therefore, I decide to delete political efficacy from the 1990 selection equation.

Based on the results of Model I, I estimate two alternative models in 1988 and three alternative models in 1990. As shown in Table 1, the estimates are quite stable across different specifications of the equations. This demonstrates that the estimated relationships between the independent variables and the voter turnout are quite stable. More noteworthy, the effects of the relationships are as expected. Specifically, socio-demographic characteristics (education, income and age), political interests (care about election outcomes, interests in campaign, and newspaper usage), and political orientation (strength of partisanship, affect for candidate, and political efficacy), are positively associated with the likelihood of voting. On the other hand, the variable "new resident" is negatively associated with voting.

The models can correctly predict respondents' turnout about 80 percent in 1988 and 75 percent in 1990. Although these percentages are not particularly high, they are actually better than usual turnout models (Rosenstone and Hansen, 1993). Model 3 in 1988 and model 4 in 1990 are the selection models I will use in later analysis.

Selection Bias in Vote Choice Models

In this section, I assess three kinds of vote choice models as to their susceptibilities to selection bias. These models include the sociological model, the social psychological model, and Jacobson's Congressional vote choice model. These models are chosen because they are able to explain voters' choices in some way. They are also quite different, and offer me a good opportunity to examine the selection bias in a variety of vote choice models.

Sociological Model

Individuals' socio-demographic characteristics have long been considered as substantial determinants of vote choice. The early research on American voting behavior has demonstrated a strong correlation between party support and socio-demographic factors (Lazarsfeld, et al. 1944). Socioeconomic status is obviously a significant determinant of individuals' political opinions and behavior. Much research combines education, income and occupation as a single indicator of socioeconomic status. However, since these three variables may have different impacts on vote choice, I place them separately into the model.[2] The proposition that the workers and lower-income people are more likely to support the Democratic party in order to become economically better off, while the middle class and higher-income groups are more likely to support the Republican party to maintain their economic advantage has been theoretically discussed and verified by evidence (Lazarsfeld, et al, 1944; Lipset, 1981). Education equips people with better ability and more opportunity to improve their socioeconomic status. This may drive more highly educated people to be more supportive of the Republican party. However, education also leads people to be more liberal on social issues (Knoke, 1979). This may drive the more highly educated people to be more supportive of the Democratic party. How these two opposite forces affect people's vote choice may depend on the individuals' other characteristics and political attitudes.

In addition, much research evidence shows that union members are more supportive of the Democratic party (Campbell, et al. 1960). This tendency does not disappear even when controlling for socioeconomic status and political attitudes (Dubin and Rivers, 1989). Therefore, I put the variable union member to the model.

Different age groups may have different party support. In general, the young tend to be more idealistic and more eager to change the status quo. Therefore, the young people are more likely to support the more liberal party-- the Democratic party; and the elderly, more likely to support the more conservative party-- the Republican party (Lipset, 1981).

Race is also an important determinant of party choice. Studies of American voters show that blacks as a group strongly support the Democratic party. In addition, religion has long been viewed as an important determinant of party support. Lazarsfeld and his associates (1944) claim that people's religion as one of the most important three factors in determining their vote choice.

Place of residence of individuals is also an important determinant of vote choice. The typical case is people who lived in southern United States strongly supported the Democratic party before the 1960s (Key, 1949; Lazarsfeld, et. al., 1944). In addition, whether an individual lives in a urban or rural area also determines his vote choice. Evidence has shown that people who live in rural areas are more likely to vote for the Republican party (Lazarsfeld, et al, 1944).

Table 2 shows the results of corrected and uncorrected estimates of the socio- demographic factors on people's vote choice. The corrected equations are estimated using

Table 2: Sociological Models with Corrected and Uncorrected Estimates

	1988 Presidential Voting		1988 Congressional Voting		1990 Congressional Voting	
	Corrected	Uncorrected	Corrected	Uncorrected	Corrected	Uncorrected
Outcome Equation						
Education	.32 (.18)	.38 (.15)	.33 (.16)	.55 (.16)	.10 (.14)	.12 (.19)
Income	.62 (.19)	.64 (.18)	.40 (.19)	.56 (.18)	1.21 (.23)	1.23 (.23)
Worker	-.23 (.10)	-.22 (.10)	-.10 (.02)	-.09 (.10)	-.21 (.02)	-.19 (.12)
Union Member	-.41 (.10)	-.41 (.10)	-.34 (.10)	-.33 (.11)	-.22 (.12)	-.23 (.12)
Age	.12 (.30)	.15 (.20)	.26 (.22)	.54 (.20)	.32 (.27)	.39 (.24)
Black	-1.63 (.20)	-1.65 (.18)	-1.08 (.21)	-1.14 (.22)	-.97 (.22)	-.96 (.22)
South	.001(.15)	.02 (.10)	-.63 (.10)	-.63 (.10)	-.17 (.12)	-.15 (.12)
Rural	.04 (.13)	.03 (.09)	.35 (.02)	.36 (.09)	.32 (.02)	.32 (.11)
Protestant	.44 (.09)	.45 (.09)	.43 (.09)	.46 (.09)	.31 (.10)	.29 (.10)
Constant	-.48 (.26)	-.59 (.17)	-.79 (.24)	-1.30 (.17)	-1.34 (.28)	-1.46 (.22)
ρ	-.12 (.15)		-.43 (.14)		-.10 (.13)	
Selection Equation						
Education	.98 (.15)		.98 (.15)		.73 (.14)	
Income	.92 (.16)		.94 (.15)		.49 (.14)	
Age	1.29 (.20)		1.31 (.20)		1.28 (.16)	
New Resident	-.28 (.13)		-.27 (.13)		-.72 (.15)	
Care about Elections	.25 (.08)		.24 (.08)		.39 (.07)	
Interest in Campaign	.77 (.13)		.74 (.13)		.99 (.12)	
Newspaper Usage	.41 (.08)		.40 (.08)		.41 (.08)	
Strength of Partisanship	.66 (.12)		.69 (.12)		.39 (.11)	
Affect for Candidates	.45 (.45)		.48 (.18)		.78 (.28)	
Political Efficacy	.46 (.14)		.45 (.13)		———	
Constant	-2.21 (.15)		-2.23 (.15)		-2.34 (.12)	
N	1,595	1,117	1,595	1,117	1,750	816

Note: 1. All variables are scaled from 0 to 1, with 0 the lowest and 1 the highest.

