

Was Ralph Nader a Spoiler?
A Study of Green and Reform Party Voters in
the 2000 Presidential Election¹

***** PRELIMINARY/INCOMPLETE DRAFT *****

Michael C. Herron²

Jeffrey B. Lewis³

April 21, 2004

¹An earlier version of this paper was presented at the 2004 annual meeting of the Midwest Political Science Association, Chicago, IL

²Department of Government, Dartmouth College, and Department of Political Science, Northwestern University. HB 6108, Hanover NH 03755 (Michael.C.Herron@Dartmouth.edu).

³Department of Political Science, University of California, Los Angeles. 4289 Bunche Hall, Los Angeles, CA 90095-1472 (jblewis@ucla.edu).

Abstract

The 2000 presidential race included two major party candidates—Republican George W. Bush and Democrat Al Gore—and two prominent third party candidates—Ralph Nader of the politically left Green Party and Pat Buchanan of the politically right Reform Party. While it is often presumed that Nader spoiled the 2000 election for Gore by siphoning away votes that would have been cast for him in the absence of a Nader candidacy, we show that this conventional wisdom is quite misleading. While Nader voters in 2000 were left of center and Buchanan voters right of center, both of these two types of voters were surprisingly close to being partisan centrists. Many Nader voters, that is, supported Republican candidates in non-presidential races and almost 40% would have voted for George W. Bush had they turned out in a Nader-less election. Our results are based on studying over 46 million vote choices from approximately three million ballots cast across Florida in the 2000 general election. They do not rely on voter self-reports, and the results show how ballot-level studies are capable of illuminating aspects of third party presidential voters that are otherwise beyond scrutiny.

1 Introduction

Running for president in 2000 were two major party candidates—Republican George W. Bush and Democrat Al Gore—along with a pair of prominent third party candidates—Ralph Nader and Pat Buchanan of the left-leaning Green Party and right-leaning Reform Party, respectively. Following a month-long period of recounts, legal wrangling, and court rulings in both Florida and Washington, D.C., Bush emerged victorious in December, 2000 over Gore by five Electoral Votes. This is a miniscule margin insofar as there were in the 2000 general election six states worth a total of 60 Electoral Votes in the aggregate that had Bush–Gore margins of fewer than two percentage points. The states in question were Florida, Iowa, New Hampshire, New Mexico, Oregon, and Wisconsin. Clearly, then, a handful of third party presidential votes could have altered the outcome of the 2000 presidential election had these votes been cast in key places for major party candidates instead of wasted, so to speak, on relatively minor candidates who had no chance of becoming president. Indeed, Nader and Buchanan together received a grand total of zero Electoral Votes, the same number that all third party candidates combined have earned in the presidential elections since 1992.

If either of the prominent third party presidential candidates in 2000 played the role of spoiler, this appellation belongs to Ralph Nader insofar as Gore lost the 2000 presidential race and Nader’s Green Party was perceived publicly as being more Democratic than Republican. The belief that voters who supported Nader would have almost certainly cast their votes for Gore had Nader not been a candidate for president in 2000 is evident in comments made by Democratic National Committee Chairman Terry McAuliffe while discussing a potential Nader candidacy in 2004: “We can’t afford to have Ralph Nader in the race... This is about the future of our country. If you care about the environment, if you care about job growth, you’ve got to support the Democratic nominee.”¹

The possibility of a 2004 Nader run for The White House has alarmed Florida Democrats in particular. In November, 2000 there were 97,488 votes cast for Nader in Florida, a pivotal state where the official Bush–Gore vote margin was a scant 537 votes; the minuteness of this number is highlighted by the fact that statewide turnout in Florida was 6,138,765. In commenting on a

¹Quoted in a February 22, 2004 report on CNN.com.

proposed 2004 Nader candidacy, Scott Maddox, the chairman of the Democratic Party in Florida, said, “I think that Ralph Nader is proving that the only master that he serves is his enormous ego. . . I have nothing nice to say about him: 2000 should have proved to him that he’s going to be nothing but a spoiler.”²

Was, in fact, Nader a spoiler in 2000? And, similarly, did Pat Buchanan siphon a significant number of votes away from George W. Bush, thus converting a solid Bush victory in Florida and other states into a tight Bush versus Gore contest at the Electoral College level?

