Discussion Paper Series
No. 0907
March 2009

Strategic Voting and Multinomial Choice In US Presidential Elections

Myoung-jae Lee and Sung-jin Kang

The Institute of Economic Research - Korea University
Anam-dong, Sungbuk-ku, Seoul, 136-701, South Korea, Tel: (82-2) 3290-1632, Fax: (82-2) 928-4948

STRATEGIC VOTING AND MULTINOMIAL CHOICE IN US PRESIDENTIAL ELECTIONS

(December 16, 2008)

Myoung-jae Lee*
Department of Economics
Korea University
Anam-dong, Sungbuk-gu
Seoul 136-701, South Korea
myoungjae@korea.ac.kr

Sung-jin Kang
Department of Economics
Korea University
Anam-dong, Sungbuk-gu
Seoul 136-701, South Korea
sjkang@korea.ac.kr

Ross Perot was a relatively viable third party candidate in the 1992 US presidential election, but he was not any more in the 1996 election. This provides a good opportunity to analyze strategic voting behavior—voting for a candidate not most preferred by the voter—in the US presidential elections with panel data drawn from NES (National Election Studies). First, the 1992 election is analyzed with multinomial choice estimators. Second, using the estimates, each individual's choice is predicted for the 1996 election. Third, those who were predicted to vote for Perot in 1996 but did not are identified as strategic voters and their profile is drawn. In addition to the main task of analyzing the strategic voting behavior, this paper does two additional tasks. First, analyzing the 1992 data with multinomial choice estimators, it is found that the following variables mattered significantly for the US presidential election: respondent and candidate ideology, personal finance, age, education, income, sex, abortion stance, health insurance policy, and welfare program policy. Second, critical mistakes in the literature in applying multinomial probit to election data are pointed.

Key Words: strategic voting, presidential election, multinomial logit, multinomial probit.

* Corresponding Author.

** The research for this paper of the first author, Myoung-jae Lee, has been supported by a Korea University Grant.

1 Introduction

When there are more than two parties in an election, voters may not cast their vote following their true preference. Rather, to avoid their vote going wasted, they may vote for another party more likely to win, which is a strategic voting; see, e.g., Mueller (2003), Kousser (2004), and the references therein. It is interesting to know whether strategic voting indeed occurs, and if so, what kind of people vote strategically. A main goal of this paper is to *examine strategic voting behavior for the US presidential election using panel data* drawn from the National Election Studies (Warren et al (1999)); the panel data enables identifying strategic voters, which is new in the literature for strategic voting.

In the literature of political science, estimates for the percentage of strategic voters vary much; e.g., as lows as 5% in Johnston and Pattie (1991) for UK and as high as 14% in Abramson et al (1992) for US. Some empirical studies estimated strategic voting extent with aggregate data (e.g. Galbraith and Rae (1989) and Johnston and Pattie (1991)); some relied on individual data stating the reason for their choices (e.g. Niemi et al (1992)); and some compared the actual vote to the preference ranking of parties or candidate (e.g., Abramson et al (1992) and Blais and Nadeau (1996)). But aggregate data can deliver, at best, only limited results such as strategic voter percentage, and there is a reliability problem using stated preferences rather than the observed ones. The main difficulty in uncovering strategic voting is how to come up with the 'counter-factual' choice of voters when, contrary to the fact, there were no incentive for strategic voting; the counter-factual choice is then to be compared with the actual choice. The aforementioned studies either could not get the counter-factual or had to rely on stated reason/preference.

Facing these shortcomings in the literature, Alvarez and Nagler (2000) apply 'multinomial probit' to individual data. They postulate that the latent utility by choosing party (or candidate) j for individual i is

$$s_{ij} = w'_{ij}\beta + \zeta\nu_{ij} + u_{ij}$$

where w_{ij} is a regressor vector, ν_{ij} is a measure of incentives to vote strategically (ν_{ij} is a function of expected vote shares of the parties in the election), u_{ij} is an error term, and β and ζ are parameters. In this framework, person i will choose alternative j if $s_{ij} > s_{im}$ $\forall m \neq j$. Alvarez and Nagler (2000) estimate β and ζ with multinomial probit. Then, to

get the counter-factual, they set $\nu_{ij} = 0$ and use only $w'_{ij}\beta$. Since the error terms are not observed, the counter-factual choice for individual i is set to be the alternative with the largest value of $w'_{ij}\beta$. This attempt by Alvarez and Nagler is a step in the right direction, but has two problems—one in theory and the other in implementation.

The problem in implementation is that Alvarez and Nagler (2000) estimate non-identified parameters in their multinomial probit; this problem is in fact pervasive in their other papers as shown in detail later; the literature on multinomial probit is riddled with similar mistakes (trying to estimate non-identified parameters). The theoretical problem—in fact, a shortcoming rather than a problem—is possible biases from various misspecifications. Multinomial probit, although it is more general than other popular parametric methods for multinomial choice, is a tightly specified model. If there is a misspecification, then the estimator is inconsistent in general. This implies that the counter-factual based on $w'_{ij}b_N$ where b_N is an inconsistent estimator for β is misleading. This is, however, less problematic in panel data as explained in the following.

Suppose we have two-wave panel data with latent utility

$$s_{ijt} = w'_{ijt}\beta + u_{ijt}, \;\; t = 1, 2$$

where the subscript t indexes two periods $1, 2$ and there is no incentive for strategic voting at period 1 whereas there is some at period 2. Suppose β is estimated with b_N at period 1 and $w'_{ij2}b_N$ is used to construct the counter-factual at period 2. In this approach, potential inconsistency of b_N for β does not pose much of a problem, because b_N is used only for prediction purpose for period 2 rather than for "structural policy" purpose.

To see the point better, consider the usual least squares estimator (LSE) for a cross-section linear model $y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + u_i$ where u is uncorrelated with x_2 and x_3 and $COR(x_2, x_3) \neq 0$. This model can be used for two different purposes: structural policy and prediction. In the former, we ask questions such as "what is the effect on $E(y|x_2, x_3)$ of changing x_2 while holding x_3 constant". In the latter, we want to predict a future y for given values of x_2 and x_3 as closely as possible. Suppose now that x_3 is not observed. Then the LSE of y on x_2 incurs the well known omitted variable bias. Without loss of generality, x_3 can be decomposed as $x_{i3} = \alpha_1 + \alpha_2 x_{i2} + v_i$ where $COR(x_2, v) = 0$ and α_1 and α_2 are the linear projection coefficients. Substitute this into $y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + u_i$ to get

$$y_i = \beta_1 + \beta_3\alpha_1 + (\beta_2 + \beta_3\alpha_2)x_{i2} + (\beta_3 v_i + u_i).$$

The LSE of y on x_2 is consistent for $\beta_2+\beta_3\alpha_2$, which is however misleading for the structural policy question because x_3 is not held constant: the change in x_3 due to the change in x_2 is $\beta_3\alpha_2$. But for the prediction purpose, the best prediction for y minimizing the prediction error $E(y-\gamma_1-\gamma_2 x_2)^2$ with a given value of x_2 is $\beta_1+\beta_3\alpha_1+(\beta_2+\beta_3\alpha_2)x_2$; here $\beta_3\alpha_2$ adds to the prediction accuracy because it accounts for the unobserved x_3 as well as the observed x_2. Going back to the panel data multinomial choice, when period 1 data is used for prediction for period 2, inconsistency in estimation is not an important issue. In Alvarez and Nagler (2000), setting $\nu_{ij}=0$ in $w'_{ij}\beta+\zeta\nu_{ij}$ to get the counter-factual is exploring a structural policy question of changing ν_{ij} from non-zero to zero while holding the other variables constant.