2. The dependent variable of outcome equation is 1 if voting for the Republican candidate(s), 0 otherwise. The dependent variable of selection equation is 1 if voting, and 0 if not voting.

3. The entries are probit coefficients (with standard errors in parentheses).

4. The Ns in the corrected models are larger than those in the uncorrected models because of the following reason. I include voters and nonvoters when estimating the correted models by way of Dubin and Rivers' approach. What I need from nonvoters is the information about the probability of voting (Please see Appendix A). Therefore, even though nonvoters have missing values in the outcome equation variables, they can be induded. However, I include only those who vote and have non-missing values in the outcome equation when estimating the uncorrected models.

voters and nonvoters, while the uncorrected equations are estimated using voters only. The dependent variable of the outcome equation is dichotomous, with 1 for voting for the Republican party and 0 otherwise. Therefore, the uncorrected model is estimated by way of a probit analysis, while the corrected model is estimated by way of the Dubin and Rivers' bivariate normal selection model described earlier.

The directions and magnitudes of the estimates shown in Table 2 are as theoretically expected. Higher educated, wealthier, older people, and Protestants are more likely to vote for the Republican party. On the other hand, blacks, workers, and members of an union are less likely to vote Republican. Although the two variables "live in southern area" and "live in a rural areas" show relatively unstable estimates across elections and years, they have effects in consistent directions. Furthermore, comparing estimates of the selection equation shown in Table 2 with those in Table 1, we may find exactly the same estimates for the same variable. This means not only that the selection equation is well specified, but also that the method for correcting the selection bias is quite reliable.

The coefficients of ρ (correction of error terms of the selection equation and the outcome equation) in the 1988 presidential voting model and the 1990 congressional voting model are small relative to their standard errors (β = -.12 and SE= .15 in 1988, β = -.10, and SE=.13 in 1990). Also, comparing the corrected and uncorrected estimates across variables in these two models, I find that there is no substantial difference in the corrected and uncorrected estimates for all variables. Therefore, selection bias may not be serious for the sociological models in the 1988 presidentail voting and the 1990 congressional voting.

Comparing the corrected and uncorrected estimates in the three models, I find obvious differences occurring in three independent variables-- education, income and age. While the magnitudes of the differences may be different across elections and across years, the patterns are consistent--all of the estimates decreased when correcting for selection bias. That is, the estimates obtained from only voters overestimate the impacts of the three variables-- education, income, and age-- on vote choice. The differences between the corrected and uncorrected estimates range from .02 to .22 on the education variable; from .02 to .16 on the income variable; and from .03 to .28 on the age variable. All of the largest differences appear in the 1988 congressional voting, where the coefficient of ρ is also the largest. This shows that when the coefficient of ρ is larger, so is the bias of the model.

Remember these three variables also happen to appear in the selection equation and have substantial effects on the selection process. This shows that when an independent variable affects the dependent variable in the model of interest, and also substantially affects the selection process, it may be more susceptible to selection bias than are other variables.)

Comparing the 1988 congressional model with the same model in 1990, one may find that the former is much more susceptible to selection bias. This finding tells us that the seriousness of the selection bias in a vote choice model is not necessarily related to the turnout rate. Remember that 1988 is a presidential election year and thus has a higher turnout rate; but congressional voting model in 1988 is much more susceptible to the selection bias. Rather, the seriousness of selection bias is relatively determined by the sizes of the effects of the independent variables (in the outcome equation) which affect the selection process. If the independent variables in the outcome equation also affect the selection equation, then the larger the effects on the selection, the more serious the selection bias. From the comparison of the selection mechanisms in 1988 and 1990 described earlier, we know that education and income substantially affect people's voting in 1988, but not as much in 1990. Since education and income substantially affect the voter turnout in 1988, the vote choice model in 1988 is susceptible to selection bias.

Whether the corrected model is better than the uncorrected model? If we knew the "true" model or not, we could answer this question by comparing the corrected and uncorrected estimates with the true values. However, we do not know the true model. Therefore, the criteria of evaluation should be whether or not the corrections cause estimates to change in a sensible direction; and whether they cause a different conclusion about the relationship we are interested. To evaluate whether the changes of estimates make sense, Greene's illustration provides a good clue (Greene, 1993: 708-710). Greene indicates that the estimate of an independent variable (which appears on the outcome and also selection equations) based on an uncensored sample has two parts: One is the estimate due to its influence on the dependent variable of the outcome equation; the other, due to its influence on the probability of being observed. If the ρ is positive, and $E(y_{1i})$ is greater when it is uncensored, then researchers base on the uncensored sample will get downward biased coefficient.[3] While if the ρ is negative and $E(y_{1i})$ is greater when it is uncensored, then researchers may get upward biased coefficients. Obviously, the bias of my model is of the second situation. Take income as an example. The effect of income on vote choice based on uncensored sample (voters) has two parts. One part is due to its influence in increasing the probability of an individual's voter turnout; the other is due to its influence on increasing the probability of voting for the Republican party. Therefore, estimates based on a voter subsample may overestimate the effect of income on vote choice.

Overall, the sociological model is subject to the selection bias, especially for the 1988 congressional voting model. Education, income, and age three variables that affect people's vote choice and also determine whether they vote, are fairly susceptible to the selection bias.

Social Psychological Model

Even though the sociological model to some degree explains people's vote choice, it is neither able to fairly capture individuals' attitudes toward political objects, nor account for the radical swings between elections (Kinder and Sears, 1985). Therefore, the efforts at explaining individuals' voter behavior shift from the individuals' socio-demographic characteristics to their social psychological features. In these efforts, the vote choice model of the Michigan school is obviously the most important one. In *The American Voter* (1960), Campbell and his associates indicate that individuals' vote choice is determined by a series of factors from distal environmental factors and individuals' socio-demographic characteristics to immediate individuals' psychological attitudes-- party identification, issue preferences, and candidate evaluations. Since this influential work was published, these three immediate factors have been cited repeatedly as the most important determinants of vote choice.

In this paper, I estimate the impacts of the three factors on individuals' vote choice, and assess their sensitivities to the selection bias. Because these three variables are highly correlated with each other theoretically and empirically, a single regression equation cannot really tell the relative importance of these variables. Therefore, a better way to examine their impacts on vote choice may be by a simultaneous equation model (Jackson, 1975 ; Markus and Converse, 1979; Page and Jones, 1979; Asher, 1983). However, such a model may complicate the assessment of the effects of selection bias. To simplify the issue, I estimated a single outcome equation and focused on comparing the differences in corrected and uncorrected estimates.

In addition to the three variables, evaluations of government performance, especially in the economy, are also an important determinant of individuals' vote choice (Fiorina, 1981; Kinder and Kieweit, 1981; Kinder, Adams and Gronke, 1989). People who evaluate government performance as good are more likely to vote for the incumbent party. To estimate this effect, I also put evaluation on government performance into the model.

Table 3 shows the corrected and uncorrected estimates in the social psychological models. As expected, the three variables (in Model I) or four variables (in Model II) are positively related to the likelihood of voting Republican. Candidate evaluation is a very powerful determinant of vote choice. More noteworthy, party identification, candidate evaluation, and government performance are all susceptible to selection bias. The estimates of party identification and government performance are biased upward, while candidate evaluations are biased downward when not correcting the bias. To get a good

Table 3: Social Psychological Model with Corrected and Uncorrected Estimates (1988 Presidential Voting)

	Model I		Model II	
	Corrected	Uncorrected	Corrected	Uncorrected
Outcome Equation				
Party Identification	1.55 (.18)	1.69 (.18)	1.29 (.23)	1.41 (.18)
Issue Preferences	1.38 (.43)	1.41 (.42)	1.18 (.41)	1.21 (.43)
Candidate Evaluations	9.13 (.51)	9.02 (.94)	8.27 (.50)	8.16 (.96)
Government Performance	———	———	.80 (.16)	.91 (.16)
Constant	-5.99 (.36)	-6.10 (.49)	-5.77 (.26)	-5.94 (.51)
ρ	-.20 (.06)		-.22 (.06)	
Selection Equation				
Education	1.00 (.15)		1.06 (.15)	
Income	.95 (.15)		.92 (.15)	
Age	1.23 (.20)		1.20 (.20)	
New Resident	-.29 (.13)		-.28 (.13)	
Care about Elections	.26 (.08)		.25 (.08)	
Interest in Campaign	.85 (.13)		.86 (.02)	
Newspaper Usage	.42 (.08)		.44 (.08)	
Strength of Partisanship	.66 (.12)		.69 (.12)	
Affect for Candidate	.48 (.13)		.47 (.18)	
Political Efficacy	.42 (.13)		.43 (.13)	
Constant	-2.28 (.15)		-2.25 (.16)	
N	1,565	1,192	1,522	1,141

Note: 1. All variables are scaled from 0 to 1, with 0 the lowest and 1 the highest.

2. The dependent variable of outcome equation is 1, if voting for the Republican candidate(s), 0 if voting for the Democratic party. The dependent variable of selection equation is 1 if voting, and 0 if not voting.

3. The entries are probit coefficients (with standard errors in parentheses).

4. The Ns in the corrected models are larger than those in the uncorrected models because of the following reason. I include voters and nonvoters when estimating the correted models by way of Dubin and Rivers' approach. What I need from nonvoters is the information about the probability of voting (Please see Appendix A). Therefore, even though nonvoters have missing values in the outcome equation variables, they can be induded. However, I include only those who vote and have non-missing values in the outcome equation when estimatting the uncorrected models.

feel about the size of the effects, I calculated the probabilities of voting for the Republican party based on Table 3 Model II. (The four variables in the model can correctly predict presidential voting with 91 percent.) In this way, I can evaluate the size of the change of a variable on the likelihood of voting Republican with other variables held constant. The results are reported in Table 4.

Table 4 shows the probabilities of voting Republican when all variables are held at their mean values (the baseline model), and when the scores of variables increase by .1, .2 , and .3. (Remember all variables are 0-1 intervals.) When all variables are held at their mean values, the probabilities of voting for the Republican party are 49.71 and 45.86 in corrected and uncorrected models, respectively.

As shown in Table 4, candidate evaluation is a very powerful determinant of vote choice. After all variables are held at their mean, a .1 increase in the candidate evaluation score from its mean (.50) yields about a 30 percent increase in the probability of voting Republican. (See the first and second columns of the corrected and uncorrected models.) This increase in probability declines by about half when the candidate evaluation score changes from .6 to .7, (see the second and the third columns), and is smaller than 5 percent when the candidate evaluation score changes from .7 to .8 (see the third and the fourth columns). This dramatic change shows that the effect of an independent variable on the dependent variable is not linear. Instead, it is determined by where the change happens. Party identification, issue preferences, and government performance also show

Table 4: Relationships Between Social Psychological Variables and the Probability of Voting for the Republican Party (1988 Presidential Voting)

	Corrected Model				Uncorrected Model			
	Baseline (Mean)	.1 increase	.2 increase	.3 increase	Baseline (Mean)	.1 increase	.2 increase	.3 increase
Candidate Evaluation	49.72	79.39	95.02	99.33	45.86	76.18	93.67	99.05
Party Identification	49.72	54.85	59.91	64.80	45.86	51.48	57.06	62.51
Issue Preferences	49.72	54.42	59.05	63.57	45.86	50.68	55.49	60.22
Government Performance	49.72	52.91	56.08	59.21	45.86	49.48	53.10	56.71

Note: 1. The entries are probabilities of voting for the Republican party.
2. Throughout the calculations, I set the remaining variables at their mean values:
Candidate evaluation= .50
Party identification= .47
Issue Preference= .52
Government Performance= .51

substantial effects when other variables are held at moderate levels. But, the influence of these three variables on voting choice is less than that of candidate evaluations. Comparing the results of the corrected and uncorrected models, I find the tendency of the effect of change in the independent variable on vote choice to be the same in either model. However, it is obvious that the uncorrected model systematically underestimates the probability of voting for the Republican party.

Overall, the social psychological model is subject to selection bias. The biases in candidate evaluation, party identification and government performance tell us that a variable may be susceptible to the selection bias, even though it does not appear in the selection equation. However, unlike the biases in the sociological model, the directions of biases are not as predictable in the social psychological model. That is because these variables do not show up in the selection equation. (Remember Greene's illustration of the direction of bias as being limited to the regressor which appears in both selection and outcome equations.)

Jacobson's Congressional Vote Choice Model

The nature of the congressional election is different from that of the presidential election. Hence, the determinants of vote decision in these two types of elections are not completely the same. For example, the factor of issue preferences is substantial in a presidential vote choice. But, it is not viewed as of the same importance in the congressional vote choice. On the other hand, the incumbent advantage has long been viewed as a substantial determinant in congressional voting. However, an incumbent president may not enjoy the same advantage as incumbent congressmen/women in elections, and is even at a disadvantage if the economic situation is bad. One conventional argument about the incumbent advantage of house members is that they enjoy better name recall and recognition. However, Jacobson challenges this argument and develops a series of models to test it. When he puts the likes and dislikes about candidates into the model, he finds that incumbency and name recognition have less effects on congressional voting. Therefore, he argues that voters are not attracted by incumbency per se, nor does the incumbency advantage arise merely from greater renown. Instead, voters' evaluation of the candidates play a more important role in their vote choice (Jacobson, 1987: 132-146). Here, I estimate one of his models to assess the effects of the selection bias.

Table 5 shows the corrected and uncorrected estimates of Jacobson's model. All estimates are in the expected directions. Specifically, the variables Democratic incumbent, familiar with Democratic candidate, likes something about Democratic candidate and dislikes something about the Republican candidate are negatively associated with the likelihood of voting for the Republican congressional candidate. On the other hand, the

Table 5: Jacobson's Congressional Voting Model with Corrected and Uncorrected Estimates

	1988 Congressional Election		1990 Congressional Election	
	Corrected	Uncorrected	Corrected	Uncorrected
Outcome Equation				
Party Identification	2.33 (.21)	2.33 (.19)	2.12 (.28)	2.15 (.25)
Democratic Incumbent	-.39 (.25)	-.54 (.24)	-.25 (.26)	-.47 (.25)
Republican Incumbent	.64 (.25)	.50 (.24)	.52 (.26)	.38 (.24)
Familiar with Democratic Candidate	-.81 (.26)	-.89 (.25)	-.80 (.33)	-.85 (.32)
Familiar with Republican Candidate	1.07 (.23)	1.13 (.22)	1.30 (.28)	1.15 (.32)
Likes about Democratic Candidate	-1.40 (.21)	-1.31 (.22)	-1.72 (.24)	-1.65 (.23)
Dislikes about Democratic Candidate	1.09 (.22)	1.12 (.20)	1.17 (.26)	1.32 (.26)
Likes about Republican Candidate	1.45 (.23)	1.50 (.22)	1.30 (.28)	1.39 (.24)
Dislikes about Republican Candidate	-.45 (.24)	-.42 (.22)	-1.05 (.32)	-1.05 (.23)
Constant	-1.35 (.29)	-1.32 (.27)	-1.14 (.40)	-1.23 (.27)
ρ	-.22 (.21)		-.32 (.22)	
Selection Equation				
Education	1.04 (.16)		.75 (.14)	
Income	1.00 (.16)		.49 (.14)	
Age	1.44 (.21)		1.29 (.18)	
New resident	-.31 (.14)		-.84 (.17)	
Care about Elections	.22 (.09)		.39 (.08)	
Interest in Campaign	.72 (.13)		1.01 (.29)	
Newspaper Usage	.44 (.09)		.42 (.09)	
Strength of Partisanship	.73 (.13)		.44 (.12)	
Affect for Candidates	.55 (.19)		1.10 (.29)	
Political Efficacy	.45 (.14)		———	
Constant	-2.48 (.17)		-2.53 (.13)	
N	1,429	1,050	1,652	802

Note: 1. All variables are scaled from 0 to 1, with 0 the lowest and 1 the highest.

2. The dependent variable of outcome equation is 1, if voting for the Republican candidate(s), 0 if voting for the Democratic party. The dependent variable of selection equation is 1 if voting, and 0 if not voting.

3. The entries are probit coefficients (with standard errors in parentheses).

4. The Ns in the corrected models are larger than those in the uncorrected models because of the following reason. I include voters and nonvoters when estimating the correted models by way of Dubin and Rivers' approach. What I need from nonvoters is the information about the probability of voting (Please see Appendix A). Therefore, even though nonvoters have missing values in the outcome equation variables, they can be induded. However, I include only those who vote and have non-missing values in the outcome equation when estimatting the uncorrected models.

variables Republican incumbent, familiar with Republican candidate, likes something about Republican candidate, and dislikes something about the Democratic candidate are positively associated with the likelihood of voting for the Republican congressional candidate. Comparing the relative sizes of the coefficients, I find that in addition to party identification, the factors of likes and dislikes of the candidates play substantial roles in determining congressional voting. The coefficients are biased when not correcting selection bias.

The most obvious differences between corrected and uncorrected estimates exist in the two incumbency variables. Specifically, when not correcting selection bias, researchers may underestimate the importance of the Republican incumbent, and overestimate the importance of the Democratic incumbent. The changes in the corrected estimates are so large that one may draw different conclusions about incumbency. Specifically, the coefficients of the Democrat incumbent change from -.54 to -.39 in 1988, and from -.47 to -.25 in 1990 when correcting the selection bias. On the other hand, the change of the Republican incumbent is in the opposite direction, from .50 to .64 in 1988, and from .38 to .52 in 1990. The magnitudes of the changes are so large that the estimates go from statistically significant to insignificant (the Democratic incumbent), or from insignificant to significant (the Republican incumbent) in the 1990 model (at .95 confidence level).

Overall, the Jacobson's model is subject to the selection bias. The most susceptible variables are the incumbent variables, which may lead to different conclusions about the effects.

Estimating the Vote Choice of Nonvoters

A long standing view about American elections is that if nonvoters would vote, then the Democratic party would always get more votes than otherwise. This view is based on the assumption that nonvoters are biased toward being lower educated, poorer, and younger; and that these people tend to vote for the Democratic party. To test whether this view is true, I predicted the vote choice of the nonvoters based on the corrected models. Even though I do not have nonvoters' data on vote choice, I have their other characteristics. Based on these characteristics, I can predict their probabilities of voting for Republican. If a respondent's expected probability of voting Republican is larger than .5, I categorized this respondent as a Republican voter. Then, I calculated the proportion of voting for the Republican party for nonvoters, voters, and the whole sample. The results are reported in Table 6.

It is obvious that the social psychological model and Jacobson's model are much bet-

Table 6: Proportions of Voting for the Republican Party:
Voters, Nonvoters, and the Whole Sample (1988 and 1990)

	1988 Presidential Voting	1988 Congressional Voting	1990 Congressional Voting
Sociological Model			
Voters	66.70	31.69	15.69
Nonvoters	55.87	15.99	7.78
Whole Sample	63.38	26.88	11.40
Social Psychological Model			
Voters	51.71		
Nonvoters	44.27	——	——
Whole Sample	49.65		
Jacobson's Model			
Voters		40.70	37.96
Nonvoters	——	33.66	38.33
Whole Sample		38.62	38.16

Note: 1. The entries are proportions of voting for the Republican party in that group.
2. The Predicted probability of voting for the Republican party are based on corrected models shown in Table 2, Table 3, and Table 5.

ter than the sociological model in predicting individuals' vote choice. This is because for these two models the predicted proportions of voters who vote Republican are close to the real proportions that Republican party obtained in the elections examined.

There is a very consistent tendency across different models and elections: nonvoters tend to be less supportive of the Republican party. The proportion of nonvoters who would vote Republican is about 7 percent less than the proportion that the voters did vote for Republican in both 1988 elections. (See social psychological model and Jacob's model.) Though the distance between voters and nonvoters is not as clear in 1990, the tendency that nonvoters are more likely to vote Democratic is evident. Furthermore, in the 1988 presidential election, if the nonvoters had voted, the election outcome would have been different. (The proportion voting Republican would reduce from 51.71 percent to 49.65 percent, if nonvoters had voted.)

The findings in this section confirm that nonvoters are biased toward being lower educated, poorer, and younger; and that these people tend to vote for the Democratic party. Therefore, when researchers aiming at the relationships between the socio-demographic and psychological factors and the vote choice of a whole population, but using only the voter subsample, may obtain inconsistent estimates.

Conclusion

I start from a suspicion that general vote choice models are subject to the selection bias resulting from excluding nonvoters. Based on Dubin and Rivers' approach, I assess three basic vote choice models. Research findings show that all of the three vote choice models are subject to selection bias, especially the sociological model and Jacobson's model.

We can learn several things from this paper. First, if an independent variable affects the dependent variable of interest and also affects the selection process, then this variable is more likely to be susceptible to the selection bias than others which do not affect selection process. For example, education, income, and age are relatively more susceptible to the selection bias than others. However, even though an independent variable does not appear in the selection equation, it may also affect the selection process by its relationship with the variables in the selection equation. Therefore, even though an independent variable is not specified in selection equation, we still need to worry about the possible impacts of the selection bias. For example, in Jacobson's model, variables of incumbency may not directly affect selection; but they are still susceptible to selection bias.

Second, comparing results of 1988 and 1990, I find that selection bias in the vote choice model is not necessarily related to the voter turnout rate. The turnout rate in 1988 was higher than that in 1990, because the former was a presidential election year. In the sociological model, the 1990 congressional voting model seems not to be affected by selection bias, while the 1988 congressional model is seriously affected. However, in Jocobson's congressional voting model, selection bias in 1990 is more serious than in 1988. This tells us that the seriousness of selection bias is not necessarily related to the turnout rate.