We show that the answer to both of these two questions is no in a general sense, although the answer to the former is affirmative in light of Florida’s striking closeness. While Nader voters in 2000 were on average left of center in a partisan sense and Buchanan voters correspondingly right of center, most Nader and Buchanan supporters were surprisingly close to being nonpartisan, meaning that they were neither heavily pro-Democratic nor heavily pro-Republican. Perhaps somewhat surprisingly, at least in light of the aforementioned quotations from Democratic Party officials, many Nader voters had Republican leanings and Buchanan voters, Democratic leanings—and this is not due solely to Palm Beach County’s infamous butterfly ballot which led approximately 2,000 Gore voters to cast accidental votes for Buchanan (Wand, Shotts, Sekhon, Mebane, Jr., Herron & Brady 2001).

With respect to Nader in particular, we estimate that approximately 61% of Nader voters would have supported Gore had they turned out and voted for one of the two major presidential candidates in 2000 (the comparable figure for Buchanan supporters and Bush votes is 58%). Of course, given the tightness of the 2000 presidential election in Florida, it follows from the former percentage that Nader’s candidacy was indeed pivotal to Gore’s loss in the state. This is somewhat of a trivial statement, though, as practically any voting phenomenon that is not literally exactly neutral in terms of its effects on major party candidates will be pivotal in a state with a tiny vote margin. For instance, the 562 Florida voters who supported the presumably left-leaning Socialist Workers Party candidate for president cost Gore the presidency, but this fact has not led to popular (or Democratic)

²Quoted in “THE 2004 CAMPAIGN: THE INDEPENDENT; Nader, Gadfly to the Democrats, Will Again Run for President,” *The New York Times*, February 23, 2004.

Table 1: Reallocating Nader and Buchanan Votes in Five Close States

State	Gore	Bush	Buchanan	Nader	Gore-Bush	New Gore-Bush
Iowa	638517	634373	5731	29374	4144	9689
New Hampshire	266348	273559	2615	22198	-7211	-2748
New Mexico	286783	286417	1392	21251	366	4820
Oregon	720342	713577	7063	77357	6765	22653
Wisconsin	1242987	1237279	11471	94070	5708	24568

Note: New Bush-Gore margins are rounded to the nearest vote.

excoration of the Socialist Workers Party. Nonetheless, our broader point is that 61% is much closer to 50% than it is to 100%. It is simply incorrect to assert that a vast majority of Nader voters would have supported Gore rather than Bush had they faced a two-candidate, Bush versus Gore, contest for president.

Additional evidence of the fact that neither Nader nor Buchanan was a spoiler in a general sense can be gleaned through examining Bush-Gore margins from close states in the 2000 general election. Florida notwithstanding, of the five such states noted previously Gore won all of them but New Hampshire, which he lost by 7,211 votes. See Table 1 for details. If we reallocate Nader votes assuming that 61% are cast for Gore and, correspondingly, Buchanan votes assuming that 58% went to Bush (the origins of these two numbers will be apparent shortly), then as shown in the table none of the five close states switch insofar as having a different election winner under our reallocation scenario. New Hampshire's margin becomes much tighter when Nader and Buchanan votes are distributed among Bush and Gore: our numbers suggest that Bush would have won the state by on 2,746 votes had neither Buchanan nor Nader run for president in 2000.

Our evidence on how Nader and Buchanan supporters would have voted had they faced a dichotomous choice between Bush and Gore is based on patterns in 3,067,701 ballots cast in Florida during the 2000 presidential contest. As explained in detail later, our analysis starts by generating for each of our approximately three million ballots a cardinal measure of the ballot's partisanship, i.e., a measure of the extent to which the ballot is overall a Democratic ballot versus being a Republican ballot. These partisanship measures—and we emphasize that they measure Democratic versus

Republican partisanship rather than ideology in the sense of liberalism versus conservatism—are based on our ballots’ non-presidential votes, e.g., votes in Congressional contests, on Florida Constitutional Amendments, on local referenda, and so forth. That is, one can conceptualize every general election ballot as having a partisan identification that is revealed through votes in non-presidential contests. The assignment of a partisanship measure to each of our approximately three million ballots enables us to determine if Nader votes, and similarly Buchanan votes, appear on Democratic ballots, on Republican ballots, or on something else altogether. In fact, as we make clear, Nader votes appear on a significant number of ballots that, Nader notwithstanding, look quite Republican insofar as containing numerous Republican votes in non-presidential races. In total our ballot-level partisanship measures are based on analyzing 46,515,369 different non-presidential vote choices among approximately three million different voters.

Because our research design draws on ballots or what are called ballot images, and because these images are not aggregated to, say, the precinct or county level, our results on the types of voters who supported Nader and Buchanan in 2000 do not have to contend with ecological inference, a statistical technique that is used for studying voting returns in aggregate units like counties (e.g., Achen & Shively 1995, King 1997). Ecological inference, as the ongoing debate on this subject illustrates, is controversial, often depends on unverifiable statistical assumptions, and is prone to erroneous conclusions (Freedman, Klein, Ostland & Roberts 1998, Tam 1998, King 1999, Cho & Gaines 2004, Herron & Shotts 2004). Ecological inference should be avoided when it is not absolutely necessary, and our analysis of Nader and Buchanan voting patterns shows that the availability of presidential election ballot images obviates a dependence on this rather troublesome statistical technique.

Moreover, since our results are based on actual election ballots from the November, 2000 general election, we need not concern ourselves with the possibility of misreported votes and turnout decisions that can plague pre- and post-election surveys of vote choices like the National Election Study (NES). Of course, ballot secrecy means that we cannot associate a given ballot from our collection with a particular Florida resident; thus, we do not have access to the wealth of voter-level demographics and preference measures that the NES provides on a collection of voters via its detailed

surveys. We cannot, in light of this, assess the types of issues and voter feelings that motivate the patterns in the ballots we study. Nonetheless, we know for certain that our results do not depend on voters' remembering accurately how they cast their votes in November, 2000 or remembering (and then being honest about) whether they turned out to vote at all.

The ballots we use to study presidential election voting and in particular the partisanship of voters who supported Nader and Buchanan in the 2000 general election are drawn from a collection of ten counties in Florida. The availability of these ballots reflects the fact that scholarly and popular interest in this state, post-November, 2000, has led to the creation of an electronic archive that contains ballot images. The existence of this archive provides us with access to an unprecedented data source that can be used to analyze third party voters. Nonetheless, despite the fact that we use data from Florida, our analysis is not about the post-election controversy in the state per se (although our results have something to say about it and in particular about the butterfly ballot in Palm Beach County).

Indeed, whether particular idiosyncrasies in Florida's election administration in 2000 affected the result of the election there has been analyzed from a variety of perspectives (Sinclair, Mark, Moore, Lavis & Soldat 2000, Merzer 2001, Posner 2001, Wand et al. 2001, Smith 2002). Our objective, in contrast, is offering general results about third party presidential voters and their partisanship using Florida voting data as a lever, just as Mebane's (2003) results on presidential overvotes have implications that transcend Florida. While we would certainly prefer to have based our results on general election ballot images from across the United States, at present this is simply not feasible as archives that contain such images do not exist. Given the scale of our analysis and the fact that ballot images are just beginning to be used by voting researchers, it is fortunate, we believe, that any image archives exist.

The remainder of this paper is organized as follows. In Section 2 we make precise our research agenda, review literature, and comment on the use of ballot-level data in presidential election research. Then in Section 3 we describe the data that lies at the heart of our analysis and discuss our statistical methodology. Section 4 presents results, and Section 5 concludes.

2 Research Questions and Existing Literature

As described in the introduction, the objective of our research is understanding the partisanship of voters who supported Ralph Nader and Pat Buchanan—both of whom were prominent third party presidential candidates—in the 2000 general election. We model partisanship as a line where a voter placed on the left side of this line is relatively Democratic and a voter on the right hand side, relatively Republican. The further left (right) a given voter is on the partisanship line the more consistently he or she votes Democratic (Republican) across the contests in a given election. Centrist voters in this model of partisanship are those who sometimes vote Democratic and sometimes Republican.

It may be the case that the partisanship line for a given election is identical to the election's ideology line insofar as Democrats typically are politically liberal and Republicans, politically conservative. However, ideology and partisanship are different theoretical constructs: a given political issue on an election's agenda can split voters in a partisan way—meaning that on average Democrats vote one way and Republicans the other. An abortion rights referendum would presumably be highly partisan. However, other political issues are more ideological, such as the extent to which the federal government can monitor American citizens for anti-terrorism reasons. If a given election featured a referendum on increasing the monitoring capabilities of local police forces, Democrats with strong feelings on civil liberties and Republicans who seek to minimize the scope of the government would presumably vote against it. Because we eventually want to consider how to reallocate third party presidential votes to mainstream Democratic and Republican candidates, our interest is in partisanship rather than ideology.

The most comprehensive study of third party candidates and voters in American presidential elections is Rosenstone, Behr & Lazarus (1996), who discuss the multiple minor candidates who have run for president in the 19th and 20th Centuries and also offer a theory which purports to explain why voters choose to support third party presidential candidates. Broadly speaking, Rosenstone, Behr & Lazarus (ch. 5) argue based on a wealth of survey data from multiple general elections that third party presidential voters are motivated by three factors: major party deterioration (e.g., major

parties that are not responsive to voter concerns), attractiveness and prestige of available third party candidates, and allegiance to a third party.

Notably, the results in Rosenstone, Behr & Lazarus imply that third party presidential voters are not heavily partisan insofar as supporting major party candidates in non-presidential races. Thus, we infer, Rosenstone, Behr & Lazarus would predict that third party presidential voters are not imbued with particularly strong allegiances to Democratic or Republican candidates in non-presidential contests. From this it follows that Nader and Buchanan should not obviously be spoilers for Al Gore and George W. Bush, respectively. Rather, results in Rosenstone, Behr & Lazarus imply that Nader and Buchanan voters should be expected to be partisan centrists more than extremists.

This conclusion is consistent with analyses of H. Ross Perot's third party presidential bid in 1992. Alvarez & Nagler (1995) and Lacy & Burden (1999) argue that the relatively right-wing Perot stole more votes from Republican candidate George Bush than from his competitor Democrat Bill Clinton, but not appreciably more.

Practically everything known about third party voters and their partisanship is grounded in opinion surveys like the NES. Both Alvarez & Nagler and Lacy & Burden are NES-based, and Rosenstone, Behr & Lazarus's results are drawn from a collection of different surveys. Such a reliance on opinion surveys gives rise to two limitations within the literature on third party presidential voters.

First, opinion surveys often do not contain enough respondents so as to canvass a non-trivial number of voters who support third party candidates. For example, in 2000 the NES queried the vote choices of 1,178 individuals who voted on election day. Within this group, and ignoring the 13 surveyed individuals who either forgot their presidential vote choice or refused to provide one, there were a total of 33 people who voted for Nader and Buchanan and a total of only nine who voted for other third party candidates. A set of 33 individuals provides no meaningful leverage on the types of general election voters who supported Nader and Bush, and, therefore, the statistical model used in, say, Lacy & Burden cannot be used in a serious way to study Nader and Buchanan voters in 2000. It goes practically without saying that the 2000 NES provides essentially no information on voters who supported candidates less prominent than Nader and Buchanan.

Second, the literature's dependence on surveys means that its results are vulnerable to the types of voter misreports discussed in Wright (1990), Wright (1992), and Wright (1993). Namely, Wright shows that responses to NES vote choice questions have a winner bias insofar as more voters report voting for winning candidates than should exist. Whether such a bias is due to voter memory lapses that occur when being interviewed, confusion, or intentional dishonesty is not ultimately important here. Rather, the key point in Wright's work on survey responses, a point that extant literature on third party voters has essentially ignored, is that survey reports of vote choices are systematically biased and that these biases can have severe consequences.³

We surmount both of these limitations by analyzing vote choices actual vote choices as opposed to reports of these choices. Within our collection of approximately three million ballot images generated from ballots in Florida during the 2000 election, there are 48,238 votes for Nader and 8,384 for Buchanan. Hence, small number problems do not affect us. Furthermore, because we study electronic images of ballots which contain true votes, so we need not concern ourselves with voter misreports—regardless of why one believes these misreports exist.

Ballot-level research in political science is not particularly common, and this is presumably due to data limitations. Ballot images for almost all potential elections of interest simply do not exist. Nonetheless, four recent examples of the use of ballot-level data in political research are Lewis (2001), Wand et al. (2001), Herron & Sekhon (2003), and Mebane (2003). Lewis describes a ballot-level scaling technique (which we employ—details follow shortly) which is then applied to ballots from Los Angeles County; Wand et al. examine ballots cast in Palm Beach County in November, 2000 and study the relationship between presidential vote choice and vote choice in a U.S. Senate race; Herron & Sekhon consider ballot images from Florida's Broward and Miami Dade counties in a study of presidential overvotes and undervotes; and Mebane analyzes presidential overvotes cast in Florida using the Florida ballot images featured here in conjunction with a dataset that contains limited images of all 2000 general election ballots from Florida with invalid or uncounted presidential

³Clearly, Wright's critiques of survey results have consequences that transcend the literature on third party voting. In fact, essentially all NES-based studies of voter behavior are conceivably contaminated by the types of voter misreports that Wright describes. Most such studies ignore Wright entirely.

votes.

The availability of ballot images allows Lewis, Herron & Sekhon, and Mebane to assess within-ballot behaviors—i.e., counting the number of 2000 general election ballots from a set of Florida counties with both a presidential overvote and a vote for Florida’s Democratic U.S. Senate candidate—without recourse to ecological inference—i.e., using precinct-level overvote rates and U.S. Senate candidate support rates to estimate the fraction of ballots that contain a presidential overvote and a vote for this candidate. As far as we know, there currently exist no ballot-level studies of third party presidential voters.⁴

3 Data and Statistical Methods

For an archive of ballot images that is required for resolving our research objective, we turn to the collection of Florida county general election ballot images maintained by the National Election Study.⁵ This NES ballot image archive contains the complete records of, in theory at least, all ballots cast in ten counties.⁶ These counties used Votomatic punchcard voting technology in 2000 (none use it now due to post-2000 changes in Florida state laws), and each ballot image is a sequence of zeroes and ones where a zero reflects a punchcard chad read by an electronic card reader as not having been punched and a one indicates a chad that was read as punched. Each Votomatic punchcard contained 312 chads, and consequently each ballot image contains a sequence of 312 zeroes and ones. Note that zeros and ones are present even for those chads that do not correspond to any valid candidates.⁷

⁴The ballot images on which we draw have the potential to improve dramatically research on ticket splitting in presidential elections. To see why, note that the controversy associated with Burden & Kimball’s (1998) results on ticket splitting is almost entirely related to Burden & Kimball use of ecological inference to estimate ticket splitting rates (e.g., Cho & Gaines 2004, