PREDICTING ELECTORAL COLLEGE VICTORY PROBABILITIES FROM STATE DATA

by

Ray C. Fair

November 2004

COWLES FOUNDATION DISCUSSION PAPER NO. 1496

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

YALE UNIVERSITY

Box 208281

New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

Predicting Electoral College Victory Probabilities from State Probability Data

Ray C. Fair*

November 2004

Abstract

A method is proposed in this paper for predicting Electoral College victory probabilities from state probability data. A "ranking" assumption about dependencies across states is made that greatly simplifies the analysis. The method is used to analyze state probability data from the Intrade political betting market. The Intrade prices of various contracts are quite close to what would be expected under the ranking assumption. Under the joint hypothesis that the Intrade price ranking is correct and the ranking assumption is correct, President Bush should not have won any state ranked below a state that he lost. He did not win any such state. The ranking assumption is also consistent with the fact that the two parties spent essentially nothing in most states in 2004.

1 Introduction

The U.S. Electoral College poses an interesting predictive problem. It can happen, as in the 2000 election, that one candidate gets the largest share of the national vote

*Cowles Foundation and International Center for Finance, Yale University, New Haven, CT 06520-8281. Voice: 203-432-3715; Fax: 203-432-6167; email: ray.fair@yale.edu; website: http://fairmodel.econ.yale.edu. I am indebted to Keith Chen, Don Green, Ed Kaplan, Steve Morris, Sharon Oster, Justin Wolfers, and Eric Zitzewitz for helpful discussions and comments and to Mike Knesevitch at Intrade for supplying me with some of the data.

and yet loses the election. Predicting the winner requires more than just predicting national vote shares. Individual state probability data are used in this paper to estimate the probability of winning in the Electoral College. A "ranking" assumption is made about dependencies across states that is different from assumptions used in previous work and that greatly simplifies the analysis. The state probability data are from a new data source, the Intrade political betting market.[1] The data are discussed in Section 2, the stochastic assumptions are discussed in Section 3, and the data and ranking assumption are analyzed in Section 4 for the 2004 election. Section 5 uses stochastic simulation to analyze the consequences of uncertain probability estimates, and Section 6 considers campaign spending under the ranking assumption. Section 7 concludes.

2 The Intrade Data

Prior to the 2004 election the website *www.intrade.com* allowed one to buy and sell contracts for each state and the District of Columbia. The contract for Iowa, for example, stated "G W Bush to win the electoral votes of Iowa." The contracts were in units of ten dollars, and a price of 55.0 meant that you could buy one contract for $5.50. If Bush won Iowa, you would get back $10.00. Otherwise, you would get back nothing. You could also sell the contract, winning $5.50 if Bush lost and losing $4.50 if Bush won. There was also a national contract that stated "George W Bush is re-elected as United States President." There were also contracts for

[1] The Intrade data are sometimes referred to as Tradesports data. Intrade is a subdivision of Tradesports, and the data are the same.

various combinations of state victories. For example, there was a Bush Greatplains contract that stated "Pres George W Bush to win IA, KS, MN, NE, ND, OK, SD, & TX." The national contract was by far the most traded contract on Intrade. The markets for many of the state contracts were fairly thin. An interesting discussion of this market and others like it is in Wolfers and Zitzewitz (2004a).

Table 1 presents the prices of the state contracts that existed on five different days. The first is September 7, 2004, the day after Labor Day. The rest are two weeks apart. The time of day is 10:00 am Eastern for the first, third, and fourth, 11:00 am Eastern for the second, and 6:00 am Eastern for the last. The last day is the day of the election, and 6:00 am Eastern is the time that the first polls open.

The states are ranked in Table 1 by the price on the last day. Many of the states have prices close to 100.0, and many have prices close to 0.0. This, of course, is the red state/blue state distinction that is popular in the press. On September 7 there were 13 states that had prices between 30.0 and 70.0, and on the last day there were 6 such states. The number of electoral votes President Bush needed to win the election was 269.[2] On the last day the "pivotal" state was Ohio, with a price of 51.1. With a little rearranging, it can be seen that on September 7 the pivotal state was Florida, with a price of 60.5.