There have not been many viable third party candidate in the US presidential election. But in the 1992 and 1996 elections, Ross Perot was a third party candidate. In the 1992 election, his popularity was relatively high, particularly in the early campaign period. In the 1996 election, however, his popularity was much lower with hardly anybody giving him a realistic chance to win. Since strategic vote is thought to occur when the preferred third party has little chance to win, these two presidential elections provide a good forum to test for strategic voting in the US presidential election. That is, by comparing the voting behavior in the 1992 and 1996 elections where the latter presents the stronger possibility for strategic voting, we can test for the presence of strategic voting and possibly identify who does that.

Our main plan to achieve this goal is as follows. First, analyze the 1992 election with a multinomial choice estimator and the first wave (for 1992) of the panel data. Second, using the estimates and the second wave (for 1996) of the panel data, predict how each individual would vote in 1996. Third, select those who are predicted to vote for Perot in 1996 but did not; they are the strategic voters.

Setting out to carry out the plan, we ran into a problem. There are essentially two multinomial choice estimators used in practice: multinomial logit (MNL) and multinomial probit (MNP). MNP is more general than MNL, but MNP does not work well in practice; for our data, MNP failed to converge. But then we found out that Alvarez and Nagler (1995,1998b) already used MNP for the US presidential election data drawn from the same source. Examining these papers in detail, we found a number of critical errors: non-identified parameters have been estimated. This brings up two other (minor) goals of this paper: one is properly (re-) *analyzing the 1992 election with MNL*, and the other is *explaining the multinomial choice estimators in detail to show the identified parameters in MNP and the mistakes*

in Alvarez and Nagler (1995,1998b,2000). A fairly detailed discussion of multinomial choice estimators will be provided later; Train (2003) provides a more extensive review, and see also Lee and Kim (2007).

With the three italicized goals (i.e., tasks) in mind, the rest of this paper is organized as follows. Section 2 does the second task of analyzing the 1992 election with MNL, where it will be seen which variables and which issues for the election really mattered. Section 3 does the first task of finding strategic voters using the estimates from Section 2. Section 4 does the third task of reviewing multinomial choice estimators and pointing out the errors in the literature. Finally, Section 5 concludes.

Logically, it is fitting to review multinomial choice estimators first, but this is technically more involving than the other topics, which is why the review is deferred to Section 4 to not distract the reader; readers lacking background knowledge on multinomial choice may want to read Section 4 first.

2 Analysis of 1992 US Presidential Election

Subsection 2.1 shows the variables in our data and notes two problems (lack of cardinality and missing data). Subsection 2.2 presents the estimation results of MNL and two other estimators. Subsection 3.3 shows the 'average effects' of regressors in MNL.

2.1 Variables and Problems

The panel data set we constructed has the sample size of about 900 with many explanatory variables available to explain the voter choice. Unfortunately, there were numerous missing entries and most explanatory variables were either binary or ordinal at best. This posed some difficulties, which need to be mentioned before presenting our data analysis. Also, the questionnaire has changed somewhat over the years, which also limited our selection of regressors for the panel data construction.

Among the regressors used, only age and $age^2/100$ are cardinal, whereas all the others are either dummy or ordinal variables. We list the dummy variables first, and then the ordinal variables. The following dummy variables taking the values 0 or 1 were used: with "R" denoting the respondent,

eco.-worse: 1 if R feels the national economy is worse (and 0 otherwise);

fin.-better: 1 if R feels his/her personal finance is better;

fin.-worse: 1 if R feels his/her personal finance is worse;

returning: 1 if R voted in the 1988 election;

female: 1 if R is female;

east, south, west: residential location dummies (central is the base category).

The following ordinal variables were used (the variables taking 1 to 7 ask the respondent about the government active role on some issues):

ideology-P: 1 to 7 on R's placement of Perot ideology: liberal to conservative;

ideology-C: 1 to 7 on R's placement of Clinton ideology: liberal to conservative;

ideology-B: 1 to 7 on R's placement of Bush ideology: liberal to conservative;

ideology-R: 1 to 3 for R being liberal to conservative;

education: 1 to 7 for grade 8 or less to graduate study;

income: 1 to 24 for below \$3,000 to above \$105,000 per year;

party: 0 to 2 for Democrat to Republican;

abortion: 0 to 2 for anti-abortion to pro-abortion;

black: 1 to 7 on government active help for blacks: none to should;

defense: 1 to 7 on government defense spending: decrease to increase;

health: 1 to 7 on health insurance: private insurance to government insurance;

job: 1 to 7 on government active promotion for job: none to should;

service: 1 to 7 on government service provision: decrease to increase;

ssc (social security): 1 to 7 on social security increase: removal to increase;

welfare: 1 to 7 on welfare programs: removal to increase.

The higher the ideology variables, the more conservative. But for some variables on the government policies (black, job, service, and welfare), the higher the more liberal.

These ordinal variables pose a problem: unless they are cardinal as well, differences do not make sense; e.g., the difference $2-1$ does not mean the same magnitude as the difference $7-6$ in the variables taking 1 to 7. In principle, those 7-category ordinal variables should be used to generate 6 dummy variables for each ordered category. But, given the data size and the number of the ordinal variables, this leads to excessively many parameters to estimate.

Hence, we decided to use the ordinal variables as such, taking them as cardinal; i.e., the difference $2-1$ is taken as the same as the difference $7-6$ in the above example.

TABLE 1: Regressor Mean (SD)			
age	44.0 (16.4)	east	0.181 (0.39)
eco.worse	0.719 (0.45)	south	0.335 (0.47)
education	4.39 (1.61)	west	0.190 (0.39)
fin.-better	0.332 (0.47)	abortion	1.44 (0.65)
fin.-worse	0.331 (0.47)	black	3.32 (1.66)
ideology-R	2.16 (0.94)	defense	3.48 (1.27)
income	15.8 (5.75)	health	4.40 (1.83)
party	1.01 (0.53)	job	3.54 (1.68)
returning	0.788 (0.41)	service	4.01 (1.42)
female	0.529 (0.50)	ssc	6.35 (0.63)
		welfare	5.63 (0.81)

Other than the ordinal variable problem, the other major problem is missing data: almost all variables in the raw data have missings. In the raw data, typically, the missing percentage is around 10 to 20%, but the missing percentage is much higher for party affiliation and the choice (response) variable. Although we could have removed all observations with any missing variable, this would have entailed too much information loss. Instead, imputation was done for some ordinal regressors where missing may be construed as "neutral"; e.g., for ordinal variables taking 1 to 7 (ideology, black, defense,...) the missings were imputed with the middle category 4. For abortion, the opinion was listed in four ordered categories (1 to 4), and we merged category 3 into 2 and renamed category 4 as category 3; the missings here were then imputed with 2. For the other variables, no imputation was done.

Our panel data size with at least one response observed in the two waves is $N = 894$ where the percentage of imputed values for most variables is about 10%. Among 894, excluding the observations with any missing, we get $N_1 = 647$ for 1992 and $N_2 = 456$ for 1996. The descriptive statistics (mean and SD) for wave 1 with $N_1 = 647$ are in Table 1.

2.2 Estimators and Problems

TABLE 2: MNL, MNP, and ODR

	Clinton(MNL)	Bush (MNL)	Clinton (MNP)	Bush (MNP)	ODR
ideology-P	0.008 (0.1)		-0.011 (-0.2)		-0.058 (-1.7)
ideology-C	0.349 (3.7)		0.068 (1.2)		-0.192 (-4.3)
ideology-B		0.115 (1.2)		0.005 (0.3)	0.036 (0.9)
1	-7.566 (-3.7)	-5.807 (-2.5)	-2.263 (-2.0)	-2.483 (-1.9)	1.903 (2.2)
age	0.119 (2.2)	-0.034 (-0.6)	0.033 (1.1)	0.012 (0.3)	-0.066 (-3.2)
$age^2/100$	-0.095 (-1.7)	0.068 (1.1)	-0.017 (-0.6)	0.008 (0.2)	0.069 (3.3)
eco.-worse	0.359 (1.1)	-0.492 (-1.6)	0.032 (0.2)	-0.149 (-0.7)	-0.312 (-2.3)
education	0.085 (0.9)	0.384 (3.4)	0.103 (2.1)	0.170 (2.1)	0.128 (3.1)
fin.-better	-0.092 (-0.3)	0.756 (2.1)	0.060 (0.4)	0.229 (1.0)	0.258 (1.8)
fin.-worse	0.079 (0.2)	0.597 (1.6)	0.064 (0.4)	0.176 (0.8)	0.154 (1.2)
ideology-R	-0.542 (-3.2)	0.461 (2.3)	-0.142 (-1.5)	0.030 (0.2)	0.445 (7.0)
income	-0.055 (-2.0)	-0.031 (-1.0)	-0.026 (-1.8)	-0.024 (-1.5)	0.014 (1.3)
party	-0.207 (-0.9)	1.145 (4.0)	0.196 (1.6)	0.543 (1.8)	0.449 (4.1)
returning	0.078 (0.2)	-0.124 (-0.3)	-0.105 (-0.6)	-0.157 (-0.8)	-0.060 (-0.4)
female	0.370 (1.4)	0.639 (2.2)	0.251 (1.8)	0.319 (1.9)	0.059 (0.5)
east	0.689 (1.8)	0.622 (1.5)	0.326 (1.6)	0.398 (1.7)	-0.131 (-0.8)
south	0.425 (1.3)	0.368 (1.0)	0.166 (1.0)	0.194 (1.0)	-0.008 (-0.1)
west	0.268 (0.7)	0.413 (1.0)	0.130 (0.6)	0.198 (0.8)	0.073 (0.4)
abortion	-0.072 (-0.3)	-0.732 (-3.1)	-0.206 (-1.7)	-0.335 (-2.0)	-0.258 (-3.0)
black	0.072 (0.8)	-0.113 (-1.2)	-0.001 (0.0)	-0.031 (-0.5)	-0.079 (-2.2)
defense	-0.098 (-0.9)	0.157 (1.3)	-0.006 (-0.1)	0.042 (0.5)	0.092 (2.0)
health	-0.108 (-1.4)	-0.239 (-2.9)	-0.095 (-2.5)	-0.113 (-2.4)	-0.056 (-1.7)
job	0.151 (1.6)	0.036 (0.4)	0.050 (1.0)	0.035 (0.6)	-0.055 (-1.5)
service	0.112 (1.1)	0.083 (0.7)	0.058 (1.0)	0.050 (0.8)	-0.002 (-0.1)
ssc	0.314 (1.4)	0.315 (1.4)	0.116 (1.0)	0.136 (1.0)	-0.006 (-0.1)
welfare	0.531 (3.1)	0.192 (0.9)	0.236 (2.4)	0.188 (1.6)	-0.179 (-2.5)
			ρ: 0.99 (44)	σ: 1.30 (4.6)	t_2 : 0.68 (11)
log-like.	-459.141		-451.882		-493.428
test-stat (pv)	7.563 (0.023)		1.601 (0.449)		

In multinomial choice with three alternatives to choose from, imagine three utility levels s_1, s_2, s_3 and three equations for these; let u_1, u_2, u_3 denotes the error terms in the equations. The alternative with the maximum utility will be chosen; e.g., 3 is chosen if $s_3 > s_1$ and $s_3 > s_2$. But this ordering does not chance when s_1 is subtracted from all terms: $s_3 - s_1 > 0$ and $s_3 - s_1 > s_2 - s_1$. Hence, the three equations get reduced to two equations with one alternative becoming the normalizing ("anchoring") one; we use Perot as the normalizing alternative.

In estimating the parameters in the two equations, only two error terms $u_3 - u_1$ and $u_2 - u_1$ appear. In MNL, no parameter for the error term covariance matrix is estimated, thanks to the assumption that u_1, u_2, u_3 are iid (independent and identically distributed). In MNP, no restriction such as iid is imposed; as the consequence, two parameters for $u_3 - u_1$ and $u_2 - u_1$ should be estimated: one is the correlation ρ between the two error terms, and the other is the ratio σ of the standard deviations of the two error terms.

When we applied MNP, the log-likelihood function kept increasing: ρ approached 1 for a given σ, although estimating σ is not a problem for a given ρ. Despite this problem, however, the signs of the estimates hardly changed for a given σ. Table 2 presents MNL and MNP with $\sigma = 1.3$ and $\rho = 0.99$, which is the best we could obtain. The source of the degeneracy problem is not clear. Keane (1992) shows that the MNP log-likelihood is rather flat along the direction of σ and ρ unless there are significant "alternative-varying" variables. In our data, there is only one alternative-varying (i.e., candidate-specific) variable: ideology of the candidates. The MNL column in Table 2 shows that only the Clinton ideology is significant, whereas the MNP shows that none of the candidate ideology variables is significant. This lack of significance in the alternative-specific variables may be a reason for the MNP degeneracy problem.

Another reason for the degeneracy may be that there actually is only one error term involved, not two, in the choice process as $\rho \simeq 0.99$ in MNP suggests. Since MNP is not compatible with one error term, we estimated 'ordered probit' model taking the three alternatives (Clinton 1, Perot 2, and Bush 3) as ordered discrete response (ODR). In the ODR model, the underlying continuous response variable, say y^*, is tendency to vote for Bush; if this is weak, Clinton is chosen; if this is strong, Bush is chosen; if neither weak nor strong, Perot is chosen. This setup may not be too far-fetched as can be seen in Table 3 for the candidate ideology placement proportion in 1992.

TABLE 3: Candidate Ideology Placement							
	1	2	3	4	5	6	7
Perot	0.055	0.055	0.081	0.501	0.115	0.100	0.094
Clinton	0.065	0.224	0.234	0.379	0.058	0.030	0.010
Bush	0.022	0.029	0.048	0.314	0.159	0.349	0.078

In Table 3, the most voters put Perot on 4,5,6 (4 being the largest), Clinton on 2,3,4 (4 being the largest), and Bush on 4,5,6 (6 being the largest); this corroborates the ordering in ODR. The last column of Table 2 shows the result of ODR estimation, where t_2 in the ODR column is for the threshold between Perot and Bush.

In Subsection 2.3, we will focus on the MNL results to interpret the estimates, because MNP has the degeneracy problem and ODR is not quite appropriate for multinomial choice. Despite this, however, the signs of significant variables mostly agree across MNL and MNP in Table 2. When we tried in-sample prediction (each person's choice probability for each candidate is computed and the person is predicted to vote for the candidate with the highest probability), all three estimators returned similar prediction performance: 0.711 for MNL, 0.720 for MNP, and 0.703 for ODR.

A multinomial choice estimator falling in between MNL and MNP in terms of its generality is 'nested multinomial logit' (NEST). Roughly speaking, NEST allows correlation in one pair among u_1, u_2, u_3. For our data, this means pairing Perot either with Clinton or Bush (but not both); but neither pairing looks attractive. Due to pairing, NEST has one more parameter than MNL whereas MNP has two more than MNL. Although not reported here, we also tried the in-sample prediction using NEST: paring Perot and Clinton together yielded 0.714, while pairing Perot and Bush yielded 0.722. Hence, if we go for in-sample prediction performance, the second NEST does tiny bit better than the almost degenerate MNP, which is hardly exciting for NEST. Both NEST and MNP will not be used any further in our empirical analysis.

Finally, at the bottom of Table 2, some test statistics are given. They have to do with "long and short parameterizations" (explained in Section 4); they will appear later when we point out mistakes in Alvarez and Nagler (1995,1998b). The test with MNL rejects the short (restrictive) parametrization while the test with MNP does not; again, we would trust the MNL finding more.

2.3 Average Effects of Regressors

Table 4 repeats the MNL results of Table 2 in the first two columns. In the last two columns, the 'average effects' of regressors are presented. Averaging the individual choice probability derivatives for a regressor across individuals, we find the average effect of the regressor, which is the average change in the choice probably as the regressor changes by one unit. The analytic forms for the average derivatives are shown later in Section 4. For the female dummy, this is the effect of "going from male (0) to female (1)"; for income, this is the effect of moving only one category up. In the last two columns of Table 4, the numbers corresponding to significant estimates in the MNL columns are marked with *. As we go over the estimates in the remainder of this subsection, we will refer to some of the frequently echoed expressions in the news media for the 1992 election, which are put in " ".

One category increment (out of seven) in Clinton ideology increases the Clinton choice probability by 0.052 and decreases the Bush choice probability by 0.027; "moving to the center" seems to have worked. The two intercepts are significantly negative with substantial magnitude, which means that much of Perot's attraction is not explained by the regressors. So long as one feels that most important factors are covered by the regressors, this may be attributed to "voters angry with politics as usual". This is, however, not "anti-incumbent", because Clinton loses far more than Bush, judging from the intercept magnitudes. The notion that Bush lost because of Perot does not seem to be true—more on this in the next section.

In the MNL column for Clinton, the age function slopes upward, and probably downward beyond age 62.6 if we take the quadratic term into account; the insignificant age function for Bush shows the opposite pattern. In the average age effect, age has a positive effect 0.021 on Clinton; if we takes the quadratic term into account, since one unit increase in $age^2/100$ happens at $age = 50$ due to the derivative $age \times 2/100$, the average effect of age on Clinton changes from positive to negative around $age = 50$.

Education works strongly for Bush; one category increment out of seven increases the Bush choice probability by 0.045. If one feels financially better off, that increases the Bush choice probability by 0.108. Although eco.-worse is not significant, the average effect -0.094 of national economy is close to that of personal finance. This mere 10% magnitude of national economy and personal finance does not justify the slogan "it's the economy, stupid". The respondent ideology also has strong effects, -0.118 and 0.104, for Clinton and Bush.

Low income group prefer Clinton; the rather small magnitude -0.006 for income is because income has 24 categories. Party affiliation has a substantial effect (0.169) on the Bush choice probability.

TABLE 4: Average Effect with MNL

	Clinton (MNL)	Bush (MNL)	Clinton	Bush
ideology-P	0.008 (0.1)		-0.001	-0.000
ideology-C	0.349 (3.7)		0.052*	-0.027*
ideology-B		0.115 (1.2)	-0.009	0.015
1	-7.57 (-3.7)	-5.81 (-2.5)	-0.679*	-0.179*
age	0.119 (2.2)	-0.034 (-0.6)	0.021*	-0.014
$age^2/100$	-0.095 (-1.7)	0.068 (1.1)	-0.020	0.017
eco.-worse	0.359 (1.1)	-0.492 (-1.6)	0.093	-0.094
education	0.085 (0.9)	0.384 (3.4)	-0.018	0.045*
fin.-better	-0.092 (-0.3)	0.756 (2.1)	-0.073	0.108*
fin.-worse	0.079 (0.2)	0.597 (1.6)	-0.035	0.073
ideology-R	-0.542 (-3.2)	0.461 (2.3)	-0.118*	0.104*
income	-0.055 (-2.0)	-0.031 (-1.0)	-0.006*	0.000
party	-0.207 (-0.9)	1.145 (4.0)	-0.121	0.169*
returning	0.078 (0.2)	-0.124 (-0.3)	0.021	-0.023
female	0.370 (1.4)	0.639 (2.2)	0.005	0.056*
east	0.689 (1.8)	0.622 (1.5)	0.055	0.029
south	0.425 (1.3)	0.368 (1.0)	0.035	0.016
west	0.268 (0.7)	0.413 (1.0)	0.008	0.034
abortion	-0.072 (-0.3)	-0.732 (-3.1)	0.047	-0.092*
black	0.072 (0.8)	-0.113 (-1.2)	0.020	-0.021
defense	-0.098 (-0.9)	0.157 (1.3)	-0.027	0.029
health	-0.108 (-1.4)	-0.239 (-2.9)	0.003	-0.023*
job	0.151 (1.6)	0.036 (0.4)	0.020	-0.007
service	0.112 (1.1)	0.083 (0.7)	0.010	0.002
ssc	0.314 (1.4)	0.315 (1.4)	0.022	0.017
welfare	0.531 (3.1)	0.192 (0.9)	0.065*	-0.016

Females prefer Bush with the effect being 0.056, which may be a surprise, given the perception that Bush is unpopular among women. Abortion has a big negative impact (-0.092) on Bush. This may explain the surprising female effect on Bush: unpopularity of Bush among women is because of the abortion issue, not of gender itself. One category increase out of seven in health insurance lowers the Bush choice probability by 0.023, whereas one category increase out of seven in welfare raises the Clinton choice probability by 0.065.

3 Strategic Voting

Having analyzed the US presidential election with 1992 data and MNL, we examine strategic voting in this section. Before we discuss strategic voting, however, it is helpful to see how people voted across the two elections in 1992 and 1996.

Table 5 is the breakdown of votes of 296 individuals who showed their choices in both years; define the nine small-print cells in the middle as 'cell 11' through 'cell 33'. The numbers in parentheses are the proportions relatives to the 1992 total for each candidate; e.g., 0.300 in cell 11 is 0.051/0.169 where 0.169 is the Perot 1992 total proportion. From Table 5, first, the presence of Perot in 1996 is meager, getting just 6.1%. Second, among the 1992 Perot voters, about one third defected to Clinton in 1996 and another one third defected to Dole. Third, only 7.5% of the 1992 Clinton voters defected to Dole in 1996, whereas twice as many Bush voters defected to Clinton in 1996. This shows that Dole lost not because of Perot, but because he lost many Republican supporters to Clinton.

TABLE 5: Voting Transition Matrix 92-96

	Perot 96	Clinton 96	Dole 96	1992
Perot 92	0.051 (0.300)	0.057 (0.340)	0.061 (0.360)	0.169 (1)
Clinton 92	0.003 (0.008)	0.412 (0.917)	0.034 (0.075)	0.449 (1)
Bush 92	0.007 (0.018)	0.061 (0.159)	0.314 (0.823)	0.382 (1)
1996	0.061	0.530	0.409	1.000

Turning to strategic voting behavior, using the 1992 MNL estimates, first we predicted how people would vote in 1996. Among those who were predicted to vote for Perot in 1996,

there were 37 people who did not vote for Perot in 1996. For us, these 37 people are the strategic voters, who take 11.8% of 296 voters for both elections or 8.1% of 456 voters for election 1996.

TABLE 6: Strategic Voters vs. the Rest

	Predicted 96 vs. 96			Bootstraped 96 vs. 96			Actual 92 vs. 96		
age	51.6	40.2	(4.38)	51.5	40.1	(4.51)	47.2	49.8	(-0.95)
eco.-worse	0.13	0.16	(-0.52)	0.13	0.18	(-0.77)	0.11	0.11	(0.06)
education	4.49	3.92	(2.51)	4.50	3.73	(3.65)	4.49	4.71	(-0.81)
fin.-better	0.41	0.54	(-1.50)	0.42	0.49	(-0.75)	0.54	0.42	(1.38)
fin.-worse	0.22	0.16	(0.81)	0.21	0.18	(0.44)	0.23	0.20	(0.34)
ideology-R	2.27	2.38	(-0.76)	2.26	2.49	(-1.56)	2.60	2.20	(2.90)
income	16.3	19.2	(-4.77)	16.3	19.0	(-4.20)	18.9	17.5	(2.04)
party	1.03	0.87	(2.29)	1.03	0.88	(1.97)	1.06	1.00	(0.51)
returning	0.93	0.96	(-0.44)	0.93	0.94	(-0.23)	1.00	0.98	(2.25)
female	0.55	0.38	(2.02)	0.55	0.36	(2.09)	0.43	0.53	(-1.11)
east	0.17	0.054	(2.81)	0.17	0.03	(4.01)	0.17	0.15	(0.27)
south	0.38	0.24	(1.75)	0.37	0.24	(1.65)	0.29	0.36	(-0.94)
west	0.22	0.11	(1.96)	0.22	0.12	(1.54)	0.20	0.21	(-0.15)
abortion	2.30	2.32	(-0.22)	2.31	2.24	(0.58)	2.29	2.32	(-0.29)
black	3.19	2.70	(2.13)	3.19	2.70	(1.99)	2.89	3.15	(-0.88)
defense	4.00	3.89	(0.54)	4.01	3.82	(0.90)	4.37	3.97	(1.92)
health	3.77	3.70	(0.25)	3.77	3.70	(0.24)	3.74	3.74	(0.02)
job	3.43	2.68	(3.13)	3.43	2.64	(3.00)	2.71	3.37	(-2.70)
service	3.70	3.32	(1.72)	3.70	3.27	(1.82)	2.97	3.68	(-3.29)
ssc	5.88	5.32	(3.17)	5.88	5.27	(3.25)	5.66	5.81	(-0.92)
welfare	5.76	5.30	(2.05)	5.76	5.27	(1.93)	5.69	5.75	(-0.41)

To see whether the strategic voters are systematically different from the other people, we compare the mean of the regressors other than the candidate ideologies in Table 6. In the column 'Predicted 96 vs. 96' of Table 6, the strategic voters are compared with all the other voters in 1996: the first number is the group mean of the strategic voters, the second is the group mean of the others, and the third is the asymptotic test statistic for the group mean

difference; here, only those with the response variable available in 1996 are compared (456 in total). We can see the *profile of strategic voters: older, more than highschool-educated, relatively poorer, neutral in party affiliation, relatively more female, and more liberal in government policy involvements.* Either major candidate could have benefitted—particularly the loser—by focusing the campaign more on this group of strategic voters. This finding could not have been obtained, had we used aggregate, non-individual, data.

One may generalize this finding to say that people with those attributes may vote strategically in other elections as well. This kind of 'external validity' is one of the reasons to examine various empirical works. After all, if findings of an empirical study have only 'internal validity'—that is, applicable only to the particular data—then, what one can learn in science will be fairly limited. But such a generalization should be done with caution because strategic voting may interact with the characteristics of the alternatives. That is, if the candidate characteristics change, then people with different attributes may vote strategically.

There are two limitations in our approach to identifying strategic voters. Recall $s_{ijt} = w'_{ijt}\beta + u_{ijt}$ for the latent utility for person i, alternative j, and time t. One limitation is that the estimation error $b_N - \beta$ of the first-stage estimation is not accounted for in the second stage prediction. The other is that the error terms u_{ij2}, $j = 1, 2, 3$, at the second period 1996 are not taken into account; namely, we looked only at $w'_{ij2}\beta$, $j = 1, 2, 3$, to predict the choice at the second period, despite that the actual choice is made by $w'_{ij2}\beta + u_{ij2}$, $j = 1, 2, 3$. Although it is not clear how the two limitations can be dealt with, we went about the first using bootstrap in the following way. First, a bootstrap pseudo sample of size N_1 is generated from the 1992 data to get a pseudo estimate b^*_N for β. Second, 1996 choice is predicted with b^*_N for each voter. Third, if the voter is predicted to vote for Perot in 1996 based on b^*_N but did not in fact, then the voter gets a "score 1". We repeat the three steps for 500 times, and collect the voters with the score (averaged over 500 times) higher than 0.5. This process resulted in 33 strategic voters and they rendered the second column of Table 6 with "Boostrapped 96 vs. 96". The second column hardly differs from the first, suggesting that the error $b_N - \beta$ may not matter much.

Voting for Perot in 1992 but not in 1996 (there are 35 such individuals who are in cell 12 and cell 13 of Table 5) does not necessarily mean strategic voting, because the available alternatives are different between 1992 and 1996; call the 35 people "*pseudo* strategic voters". Comparing different elections to verify strategic voting always has this problem. It is im-

portant to see that our preceding approach of predicting the vote with MNL and panel data controls for this "menu" change. The last column of Table 6 compares the pseudo strategic voters with the other voters; here, only those with the response variable available in both 1992 and 1996 can be compared (296 in total). We can see that the profile of the pseudo strategic voters: more conservative in ideology, relatively richer, and more conservative in government service involvements. *This profile of the pseudo strategic voters is much different from that of the actual strategic voters.*

TABLE 7: Perot to Clinton vs. Perot to Dole

age	38.6	41.8	(-0.63)	east	0.11	0.00	(1.46)
eco.-worse	0.00	0.32	(-2.88)	south	0.33	0.16	(1.23)
education	4.33	3.53	(1.94)	west	0.00	0.21	(-2.19)
fin.-better	0.89	0.21	(5.53)	abortion	2.39	2.26	(0.60)
fin.-worse	0.056	0.26	(-1.76)	black	2.94	2.47	(1.08)
ideology-R	1.89	2.84	(-3.93)	defense	4.00	3.79	(0.54)
income	19.0	19.4	(-0.34)	health	4.67	2.80	(4.51)
party	0.83	0.90	(-0.44)	job	3.17	2.21	(2.17)
returning	0.94	0.95	(-0.04)	service	3.78	2.90	(2.28)
female	0.28	0.47	(-1.22)	ssc	5.39	5.26	(0.37)
				welfare	4.94	5.63	(-1.54)

Returning back to the 37 strategic voters, there are 18 individuals who voted for Clinton and 19 who voted for Dole. In Table 7, the two groups are compared. As one could have expected, the incumbent Clinton voters feel better about the nation's economy and the personal finance, and relatively liberal in ideology and government policy involvements.

4 Review, Errors in Using MNP, and Average Effect for MNL

In the preceding sections, the US presidential election has been analyzed, and we obtained the profile of strategic voters. In this section, we do three things. First, MNL and MNP are reviewed, drawing on Lee and Kim (2007). Second, critical mistakes in the literature of applying MNP to public choices are pointed out. Third, the analytic form of average derivatives for MNL that was used in Section 2 is presented.

4.1 Review on MNL and MNP

Multinomial choice models have been used extensively in many disciplines of science: transport mode choice, location choice, job choice, and automobile choice, just to name a few. The most popular estimation method for multinomial choice has been multinomial logit (MNL; McFadden (1974)) which behaves well computationally. But MNL has the well known critical shortcomings of requiring iid errors in the latent utility equations and the "independence of irrelevant alternatives" resulting from the multinomial logit probability form. Nested logit (NEST; McFadden (1981)) has been proposed to avoid the two problems of MNL to some extent by allowing correlations among the error terms in the same "branch" in a sequential "choice trees". More general than MNL and NEST is multinomial probit (MNP; Hausman and Wise (1978)) that allows, at least in principle, arbitrary dependence among all error terms.

Despite the flexibility in MNP, one of the major shortcomings for MNP has been multidimensional numerical integration for computing the choice probabilities. Although simulated method of moment (McFadden (1989)) solved this problem, MNP hardly converges well with an arbitrary error term covariance matrix even with four alternatives; e.g., Green makes note of this in his LIMDEP manual, and our experience has been the same as well. It is in fact possible for MNP to find a spurious maximum due to extra errors introduced by simulation. See Train (2003) and the references therein for more on multinomial choice estimation.

Suppose person i, $i = 1, ..., N$, gets linear utility or satisfaction s_{ij} with the alternative j, $j = 1, ..., J$, and the utility equation is

$$s_{ij} = x'_{ij}\delta_j + z'_i\eta_j + u_{ij}, \tag{4.1}$$

where x_{ij} is a $k_x \times 1$ regressor vector changing across i and j, z_i is a $k_z \times 1$ regressor vector changing only across i, and u_{ij} is a continuously distributed error term; δ_j and η_j are unknown parameter vectors. Let the first component of z_i be one for all i, so that the first component of η_j is the "alternative-specific intercept" capturing contributions from the variables varying only across j.

Person i will choose j iff

$$s_{ij} > s_{im}, \quad \text{for all } m \neq j \quad \Leftrightarrow \quad s_{ij} - s_{i1} > s_{im} - s_{i1}, \quad \text{for all } m \neq j; \tag{4.2}$$

this is a location normalization using the first alternative as the base. From this, we use not

(4.1) but

$$s_{ij} - s_{i1} = -x'_{i1}\delta_1 + x'_{ij}\delta_j + z'_i(\eta_j - \eta_1) + u_{ij} - u_{i1}, \quad j = 2, ..., J, \tag{4.3}$$

with the regression parameters (PL in the following stands for "parameterization long")

$$\delta_1, ..., \delta_J,\ \eta_2 - \eta_1, ..., \eta_J - \eta_1. \tag{PL}$$

If

$$\delta_j = \delta \quad \text{for all } j,$$

then, instead of (4.3), we get

$$s_{ij} - s_{i1} = (x_{ij} - x_{i1})'\delta + z'_i(\eta_j - \eta_1) + u_{ij} - u_{i1}, \quad j = 2, ..., J, \tag{4.4}$$

with the regression parameters (PS in the following stands for "parameterization short")

$$\delta,\ \eta_2 - \eta_1, ..., \eta_J - \eta_1. \tag{PS}$$

What is actually identified is, however, not PL nor PS, but a scalar multiple of PL or PS. We show this point in the following.

Consider PL. Let 0_k denote the $k \times 1$ zero vector. Define β, $v_{i2}, ..., v_{iJ}$, and $w_{i2}, ..., w_{iJ}$ as

$$\begin{aligned} \beta &\equiv (\delta'_1, ..., \delta'_J,\ \eta'_2 - \eta'_1, ..., \eta'_J - \eta'_1)', \quad v_{ij} \equiv u_{ij} - u_{i1}, \quad j = 2, ..., J, \\ w_{ij} &\equiv (-x'_{i1}, 0'_{k_x}, ..., 0'_{k_x}, x'_{ij}, 0'_{k_x}, ..., 0'_{k_x},\ 0'_{k_z}, ..., 0'_{k_z}, z'_i, 0'_{k_z}, ..., 0'_{k_z})', \end{aligned}$$

where the numbers of 0_{k_x} and 0_{k_z}'s about x'_{ij} and z'_i in w_{ij} are chosen to get

$$-x'_{i1}\delta_1 + x'_{ij}\delta_j + z'_i(\eta_j - \eta_1) + u_{ij} - u_{i1} = w'_{ij}\beta + v_{ij}; \tag{4.5}$$

recall (4.3). For instance, with $J = 3$,

$$\begin{aligned} \beta &\equiv (\delta'_1, \delta'_2, \delta'_3,\ \eta'_2 - \eta'_1,\ \eta'_3 - \eta'_1)', \quad v_{i2} \equiv u_{i2} - u_{i1}, \quad v_{i3} = u_{i3} - u_{i1}, \\ w_{i2} &\equiv (-x'_{i1}, x'_{i2}, 0'_{k_x},\ z'_i, 0'_{k_z})', \quad w_{i3} \equiv (-x'_{i1}, 0'_k, x'_{i3},\ 0'_k, z'_i)'. \end{aligned}$$

Then

$$\begin{aligned} -x'_{i1}\delta_1 + x'_{i2}\delta_2 + z'_i(\eta_2 - \eta_1) + u_{i2} - u_{i1} &= w'_{i2}\beta + v_{i2}, \\ -x'_{i1}\delta_1 + x'_{i3}\delta_3 + z'_i(\eta_3 - \eta_1) + u_{i3} - u_{i1} &= w'_{i3}\beta + v_{i3}. \end{aligned} \tag{4.6}$$

Define y_{ij} and w_i as

$$y_{ij} \equiv 1 \text{ if person } i \text{ chooses } j, \text{ and 0 otherwise}, \qquad w_i \equiv (x'_{i1}, ..., x'_{iJ}, z'_i)'. \tag{4.7}$$

Omitting i to simplify notations, we have

$$\begin{aligned}
P(y_1 &= 1|w) = P(s_2 - s_1 < 0,\ ...,\ s_J - s_1 < 0|w) = P(v_2 < -w'_2\beta,\ ...,\ v_J < -w'_J\beta|w), \\
P(y_2 &= 1|w) = P(s_2 - s_1 > 0,\ s_3 - s_2 < 0,\ ...,\ s_J - s_2 < 0|w) \\
&= P(v_2 > -w'_2\beta,\ v_3 - v_2 < (w_2 - w_3)'\beta,\ ...,\ v_J - v_2 < (w_2 - w_J)'\beta|w), \\
&\ \ \vdots \\
P(y_J &= 1|w) = P(s_J - s_1 > 0,\ ...,\ s_J - s_{J-1} > 0|w) \\
&= P(v_J > -w'_J\beta,\ ...,\ v_J - v_{J-1} > (w_{J-1} - w_J)'\beta|w).
\end{aligned}$$

Divide all events by an unknown positive constant σ to get

$$\begin{aligned}
P(y_1 = 1|w;\gamma) &= P(\frac{v_2}{\sigma} < -w'_2\frac{\beta}{\sigma},\ ...,\ \frac{v_J}{\sigma} < -w'_J\frac{\beta}{\sigma}|w), \\
P(y_2 = 1|w;\gamma) &= P(\frac{v_2}{\sigma} > -w'_2\frac{\beta}{\sigma},\ \frac{v_3 - v_2}{\sigma} < (w_2 - w_3)'\frac{\beta}{\sigma}, \cdots,\ \frac{v_J - v_2}{\sigma} < (w_2 - w_J)'\frac{\beta}{\sigma}|w), \\
&\ \ \vdots \\
P(y_J = 1|w;\gamma) &= P(\frac{v_J}{\sigma} > -w'_J\frac{\beta}{\sigma},\ ...,\ \frac{v_J - v_{J-1}}{\sigma} > (w_{J-1} - w_J)'\frac{\beta}{\sigma}|w)
\end{aligned} \tag{4.8}$$

where γ denotes the identified parameters which depends on the probability distribution of $(v_2, ..., v_J)|w$, which in turn depends on the distribution of $(u_1, u_2, ..., u_J)|w$. A maximum likelihood estimator (MLE) is obtained by maximizing $\sum_{i=1}^{N}\sum_{j=1}^{J} y_{ij} \ln\{P(y_{ij} = 1|w_i;\gamma)\}$ for γ.

MNL is obtained if

$\frac{u_1}{\sigma},\ ...,\ \frac{u_J}{\sigma}$ are independent of w and *iid* with Type I extreme value distribution.

With this, γ becomes β/σ; β is identified only up to the unknown scalar σ, but the ratios of the components of β are fully identified. There is no error term distribution parameter to estimate in MNL. MNP is obtained if

$u_1, ..., u_J$ are independent of w and jointly normal with mean zero and covariance C_u.

This implies the independence of $(v_2, ..., v_J)$ from w and a joint normality for $(v_2, ..., v_J)$. Even if $u_1, ..., u_J$ are independent ($\Leftrightarrow C_u$ is diagonal), $v_2, ..., v_J$ are positively correlated, for they

share $-u_1$. If $u_1, ..., u_J$ are iid ($\Leftrightarrow C_u$ is a diagonal matrix with the same element along the diagonal), then MNP is analogous to MNL in terms of the identified parameters: $\gamma = \beta/\sigma$, with σ being $SD(v_2)$ as shown in the following.

For MNP, let $\sigma = SD(v_2)$, and let $C_{v\sigma}$ denote the covariance matrix for $(v_2/\sigma, ..., v_J/\sigma)$. What is identified is then

$$\begin{aligned} \gamma &= (\frac{\beta'}{\sigma},\ \sigma_{jm},\ j, m = 2, ..., J,\ j < m,\ \sigma_j^2,\ j = 3, ..., J)' \\ &\Leftrightarrow (\frac{\beta'}{\sigma},\ \rho_{jm},\ j, m = 2, ..., J,\ j < m,\ \sigma_j,\ j = 3, ..., J)' \end{aligned}$$

where $\sigma_{jm} \equiv COV(v_j/\sigma, v_m/\sigma)$, $\sigma_j^2 \equiv V(v_j/\sigma)$, and $\rho_{jm} \equiv COR(v_j/\sigma, v_m/\sigma)$. For instance, with $J = 3$,

$$\gamma = (\frac{\beta'}{\sigma}, \sigma_{23}, \sigma_3^2)' \Longleftrightarrow (\frac{\beta'}{\sigma}, \rho_{23}, \sigma_3)'. \tag{4.9}$$

Train (2003, p.106) shows essentially the same identification result (for $J = 4$).

As for PS, we need to redefine β and w_{ij}'s as

$$\beta \equiv (\delta',\ \eta_2' - \eta_1', ..., \eta_J' - \eta_1')', \quad w_{ij} \equiv (x_{ij}' - x_{i1}',\ 0_{k_z}', ..., 0_{k_z}', z_i', 0_{k_z}', ..., 0_{k_z}')'$$

where the number of 0_{k_z}'s about z_i' is chosen to get

$$(x_{ij} - x_{ij})'\delta + z_i'(\eta_j - \eta_1) + u_{ij} - u_{i1} = w_{ij}'\beta + v_{ij};$$

recall (4.4). For instance, with $J = 3$, β and w_{ij} are

$$\beta \equiv (\delta',\ \eta_2' - \eta_1', \eta_3' - \eta_1')', \quad w_{i2} \equiv (x_{i2}' - x_{i1}',\ z_i', 0_{k_z}')', \quad w_{i3} \equiv (x_{i3}' - x_{i1}',\ 0_{k_z}', z_i')';$$

identification for the error term distribution parameters does not change, while β now has far fewer parameters compared with PL.

As mentioned already, MNP does not converge well, and one way to alleviate this problem for $J = 3$ is as follows; cases with $J > 3$ can be dealt with doing analogously. Some grid points are selected for ρ_{23} in $(-1, 1)$ and for σ_3 around 1. For each grid point of ρ_{23} and σ_3, MNP is obtained for the remaining parameters. Finally, the maximized log-likelihood values are compared across the grid points to get the desired MNP. If this kind of grid search fails to converge, then one may impose a plausible restriction such as $\rho_{23} = 0.5$ or $\sigma_3 = 1$. In either case, there is still one more parameter than in MNL, and thus MNP is still more flexible than MNL; note, however, that neither of MNL and MNP nests the other as a special case, for the error term distribution differs between MNL and MNP.

If one has a panel data, panel multinomial choice estimators are applicable. Lee (2002) explains panel multinomial logit in detail. In fact, we applied panel multinomial logit to our data, but found no significant variable because temporal variation of the regressors were too small. Panel multinomial probit is possible in principle, but its implementation is troublesome, because there are too many parameters to estimate in the error term covariance matrix across both alternatives and time.

So far, we discussed constant coefficient models, but it is also possible to allow the model coefficients to be random, which, however, leads to heteroskedastic error terms and far more parameters to estimate. For instance, suppose we use PS with scalar x_{ij} and z_i to have

$$s_{ij} = x_{ij}(\delta + e) + z_i(\eta_j + \varepsilon_j) + u_{ij} = x_{ij}\delta + z_i\eta_j + (x_{ij}e + z_i\varepsilon_j + u_{ij}),$$

where e and ε_j's are random with zero means. In this case, under the assumption that $(e, \varepsilon_1, ..., \varepsilon_J, u_1, ..., u_J)$ is jointly normal independently of $(x_{i1}, ..., x_{1J}, z_i)$, a MNP is obtained where the error term $x_{ij}e + z_i\varepsilon_j + u_{ij}$, $j = 1, ..., J$, are heteroskedastic and there are more parameters to estimate due to more random terms entering the linear utility equation. Hausman and Wise (1978) explored this kind of random coefficient models for MNP. As in constant coefficient models, they looked at two types of MNP's: "covariance MNP" where the correlations of the error terms in $s_{i1}, ..., s_{iJ}$ are not constrained, and "independence MNP" where the correlations are assumed to be zero. Note that correlations of the error terms in $s_{i1}, ..., s_{iJ}$ can occur through e, even if $u_{i1}, ...u_{iJ}$ are uncorrelated. MNP random coefficient models are computationally troublesome, because too many error term (co-) variances have to be estimated. A better alternative for random coefficients is 'Mixed MNL' in McFadden and Train (2000).

4.2 Errors in Using MNP

In political science, MNP has been applied to the 1992 and 1996 US presidential elections by Alvarez and Nagler (1995, 1998a, 1998b). These studies, however, have critical errors that non-identified parameters are estimated. The identification problem with MNP is not uncommon in the literature: Keane (1992, p.194) cites a working paper by Bunch and Kitamura stating that nearly half of the existing applications of MNP have used formally non-identified models; see Train (2003, p.105) for more on this. In the following, we show the errors in detail.

Alvarez and Nagler (1995) did MNP with PS and $J = 3$. In their Table 3, they show *three* estimates for the error term covariance matrix. This is simply impossible as (4.9) shows. In Footnote 20 of Alvarez and Nagler (1998a), they acknowledge this problem; also there, they state that they use PS to avoid the "fragile identification problem" of PL, but there is nothing fragile about PL, which is not a source of identification problems in MNP. Table 6 of Alvarez and Nagler (1998a) and Table 1 of Alvarez and Nagler (1998b) show *two correlation estimates* for the error term covariance matrix. The number two is right, but as (4.9) shows, *only one correlation and one standard deviation* are identified: recall that, with $J = 3$, there are only two error terms $v_2 = u_2 - u_1$ and $v_3 = u_3 - u_1$, and consequently only one correlation $\rho_{23} = COR(v_2, v_3)$. Alvarez and Nagler (2000) repeat the same mistake, which is their "implementation error" as alluded to in Section 1.

In Footnote 20 of Alvarez and Nagler (1998a), they state that they follow Hausman and Wise (1978) specifications, but Hausman and Wise (1978) never used the specification with two correlation coefficients in a constant coefficient model as Alvarez and Nagler (1998a) did. When Hausman and Wise (1978) allowed for multiple correlations, it was for random coefficient models of the type in the preceding paragraph. Alvarez and Nagler (1998b, p.1354) state that their 1996 election results are similar to those for the 1992 election: since their 1992 election results are not trustworthy, the statement further shows that their 1996 election results are not trustworthy either.

When non-identified parameters are estimated, the algorithm does not converge; it is puzzling how they came up with the numbers presented in their papers. Footnote 22 in Alvarez and Nagler (1998a) states "There are three possible disturbance covariance matrices which can be considered if two correlations between disturbances are estimated", which also illustrates their false parametrization.

4.3 Analytic Form of Average Derivatives for MNL

Recall w_{i2} and w_{i3}, and define

$$\alpha \equiv \frac{\beta}{\sigma}, \quad \frac{\delta_j}{\sigma} \equiv \bar{\delta}_j,\ j = 1, 2, 3, \quad \eta_j - \eta_1 \equiv \bar{\eta}_j,\ j = 2, 3.$$

In MNL, the choice probabilities P_{i1}, P_{i2}, and P_{i3} are

$$\begin{aligned} P_{i1} &= \frac{1}{1+\exp(w_{i2}'\alpha)+\exp(w_{i3}'\alpha)}, \quad P_{i2} = \frac{\exp(w_{i2}'\alpha)}{1+\exp(w_{i2}'\alpha)+\exp(w_{i3}'\alpha)}, \\ P_{i3} &= \frac{\exp(w_{i3}'\alpha)}{1+\exp(w_{i2}'\alpha)+\exp(w_{i3}'\alpha)}. \end{aligned}$$

Even if a variable appears only in w_{i2}, its change influences both P_{i2} and P_{i3}, because of the denominator in P_{i2} and P_{i3}.

Define $S_i \equiv 1+\exp(w_{i2}'\alpha)+\exp(w_{i3}'\alpha)$ and observe

$$\begin{aligned} \partial P_{i2}/\partial x_{i1} &= [S_i \exp(w_{i2}'\alpha)\cdot(-\bar{\delta}_1) - \exp(w_{i2}'\alpha)\{\exp(w_{i2}'\alpha)+\exp(w_{i3}'\alpha)\}(-\bar{\delta}_1)]/S_i^2 \\ &= P_{i2}(-\bar{\delta}_1) - P_{i2}(P_{i2}+P_{i3})(-\bar{\delta}_1) = -\bar{\delta}_1 P_{i2}P_{i1}; \\ \partial P_{i2}/\partial x_{i2} &= [S_i \exp(w_{i2}'\alpha)\cdot\bar{\delta}_2 - \exp(w_{i2}'\alpha)\exp(w_{i2}'\alpha)\cdot\bar{\delta}_2]/S_i^2 \\ &= P_{i2}\bar{\delta}_2 - P_{i2}^2\bar{\delta}_2 = \bar{\delta}_2 P_{i2}(1-P_{i2}); \\ \partial P_{i3}/\partial x_{i2} &= [-\exp(w_{i3}'\alpha)\exp(w_{i2}'\alpha)\cdot\bar{\delta}_2]/S^2 = -\bar{\delta}_2 P_{i3}P_{i2}; \end{aligned}$$

∂P_{i1} is omitted, because a derivative for P_{i1} can be found from those for P_{i2} and P_{i3} owing to the restriction $P_{i1}+P_{i2}+P_{i3}=1$. Doing "symmetrically", we find

$$\partial P_{i3}/\partial x_{i1} = -\bar{\delta}_1 P_{i3}P_{i1}, \quad \partial P_{i3}/\partial x_{i3} = \bar{\delta}_3 P_{i3}(1-P_{i3}), \quad \partial P_{i2}/\partial x_{i3} = -\bar{\delta}_3 P_{i2}P_{i3}.$$

Also observe

$$\begin{aligned} \partial P_{i2}/\partial z_i &= [S_i \exp(w_{i2}'\alpha)\cdot\bar{\eta}_2 - \exp(w_{i2}'\alpha)\{\exp(w_{i2}'\alpha)\cdot\bar{\eta}_2 + \exp(w_{i3}'\alpha)\cdot\bar{\eta}_3\}]/S_i^2 \\ &= P_{i2}\bar{\eta}_2 - P_{i2}(P_{i2}\bar{\eta}_2 + P_{i3}\bar{\eta}_3) = P_{i2}(1-P_{i2})\bar{\eta}_2 - P_{i2}P_{i3}\bar{\eta}_3; \\ \partial P_{i3}/\partial z_i &= P_{i3}(1-P_{i3})\bar{\eta}_3 - P_{i3}P_{i2}\bar{\eta}_2. \end{aligned}$$

5 Conclusions

In this paper, we did three tasks. First, we analyzed the 1992 US presidential election to find the following variables relevant: respondent and candidate ideology, personal finance, age, education, income, sex, abortion stance, health insurance policy, and welfare program policy. Second, using panel data for 1992 and 1996, we derived the profile of strategic voters in 1996: average age of 52, more than highschool-educated, more female, and liberal in government policy involvements. Third, we reviewed multinomial choice estimators and showed

the critical errors in Alvarez and Nagler (1995, 1998a, 1998b,2000) in applying multinomial probit to elections. Doing the three tasks, we hope to have shown the reader that a proper use of multinomial choice estimators can lead to fruitful findings in analyzing voting behavior in elections with more than two parties.

REFERENCES

Abramson, P.R., J.H. Aldrich, P. Paolino, and D.W. Rohde, 1992, Sophisticated voting in the 1988 presidential primaries, American Political Science Review 86, 55-69.

Alvarez, R.M. and J. Nagler, 1995, Economics, issues and the Perot candidacy: voter choice in the 1992 presidential election, American Journal of Political Science 39, 714-744.

Alvarez, R.M. and J. Nagler, 1998a, When politics and models collide: estimating models of multiparty elections, American Journal of Political Science 42, 55-96.

Alvarez, R.M. and J. Nagler, 1998b, Economics, entitlements, and social issues: voter choice in the 1996 presidential election, American Journal of Political Science 42, 1349-1363.

Alvarez, R. M. and J. Nagler, 2000, A New approach for modelling strategic voting in multiparty elections, British Journal of Political Science 30, 57-75.

Blais, A. and R. Nadeau, 1996, Measuring strategic voting: a two-step procedure," Electoral Studies 15, 39-52.

Hausman, J.A. and D.A. Wise, 1978, A conditional probit model for qualitative choice: discrete decisions recognizing interdependence and heterogenous preferences, Econometrica 46, 403-426.

Galbraith, J. W. and N. C. Rae, 1989, A Test of the importance of tactical voting: Great Britain, 1987, British Journal of Political Science 19, 126-136.

Johnston, R. J. and C.J. Pattie, 1991, Tactical voting in Great Britain in 1983 and 1987: an alternative approach, British Journal of Political Science 21, 95-128.

Keane, M.P., 1992, A note on identification in the multinomial probit model, Journal of Business and Economic Statistics 10, 193-200.

Kousser, T., 2004, Retrospective Voting and Strategic Behavior in European Parliament Elections, Electoral Studies 23, 1-21.

Lee, M.J., 2002, Panel data econometrics: methods-of-moments and limited dependent variables, Academic Press.

Lee, M.J. and Y.S. Kim, 2007, Multinomial choice and nonparametric average derivatives, Transportation Research Part B: Methodological 41, 63-81.

McFadden, D., 1974, Conditional logit analysis of qualitative choice behavior, in Frontiers in Econometrics, edited by P. Zarembka, Academic Press.

McFadden, D., 1981, Econometric models of probabilistic choice, in Structural Analysis of Discrete Data, edited by C. Manski and D. McFadden, MIT Press.

McFadden, D., 1989, A Method of simulated moments for estimation of discrete response models without numerical integration, Econometrica 57, 995-1026.

McFadden, D. and K. Train, 2000, Mixed MNL models for discrete response, Journal of Applied Econometrics 15, 447-470.

Mueller, D.C., 2003, Public choice III, Cambridge University Press.

Niemi, R.G., G. Whitten, and M.N. Franklin, 1992, Constituency characteristics, individual characteristics and tactical voting in the 1987 British General election, British Journal of Political Science 22, 229-254.

Train, K.E., 2003, Discrete choice methods with simulation, Cambridge Univ. Press.

Warren, E.M., D.R. Kinder, and S.J. Rosenstone, 1999, National Election Studies 1996, Center for Political Studies, University of Michigan, U.S.A.