Finally, even though selection bias may not always cause different directions or inferences about the effects, it does cause overestimation or underestimation of the effects. Furthermore, in some situations, it does cause different inferences about the effects of some variables of interest. Because political scientists are not only interested in whether a relationship exists and the direction of the relationship, they also care about the magnitude of this relationship. Therefore, a precise coefficient is necessary, and hence correction of selection bias is also necessary.

Appendix A:
Dubin and Rivers' Approach of Dealing With Selection Bias

Dubin and Rivers estimate the selection equation and outcome equation simuetaneously by way of Maximum Likelihood Estimation.

The outcome equation is:

$$y_{1i}^* = \beta_1' x_{1i} + u_{1i}$$

Define a dummy variable y_{1i},

$$y_{1i} = \begin{cases} 1, \text{ if } y_{1i}^* > 0 \\ 0, \text{ otherwise} \end{cases}$$

The selection equation is:

$$y_{2i}^* = \beta_2' x_{2i} + u_{2i}$$

Define a dummy variable y_{2i},

$$y_{2i} = \begin{cases} 1, \text{ if } y_{2i}^* > 0 \\ 0, \text{ if } y_{2i}^* \leqq 0 \end{cases}$$

Because the probability of y_{1i} is conditional on y_{2i}, we have to consider the joint normal distribution function. The joint cumulative density function takes the form $F(u_1, u_2; \rho)$, and the marginal distributions $H(u_1) = F(u_1, \infty; \rho)$ and $H(u_2) = F(-\infty, u_2; \rho)$. Here, ρ is the correlation between u1 and u_2.

There are three possible outcomes: an uncensored success ($y_{1i} = 1$ and $y_{2i} = 1$), an uncensored failure ($y_{1i} = 0$ and $y_{2i} = 1$), and a censored observation ($y_{2i} = 0$). Then, the next step is to calculate the probability of these three possible outcomes. Let $G(\cdot, \cdot; \rho)$ denote the upper tail probability of $F(\cdot, \cdot; \rho)$; i.e.:

$$\begin{aligned} G(u_1, u_2, \rho) &= \Pr(u_{1i} > u_1, u_{2i} > u_2) \\ &= 1 - H(u_1) - H(u_2) + F(u_1, u_2; \rho) \end{aligned}$$

The probability of an observation not being censored is given by

$$\begin{aligned} Q_i(\beta_2) &= \Pr(y_{2i} = 1 \mid x_{1i}, x_{2i}) = \Pr(Y_{2i}^* > 0) \\ &= 1 - H(-\beta_2' X_{2i}) \end{aligned}$$

Therefore, the probabilities of these three outcomes are as follow:

(Censored) $1- Q_i (\beta_2)$

(Uncensored success) $P(\beta_1, \beta_2, \rho)$

$$= Pr(y_{1i} = 1, y_{2i} = 1 | x_{1i}, x_{2i})$$

$$= Pr(y_{1i}{}^* > 0, y2i^* > 0 | x_{1i}, x_{2i})$$

$$= G(-\beta_1{}'X_{1i}, -\beta_2{}'X_{2i})$$

(Uncensored failure) $Q_i (\beta_2)- P(\beta_1, \beta_2, \rho)$

Combining these three parts, we obtain the log likelihood function as follows:

$$L(\beta_1, \beta_2, \rho) = \sum_{i=1}^{n} y_{2i}\, y_{1i} \log P_i (\beta_1, \beta_2, \rho) + \sum_{i=1}^{n} y_{2i} (1- y_{1i}) \log (Q_i (\beta_2)- P_i(\beta_1, \beta_2, \rho))$$
$$+ (1-y_{2i}) \log (1-(Q_i (\beta_2))$$

The Maximum likelihood estimator of $\theta = (\beta_1, \beta_2, \rho)$ is obtained by maximizing the Log likelihood function with respect to θ. The detailed solution for this particular case is shown in Dubin and Rivers' article (1989). I do not repeat the procedure here.

Appendix B:
Measurement and Coding of Variables

All data used in this paper are taken from the American National Election Studies Cumulative Data File, 1952-1990.

Voter Turnout. Respondents who vote in the election are coded 1; respondents who do not vote are coded 0.

Education. Respondents with 8 grades or less education are coded 0; respondents with 9 -11 grades of education are coded .2; respondents with high school diploma are coded .4; respondents with more than 12 years of education but no higher degree are coded .6; respondents with junior or community college level degree are coded .8; respondent with bachelor or higher degree are coded 1.

Income. This variable reports respondents' family income. Respondents' family income of less than 9,999 are coded 0; 10,000-16,999 are coded .25; 17,000 -34,999 are coded .5; 35,000 -89,999 are coded .75; more than 90,000 are coded 1.

Age. This variable recode respondents' age to a 0-1 interval. With 0 the youngest, and 1 the oldest.

New Resident. Respondents who live in the current community less than 1 year are coded 1, 0 otherwise.

Care about Elections. Respondents who personally care about which party wins the presidential election (in 1988), or care about the way the election of the House of Representatives came out (in 1990) are coded 1, otherwise 0.

Interest in Campaign. Respondents who are not much interested in the political campaign in the particular year are coded 0; somewhat interested are coded .5; and very much interested are coded 1.

Newspaper Usage. Respondents who read campaign in the newspaper are coded 1, 0 otherwise.

Strength of Partisanship. Respondents who are independent or apolitical are coded 0; independent leaning toward a party are coded .333; weak partisans are coded .667; strong partisans are coded 1.

Affect for Candidates. This variable is based on four questions about likes and dislikes about the candidates of the two major parties (presidential candidates in 1988, and congressional candidates in 1990). Take an absolute value of the difference of two sums then coded to a 0-1 interval: the sum of Democratic candidate "likes" and Republican candidate "dislikes" minus the sum of Democratic candidate "dislikes" and Republican candidate "likes."

Political Efficacy. This variable is based on three items: " Public officials don't care much what people like me think." "People like me don't have any say about what the government does." "Sometimes politics and government seem so complicated that a person like me can't really understand what's going on." Add the answers up and recode to a 0-1 interval, with 0 the lowest efficacy, and 1 the highest efficacy.

Race. Respondents who are blacks are coded 1, 0 otherwise.

Married. Respondents who are married and live with spouses are coded 1, 0 otherwise. **Worker.** Respondents who are clerical workers, sales workers, skilled or semi-skilled workers, service workers, or laborers are coded 1, 0 otherwise.

Union Member. Respondents or their families belong to a labor union are coded 1, 0 otherwise.

Rural Resident. Respondents who live in a rural area are coded to 1, 0 otherwise.

Region. Respondents who live in southern region are coded 1, 0 otherwise.

Protestant. Respondents who are Protestant are coded 1, 0 otherwise.

Vote Choice. Respondent who vote for the Republican candidate are coded 1, 0 otherwise. Party Identification. This variable is based on a traditional SRC/CPS ANES measure, and recode to a 0-1 interval: strong Democrats 0; weak Democrats .167, independent Democrats .333, independent .5, independent Republicans .667, weak republicans .88 4, and strong Republicans 1.

Issue Preferences. This variable is based on five 1-7 scaled issues: (1) Government or private insurance plan; (2) government sees to job and good standard of living or government lets each person go ahead on his own; (3) government should help blacks or blacks should help themselves; (4) government should provide more services and increase spending or government should provide fewer services and reduce spending; and (5) government should decrease or increase defense spending. Run factor analysis on these five issues and save the factor scores for each of the respondents. Then, recode the factor scores to a 0-1 interval, with 0 the more likely to prefer the position of the Democratic party, and 1 the more likely to prefer the position of the Republican party.

Candidate Evaluations. This variable is based on respondents' comparisons on two major presidential candidates in three respects: (1) the difference between evaluation on two candidates' personalities; (2) the difference between evaluations on two candidates' abilities; and (3) the differences between affect for the two candidates. Indicators of candidate personality and ability include a series of survey questions which asked respondents to evaluate how well the phrase "____" describes the candidate. Indicators of personality include the following four phrases: compassionate, decent, moral, and cares about people. Indicators of ability include four phrases: intelligent, inspiring, knowledgeable, and strong leadership. On the other hand, the indicator of affect for the candidates is based on the survey questions of likes and dislikes about the candidates. The candidate evaluation scaling is built by the following steps: (1) Sum up the four items of personality for each of the two candidate and then divide the sum by the number of questions answered,[4] so as to get evaluation scores on each of the two candidates' personalities; (2) subtract the Democratic candidate personality scores from the Republican candidate personality scores to get the differences between evaluations on two candidates' personalities; (3) obtain the differences between evaluations on two candidates' abilities in the same way as steps (1) and (2), except that the items are replaced by the ability items ; (4) obtain the scores of affect for each of the two candidates by subtracting candidate "likes" by candidate "dislikes;" (5) take a subtraction of the differences between the affect for the two candidate: Affect for the Republican candidate minus affect for the Democratic candidate; (6) run factor analysis on the three indicators: comparative evaluations on candidate personality, ability and affect, and save the factor scores for each of the respondents; (7) recode the factor scores to a 0-1 interval, with 0 the more likely to prefer the Democratic candidate, and 1 the more likely to prefer the Republican candidate.

Democratic Incumbent. Respondents who live in election districts with Democratic incumbents in the House of the Representatives are coded 1, 0 otherwise.

Republican Incumbent. Respondents who live in election districts with Republican incumbents in the House of the Representatives are coded 1, 0 otherwise.

Familiar with Democratic Candidate. Respondents who can recall the Democratic congressional candidate's name are recoded 1; respondents who recognize the name from a list are coded .5; 0 otherwise.

Familiar with Republican Candidate. Respondents who can recall the Republican congressional candidate's name are recoded 1; respondents who recognize the name from a list are coded .5; 0 otherwise.

Likes about Democratic Candidate. Respondents who like something about the Democratic congressional candidate are coded 1, 0 otherwise.

Dislikes about Democratic Candidate. Respondents who dislike something about the Democratic congressional candidate are coded 1, 0 otherwise.

Likes about Republican Candidate. Respondents who like something about the Republican congressional candidate are coded 1, 0 otherwise.

Dislikes about Republican Candidate. Respondents who dislike something about the Democratic congressional candidate are coded 1, 0 otherwise.

Presidential Vote. Respondents who vote for the Republican presidential candidate are coded 1, 0 otherwise.

Notes

1. I use respondents' self-report turnout rather than "valid turnout," though I understand some respondents may overreport their voting. There are two major advantages using self-report turnout. First, the major purpose of this paper is people's vote choice, rather than people's turnout. Therefore, even though people may overreport their turnout, they may possibly give true answers about vote choice. If I use "valid turnout," I may have to sacrifice the vote choice answers of the people who give answers about their vote choices but misreport their turnout. Based on the understanding about American voters, I believe that the possibility of misreporting voter turnout is larger than that of vote choice. This is because people who view voting as a civic duty may feel embarrassed if they do not vote. Therefore, they may overreport their turnout. But, it is quite impossible that when they actually vote for the Republican party they say that they vote for the Democratic party. Second, using the same explanatory variables in Table 1, the self-report turnout can be predicted better than the valid turnout (with about 8 percent better prediction in 1988, and the same percent correct prediction in 199 0).

2. For example, Knoke well illustrates that higher income leads people to greater economic conservatism, while higher education leads people to liberalism on social issues (Knoke, 1979).

3. The formula that Greene derives to show the full effect of a regressor x_k that appears in both outcome equation and selection equation is as follows: (Notice: In order to make the symbol consistent in this paper, I use different symbols from Greene's.)

$$\frac{E[y_{1i} \mid y_{2i}{}^* > 0]}{x_{ik}} = \beta_{1k} - \beta_{2k}\, \rho \frac{\sigma_{u1}}{\sigma_{u2}}\, \delta_i\, (\alpha_{u2})$$

Where, σ_{u1} , σ_{u2} , and $\alpha_{u2} > 0$, and $0 < \delta_i < 1$. So, the bias based on observed sample is determined on two values, β_{2k} and ρ. If $\beta_{2k} > 0$ and $\rho > 0$, then the second term is positive and thus the estimate is biased upward. On the other hand, if $\beta_{2k} > 0$ and $\rho < 0$, then the second term is negative and thus the estimate is biased downward.

4. That is, if respondents answer four of the questions for a particular candidate, the sum of the candidate personality is divided by 4, answer two questions, divided by 2. If respondents do not answer all of four questions, they are missing. In this way, I can keep the respondents who answer some questions but do not answer all of the questions.

Bibliography

✓ Achen, Christopher H.

1986 *The Statistical Analysis of Quasi- Experiments.* Berkeley: The University of California Press.

Aldrich, John H. and Forrest D. Nelson.

1984 *Linear Probability, Logit, and Probit Models.* Newbury Park: Sage Publications.

Asher, Herbert B.

1983 "Voting Behavior Research in the 1980s: An Examination of Some Old and New Problem Areas." in Ada Finifter, eds. *Political Science: the State of the Discipline.* American Political Science Association. pp. 339-388.

✓ Brehm, John.

1993 *The Phantom Respondents.* Ann Arbor: The University of Michigan Press.

Campbell, Angus, Philip E. Converse, Warren Miller, and Donald E. Stokes,

1960 *The American Voter.* New York: John Wiley and Sons, Inc.

✓ Dubin Jeffrey A. and Douglas Rivers.

1989 "Selection Bias in Linear Regression, Logit and Probit Models." *Sociological Methods and Research.* 18:360-390.

Fiorina, Morris P.

1981 *Retrospective Voting in the American National Elections.* New Haven and London: Yale University Press.

Finkel, Steven E.

1985 "Reciprocal Effects of Participation and Political Efficacy: A Panel Analysis." *American Journal of Political Science.* 29:891-913.

Gant, Michael and Norman R. Luttbeg.

1991 *American Electoral Behavior.* Itasca, Illinois: F.E. Peacock Publishers, Inc.

Greene, William H.

1993 *Econometric Analysis*. 2nd edition. New York: Macmillan Publishing Company.

Heckman, James J.

1979 "Sample Selection Bias as a Specification Errors." *Econometrica*. 47:153 -161.

Jackson, John E.

1975 "Issues, Party Choices and Presidential Votes." *American Journal of Political Science*. 19:161-185.

Jacobson, Gary C.

1987 *The Politics of Congressional Elections*. 2nd edition. Boston: Little, Brown, and Company.

Key, V.O. Jr.

1949 *Southern Politics in State and Nation*. New York: Knoph.

Kinder, Donald and D. Roderick Kiewiet.

1981 "Sociotropic Politics: The American Case." *British Journal of Political Science*. 11:129-161.

Kinder, Donald and D. O. Sears.

1985 "Public Opinion and Political Action. " in Gardner Lindzey and Elliot Aronson, *Handbook of Social Psychology*. 3rd edition. Vol. II. New York: Random House.

Kinder, Donald, Gordon S. Adams, and Paul W. Gronke.

1989 "Economics and Politics in the 1984 American Presidential Elections." *American Journal of Political Science*. 33: 491-515.

King, Gary.

1989 *Unifying Political Methodology*. New York: Cambridge University Press.

Knoke, David.

"Stratification and the Dimensions of American Political Orientation." *American Journal of Political Science*. 23: 772-791.

Lazarsfeld, P. Berelson, and H. Gaudet.

1944 *The People's Choice*. New York: Columbia University Press.

Lipset, Seymour Martin.

1981 *Political Man: The Social Bases of Politics.* Expanded edition. Baltimore: The John Hopkins University Press.

Markus, Gregory B. And Philip E. Converse

1979 "The Dynamic Simultaneous Equation Model of Electoral Choice." *The American Political Science Review.* 73:1055-1070.

Markus, Gregory B.

1982 " Political Attitudes during an Election Year: A Report on the 1980 NES Panel Study." *The American Political Science Review.* 76: 538-560.

Nie, Norman H., Sidney Verba and John Petrocik.

1976 *The Changing American Voter.* Cambridge, MA: Harvard University Press.

Page, Benjamin I. and Calvin C. Jones.

1979 "Reciprocal Effects of Policy Preferences, Party Loyalties and the Vote." *The American Political Science Review.* 73:1071- 1090.

Rosenstone, Steven J. and John Mark Hansen.

1993 *Mobilization, Participation, and Democracy in America.* New York: Macmillan Publishing Company.

Squire, Peverill, Raymond E. Wolfinger, and David Glass.

1987 "Residential Mobility and Voter Turnout." *American Political Science Review.* 81: 45-65.

Verba, Sidney and Norman Nie.

1972 *Participation in America: Political Democracy and Social Equality.* New York: Harper & Row, Publishers.

Wolfinger, Raymond E., and Steven J. Rosenstone.

1980 *Who Votes?* New Haven: Yale University Press.