[2]The electoral vote is tied if each candidate gets 269, but a tie goes to the House of Representatives, which is controlled by the Republicans.

Table 1
Intrade Data

State	Intrade Price 9/7	9/21	10/5	10/19	11/2	# Votes	$\sum$ Votes
Montana	95.0	94.0	95.0	96.3	99.0	3	3
Oklahoma	97.0	97.0	97.0	97.0	98.4	7	10
Utah	96.0	97.0	97.0	97.5	98.0	5	15
Idaho	95.5	96.0	95.0	95.5	98.0	4	19
Texas	98.0	98.0	98.0	98.0	97.9	34	53
Wyoming	97.0	97.0	97.0	97.5	97.6	3	56
Indiana	96.0	96.0	91.2	94.4	97.4	11	67
Alaska	96.0	96.0	98.0	95.5	97.4	3	70
Louisiana	92.5	91.9	92.0	92.6	97.0	9	79
Tennessee	78.7	85.0	89.0	92.0	96.5	11	90
Kentucky	92.5	92.0	92.0	93.1	95.8	8	98
Kansas	96.0	96.0	93.5	94.1	95.8	6	104
Mississippi	96.0	96.0	94.0	94.5	95.6	6	110
Georgia	96.5	97.0	92.2	95.7	95.2	15	125
Alabama	98.0	96.0	94.0	96.5	95.2	9	134
Nebraska	96.0	97.5	94.0	95.7	95.2	5	139
South Carolina	95.0	97.0	91.0	93.7	95.1	8	147
North Dakota	96.0	96.0	92.5	95.5	95.1	3	150
South Dakota	96.0	96.0	92.0	95.7	95.1	3	153
North Carolina	81.0	93.0	87.5	89.0	94.7	15	168
Arizona	78.0	83.0	83.0	90.0	94.0	10	178
Virginia	86.0	91.0	87.5	87.8	93.2	13	191
West Virginia	67.7	77.0	77.0	79.9	92.0	5	196
Arkansas	73.0	78.0	84.0	82.0	90.0	6	202
Missouri	67.0	85.0	84.0	81.0	87.1	11	213
Colorado	75.5	76.0	75.0	79.4	77.0	9	222
Nevada	60.0	69.9	74.5	67.5	76.8	5	227
New Mexico	43.0	40.0	37.7	37.2	56.5	5	232
Florida	60.5	70.0	63.5	66.0	53.9	27	259
Ohio	63.0	72.0	67.5	57.8	51.1	20	279
Iowa	43.0	55.0	57.0	55.2	51.0	7	
Wisconsin	57.0	62.0	64.0	54.5	41.0	10	
New Hampshire	42.0	55.0	51.0	43.0	31.0	4	
Pennsylvania	43.4	43.0	35.0	38.0	28.9	21	
Hawaii	10.0	10.0	8.0	5.5	26.1	4	
Minnesota	40.0	40.5	35.5	38.5	24.0	10	
Michigan	33.0	29.9	23.0	19.9	11.1	17	
New Jersey	15.9	24.0	18.0	16.5	10.0	15	
Oregon	36.3	35.0	26.9	21.9	10.0	7	
Maine	27.4	26.2	26.5	24.0	9.2	4	
Delaware	16.0	18.0	13.0	9.6	5.1	3	
California	9.6	11.4	8.0	6.0	3.3	55	
Connecticut	8.0	7.0	7.0	5.7	3.3	7	
Washington	28.0	25.0	19.0	8.0	3.0	11	
Vermont	7.0	8.0	8.0	3.3	2.5	3	
Illinois	8.8	12.0	8.8	6.8	2.0	21	
Maryland	14.0	16.0	17.9	9.0	2.0	10	
New York	7.0	9.9	8.4	4.9	1.7	31	
Massachusetts	3.0	4.0	2.0	2.8	1.7	12	
Rhode Island	4.0	4.0	4.0	3.5	1.7	4	
DC	1.0	1.0	2.0	1.5	0.8	3	

- Votes are electoral votes. 269 votes are needed to win for President Bush.
- President Bush won Iowa, all the states above it, and none below it.

3 Stochastic Assumptions

The postulated probability structure in this paper is as follows. Assume that on election day there are n possible "states" of nature (to be called "conditions" of nature to avoid confusion with U.S. states), each with probability $1/n$ of occurring. If in p_i percent of the n conditions Bush wins state i, then p_i is the probability that Bush wins state i. Rank the states by p_i, as was done in Table 1 using the Intrade data. The key assumption in this paper, called the "ranking" assumption, is that there is no condition of nature in which Bush wins state i and loses a state ranked higher than i. If, for example, Texas is ranked higher than Massachusetts, then in none of the n conditions of nature does Bush win Massachusetts and lose Texas. There may be conditions in which Bush wins Massachusetts (Kerry makes some serious error), but in these conditions Bush also wins Texas.

Under the ranking assumption it is trivial to compute the probability that Bush wins in the Electoral College. Just go down the ranking, adding electoral votes, until 269 is reached. If this is state j, then state j is "pivotal," and the probability that Bush wins the election is simply the probability that he wins state j.

It is common in previous work to assume some form of independence. Kaplan and Barnett (2003) assume that the state outcomes are independent, that "the events that the candidate is leading in various states are mutually independent" (p. 33). Snyder (1989) analyzes districts and assumes that the elections in the districts are all statistically independent. He points out that this rules out "uncertainty about national variables that may affect the electoral outcomes in all districts simultaneously, such as changes in aggregate output or foreign policy crises" (p. 646).

Brams and Davis (1974) assume that "the voting of uncommitted voters within each state is statistically independent" (p. 120). Strömberg (2002) assumes that the state level popularity parameters of a candidate are independent, although he also has a national popularity parameter.

What would it mean in the present context for the state probabilities to be independent? On election day the probability of Bush winning state i is simply the percent of his state i wins in the n possible conditions of nature. The probabilities will, of course, change if the n possible conditions of nature change. Consider as a thought experiment different *sets* of n possible conditions of nature on election day. Say that Bush has done poorly in the debates in set 1 and well in set 2. One would expect all the state probabilities to be higher for Bush in set 2. In set 2 there would fewer conditions of nature in which Bush loses any given state. The state probabilities in this case would be positively correlated. In order for the probabilities to be uncorrelated, the sets must differ in state-specific ways. For example, the Republican party might be better organized in California in set 1 than in set 2, but everything else the same. The two sets would then differ only regarding the probability for California. These state-specific differences across different sets of the n possible conditions of nature seem less likely to occur than differences that affect all the state probabilities.

The ranking assumption does not, of course, directly concern different sets of the n possible conditions of nature. It simply puts restrictions on the n possible conditions of nature that exist on election day. If state i is ranked ahead of state j, then in no condition of nature does Bush win j and lose i. The concept of different *sets* of the n possible conditions of nature is not needed.

4 Analysis of the Intrade Data and the Ranking Assumption

Price Predictions

How should the prices in Table 1 be interpreted in light of the setup in Section 3? It is assumed in this paper that the prices are *estimates* of what the probabilities will be on election day—of what the n possible conditions of nature will be on election day.[3]

Given the individual state prices in Table 1, the Intrade prices of various combination contracts are quite close to what one would expect under the ranking assumption. This can be seen in Table 2, which presents prices for various combination contracts along with what the ranking assumption would predict the prices should be and what the independence assumption would predict. For the Bush Greatplains contract, for example, the price predicted by the ranking assumption is the price of the lowest ranked state in the contract, which for September 7 is Minnesota with a price of 40.0. The price predicted by the independence assumption is simply the product of the state prices (after dividing each price by 100 and multiplying the final product by 100).

It is clear from Table 2 that the predictions are much closer under the ranking assumption than under the independence assumption. The worst case for the independence assumption is Bush South, where for September 7 the ranking-

[3]Manski (2004) has shown that under certain assumptions about the beliefs of traders the market price of a contract is not necessarily the mean belief of the traders. However, under what appear to be plausible assumptions, this bias is either zero or small—see Wolfers and Zitzewitz (2004b). This paper is based on the assumption that the bias is zero.

Table 2
Intrade Prices for Various Contracts

	September 7, 2004			November 2, 2004		
Contract	Intrade Price	Predicted by Ranking Assumption	Predicted by Independ. Assumption	Intrade Price	Predicted by Ranking Assumption	Predicted by Independ. Assumption
Bush Greatplains	35.0	40.0	13.9	23.0	24.0	9.7
Bush OH+FL	56.9	60.5	38.1	37.0	51.1	27.5
Bush South	55.0	60.5	18.9	53.0	53.9	32.3
Bush Southwest	36.0	43.0	18.7	53.8	56.5	32.7
Kerry New England	53.7	58.0	33.5	70.0	69.0	57.1
Kerry Rustbelt	32.0	37.0	14.0	42.5	48.9	30.9
Kerry Westcoast	63.5	63.7	41.5	87.5	90.0	84.4

Notes:
- Greatplains: IA, KS, MN, NE, ND, OK, SD, & TX.
- South: SC, MS, FL, AL, GA, LA, TX, VA, AR, NC, & TN.
- Southwest: NV, NM, UT, & CO.
- New England: CT, RI, ME, VT, MA, & NH.
- Rustbelt: PA, OH, & MI.
- Westcoast: CA, OR, & WA.

assumption price is 60.5, the price for Florida, and the independence-assumption price is 18.9. These compare to the actual price of the contract of 55.0. The only weak case for the ranking assumption is Bush OH+FL for November 2, where the contract price is 37.0 and the price predicted by the ranking assumption is 51.1. Although the results in Table 2 have to be taken with some caution because the markets are thinly traded, they are strikingly supportive of the ranking assumption.[4]

Table 3 shows the price of the national contract on each of the five days and the price of the pivotal state. Remember that under the ranking assumption the two prices should be the same. The table shows that the prices are quite close. On the last day the prices differ by 4.4, but the bid/ask spread for Ohio was quite large, and so the Ohio price may not be reliable.

[4]Ed Kaplan has pointed out to me that given a ranking like in Table 1, under the ranking assumption there are only 52 possible outcomes: Bush takes all 51, Bush takes all but the last one, Bush takes all but the last two, etc. This compares to 2^{51} possible outcomes, about 2.25 million billion. A remarkable economy of outcomes has been achieved by the ranking assumption!

Table 3
Intrade Data on the National Contract

	9/7	9/21	10/5	10/19	11/2
National Contract	60.2	70.0	60.0	58.5	55.5
Pivotal State	60.5	70.0	63.5	57.8	[a]51.1
	FL	FL	FL	OH	OH

[a]Bid/ask spread was 50.0/55.5.

The Actual Outcome

Tables 2 and 3 show that the ranking assumption is a good predictor of the Intrade prices of the combination contracts, including the national contract. Market participants appear to be using the ranking assumption in pricing these contracts. These results say nothing about the accuracy of the Intrade prices in predicting the actual outcome. After the outcome, however, one can consider the joint hypothesis that the Intrade price ranking on the last day is correct and the ranking assumption is correct. Under this hypothesis President Bush should not have won any state ranked below a state that he lost. Table 1 shows that he did not win any such state. Bush won Iowa, all the states above Iowa, and none below Iowa. The results are 100 percent in favor of the joint hypothesis!

Note from Table 1 that Bush won all the states with a price above 50 on the last day and lost all the states with a price below 50. Although Intrade is quite happy about this result, it is not necessary for the joint hypothesis to be true. If, say, all the prices on the last day were 10 percent lower, so that the price of Iowa were 45.9 rather than 51.0, the results would still be consistent with the joint hypothesis even though Bush would have won Iowa with a price below 50.

5 Uncertain Probability Estimates

It is clear from Table 1 that the ranking of the states varies somewhat across time. This result is not, however, inconsistent with the ranking assumption because the assumption pertains only to the ranking on the last day. For example, the Intrade estimates in Table 1 on September 7 are uncertain because unexpected events can happen between September 7 and election day. Let P_i^a be the probability that Bush wins state i on election day, and let $\hat{P}_i^a$ be the estimated probability on September 7. Let u_i be the difference between the two:

$$u_i = \hat{P}_i^a - P_i^a. \tag{1}$$

It is important to note that u_i is an estimation error. $\hat{P}_i^a$ is uncertain, but P_i^a is not. As discussed in Section 3, P_i^a is simply the percent of Bush wins in state i in the n conditions of nature that exist on election day.

Surprises that happen before election day will change the estimated probabilities as people update their views about the conditions of nature that will exist on election day. A surprise negative performance by Bush in the debates would likely lower nearly all the estimated probabilities. The fact that the ranking in Table 1 changes somewhat across time means in the present context that the u_i vary across states. There is obviously a positive correlation, since most probabilities change in the same direction, but the correlation is not perfect. Some of a state estimation error is thus state specific.

Stochastic Simulation

To the extent that some of the variation in the u_i is state specific, it is of interest to examine the effects of this variation. Table 4 presents results of some stochastic simulations that get at this question. To focus on state-specific variation, the error terms are taken to be uncorrelated across states for the simulation work. The states used are the 13 states with prices between 30.0 and 70.0 on September 7. For each state i, u_i is assumed to be normally distributed with mean 0 and variance σ^2. σ^2 is assumed to be the same across states.

The stochastic-simulation experiments were performed as follows. For each trial 13 errors were drawn from the $N(0, \sigma^2)$ distribution, one per state, where σ varied from zero for the first experiment to 0.05 for the sixth experiment. Consider a given experiment, i.e., a given value of σ. Let $u_i^{(k)}$ denote the error drawn for state i on the kth trial. The probability for state i on the kth trial was computed as:

$$P_i^{(k)} = \hat{P}_i^a + u_i^{(k)}. \tag{2}$$

In this context $\hat{P}_i^a$ is the "base" probability. For each trial k the values of $P_i^{(k)}$ were ranked, the pivotal state was determined,[5] and its probability, denoted $P_p^{(k)}$, was recorded. This was done 10,000 times, resulting in 10,000 values of $P_p^{(k)}$. The number of times a particular state was the pivotal state was also recorded, as was the number of times a state was above the pivotal state. Summary results are presented in Table 4. Presented in the table are the minimum value of $P_p^{(k)}$, the value below which 5 percent of the trial values lie, and the median. Also presented

[5] For this work 270, not 269, was taken to be the number of electoral votes needed to win.

Table 4
Stochastic Simulation Results
Data for September 7, 2004

	Value of σ					
	0.00	0.01	0.02	0.03	0.04	0.05
$P_p^{(k)}$						
median	.600	.597	.592	.588	.582	.576
minimum	.600	.559	.522	.481	.439	.409
.05	.600	.583	.567	.551	.533	.515
# times pivotal state						
WV	0	0	0	48	111	180
MO	0	0	9	91	254	400
OH	0	50	705	1416	1962	2199
FL	0	3560	4185	4278	4185	4057
NV	10000	6218	4113	2913	2236	1814
WI	0	172	988	1254	1235	1197
PA	0	0	0	0	8	84
IA	0	0	0	0	3	26
NM	0	0	0	0	2	18
NH	0	0	0	0	4	10
MN	0	0	0	0	0	13
OR	0	0	0	0	0	1
MI	0	0	0	0	0	1
# times pivotal state or above						
WV	10000	10000	9999	9978	9906	9783
MO	10000	10000	10000	9982	9927	9807
OH	10000	10000	10000	10000	9997	9955
FL	10000	10000	10000	10000	9996	9943
NV	10000	9819	8733	8122	7869	7753
WI	0	248	2100	3616	4556	5208
PA	0	0	0	0	10	101
IA	0	0	0	0	12	104
NM	0	0	0	0	11	97
NH	0	0	0	0	8	50
MN	0	0	0	0	1	21
OR	0	0	0	0	0	2
MI	0	0	0	0	0	1

- The prices (base probabilities) from Table 1 for September 7 are: WV 67.7, MO 67.0, OH 63.0, FL 60.5, NV 60.0, WI 57.0, PA 43.4, IA 43.0, NM 43.0, NH 42.0, MN 40.0, OR 36.3, MI 33.0.
- 10000 trials per value of σ.
- $P_p^{(k)}$ = probability of winning the election for the kth trial, which is the probability of winning the pivotal state.
- .05 for $P_p^{(k)}$ means the value below which 5 percent of the trial values lie.

are the number of times each state was pivotal and the number of times each state was pivotal or above the pivotal.

Before discussing the results it should be noted that it can be the case in the stochastic simulations that $P_i^{(k)}$ for a particular state i is greater than the base probability for states above the highest ranked state used (West Virginia) or less than the base probability for states below the lowest ranked state used (Michigan). This does not matter for the results, however, because the solutions that matter are around the pivotal state. The stochastic simulation could have been set up using all the states, but, as just noted, this is not necessary. If all states were used, the assumption that the variance of the error term is the same across states would have to be changed. The variance is obviously smaller when the base probability is near one or zero than when it is near one half.

The results in Table 4 are easy to explain. When the variance is zero, Nevada is always pivotal and the probability of winning the election is always .600.[6] As the variance increases, more and more states are sometimes pivotal or above the pivotal. The median of $P_p^{(k)}$ falls from .600 when σ is zero to .576 when σ is 0.05. The median falls because, except for Wisconsin, the states below Nevada have base probabilities that are considerably below .600. There is not symmetry around .600, and so negative draws for states above Nevada are on average not completely offset by positive draws for states below Nevada. When the calculations were repeated using .570 for the base probabilities for the states below Wisconsin (instead of the values in Table 1 for September 7), the median of $P_p^{(k)}$ rose as the variance

[6] In Table 3 Florida is listed as the pivotal state for September 7, whereas in Table 4 Nevada is listed as pivotal. This difference is due to the use of 270 electoral votes to win rather than 269.

increased. For $\sigma = 0.01$ the median was .597. The values of the median for the increasing values of σ were, respectively, .598, .600, .603, and .605.

When σ is zero, i.e., no state-specific variation, all that matters in terms of predicting the probability of winning the election is the probability for the pivotal state. It does not matter, for example, how much larger the probabilities for the states above the pivotal state are or how much smaller the probabilities for the states below the pivotal state are. As just seen, this changes when σ is non zero—the sizes of the probabilities around the pivotal state now matter.

The stochastic simulations were repeated using the September 21 data, and the results are presented in Table 5. These results are similar to those in Table 4, although with higher probabilities, except that some states are now never pivotal nor above the pivotal. The fact that the base probabilities for Iowa and New Hampshire have risen substantially leads to these states doing all the extra work. Even with its 21 electoral votes, Pennsylvania is never used.

6 Campaign Spending

The ranking assumption has important implications for campaign spending across states. On election day there are postulated to be n possible conditions of nature, one of which is drawn. Each condition is based on everything that has happened up to the day of the election (i.e., up to the time of the draw). "Everything" includes all the campaigning that has been done in each state. After all the campaigning is over, the ranking assumption says that there is no possible condition of nature in which Bush wins a state ranked below a state he loses. This is not to say, of course,

Table 5
Stochastic Simulation Results
Data for September 21, 2004

	Value of σ					
	0.00	0.01	0.02	0.03	0.04	0.05
$P_p^{(k)}$						
median	.699	.694	.688	.680	.673	.667
minimum	.699	.658	.617	.576	.534	.492
.05	.699	.680	.660	.642	.623	.606
# times pivotal state						
MO	0	0	0	0	2	7
WV	0	0	4	78	187	296
OH	0	219	1100	1733	2103	2333
FL	0	4553	4264	4016	3870	3819
NV	10000	5228	4610	3898	3265	2648
WI	0	0	22	268	532	743
IA	0	0	0	3	27	80
NH	0	0	0	4	14	74
# times pivotal state or above						
WV	10000	10000	10000	9998	9980	9908
MO	10000	10000	10000	10000	10000	10000
OH	10000	10000	10000	10000	9999	9971
FL	10000	10000	10000	10000	10000	10000
NV	10000	10000	9977	9683	9277	8838
WI	0	0	48	624	1543	2456
IA	0	0	0	5	74	285
NH	0	0	0	7	54	300

- See notes to Table 4.
- The prices (base probabilities) from Table 1 for September 21 are: MO 85.0, WV 77.0, OH 72.0, FL 70.0, NV 69.9, WI 62.0, IA 55.0, NH 55.0, PA 43.0, MN 40.5, NM 40.0, OR 35.0, MI 29.9.
- PA, NM, MN, OR, and MI were never used.

that campaigning has no effect on the possible conditions of nature. It is just that once campaigning is over, the ranking assumption holds.

Consider now the strategy of the Republican party on September 7. Assume for now that the Republican party does not take into account any Democratic-party response to its actions. As in Section 5, let $\hat{P}_i^a$ denote the Intrade price on September 7. In Section 5 this price was taken to be market's estimate of what the actual probability will be on election day (P_i^a). This estimate obviously takes into account market participants' views about how much campaigning there will be in each state. Let z_i^e denote the market's expectation of the amount the Republican party will spend in state i between September 7 and election day. The following equation is then postulated:

$$P_i^a = \hat{P}_i^a + f_i(z_i - z_i^e) + u_i \tag{3}$$

where z_i is the actual amount the Republican party spends in state i between September 7 and election day. Equation (3) says that spending in a state affects the probability of winning the state. The Republican party faces a budget constraint that the sum of the z_i's across all the states cannot exceed some amount.

Consider first the case in which u_i in equation (3) is zero for all i. If the Republican party wants to maximize the probability of winning the election, what should it do? Under the ranking assumption, it simply maximizes the probability of winning the pivotal state. In Table 1 for September 7 the pivotal state is Nevada (assuming 270 electoral votes needed to win), which has a price of 60.0. The state above it is Florida, with a price of 60.5. The next state is Ohio, with a price of 63.0, and the next state is Missouri with a price of 67.0. To take an example, say the Republican party's budget constraint is such that the party can spend in Nevada, Florida, and Ohio to raise P_i^a to 65.0 each. The probability of winning has thus increased from .60 to .65, and there has been spending in just three states. (In this example there would be in the end no conditions of nature on election day in which Bush won one or two of these states and lost the other.)

Consider next the case in which u_i is not zero in equation (3). Remember that these are state-specific errors of estimation. On September 7 the Republican party knows that it can change the actual probabilities that will exist on election day, but when there are estimation errors it does not know the actual values that will exit. What should be the objective of the party in this case? Go back to the stochastic-simulation setup in Section 5, and assume that the 13 states in Table 4 are in play. Let z denote the vector of the 13 z_i values, and let u denote the vector of the 13 u_i values. Given z and u, it is straightforward to compute the probability that the Republican party wins the election. The values of P_i^a can be computed from equation (3) (assuming also knowledge of the z_i^e) and then the values ranked to determine the pivotal-state value. For the given value of z this can be done, say, for 10,000 draws of u. This gives 10,000 values of the probability of winning the election, from which summary measures like those in Table 4 can be computed.

One can think of the Republican party considering many values of z and for each value computing 10,000 probabilities and summary measures like those in Table 4. Its objective might be to choose z to maximize the median of the probability values, the minimum of the values, or the value below which 5 percent of the trial values lie. This last option means that there would be a 95 percent chance that the actual probability of winning on election day is above the maximized value. Whatever is maximized, Table 4 shows that the optimal strategy for the party would be to allocate some of its spending to states below Nevada, the pivotal state when the errors are zero. Some states that are below Nevada now have, depending on the draw for u, some chance of being pivotal, and so it would be optimal to spend something on these states.

The addition of uncertainty has thus increased the number of states in which spending is done. Table 4 shows that as the variance of the error term increases, the number of states that are sometimes pivotal increases. Thus, the larger the variance, the larger the number of states in which spending is done.

Consider finally the Democratic-party response to a Republican-party move. In any given presidential election the two parties generally have similar resources and similar information. It also seems likely that the effects of spending on votes are similar between the two parties. If there is complete symmetry between the two parties and, say, the Republicans move first, then the Democrats can merely offset whatever the Republicans do. In practice this seems to be roughly the case. Both parties focus their spending on the swing states and come close to matching each other by state in terms of number of visits by the candidates and advertising spending. If one party begins to do more in a key state, the other party tends to respond. Also, there is essentially no spending in many states, which, as discussed next, is consistent with the ranking assumption but not the independence assumption.

No attempt is made in this paper to set up a formal game between the two parties under the ranking assumption. This is a possibly interesting area for future work. With a probability structure like that in Table 1, where many states are close to zero or one, it seems clear from the results in Table 4 that if a game is set up using the ranking assumption, there are likely to be many states in which there is no spending by either party. This is contrary to results in the literature that are based on the independence assumption. In the model of Snyder (1989), for example, spending is high in states that are close and that have a high probability of being pivotal, but there is some spending in all states. The same is true for the model in Strömberg (2002). In the model of Brams and Davis (1974) there is spending in all states, where spending is in proportion to the 3/2's power of the number of electoral votes in each state.

7 Conclusion

Although the ranking assumption is obviously only an approximation, it appears to be a very good one. It is consistent with the way combination contracts are priced on Intrade, and the actual outcome of the 2004 election is completely in line with the joint hypothesis that the Intrade ranking is correct and the ranking assumption is correct. The ranking assumption is also consistent with the fact that the two parties spent essentially nothing in most states in 2004.

References

[1] Brams, Steven J., and Morton D. Davis, 1974, "The 3/2's Rule in Presidential Campaigning," *The American Political Science Review*, 68 (March), 113-134.

[2] Kaplan, Edward H., and Arnold Barnett, 2003, "A New Approach to Estimating the Probability of Winning the Presidency," *Operations Research*, 51 (January/February), 32-4.

[3] Manski, Charles F., 2004, "Interpreting the Predictions of Prediction Markets," February.

[4] Snyder, James M., 1989, "Election Goals and the Allocation of Campaign Resources," *Econometrica*, 57 (May), 637-660.

[5] Strömberg, David, 2002, “Optimal Campaigning in Presidential Elections: The Probability of Being Florida,” Seminar Paper No. 706, Institute for International Economic Studies, Stockholm University.

[6] Wolfers, Justin, and Eric Zitzewitz, 2004a, ”Prediction Markets,” *Journal of Economic Perspectives*, 18 (Spring), 107-126.

[7] Wolfers, Justin, and Eric Zitzewitz, 2004b, “Using Markets to Inform Policy: The Case of the Iraq War,” June.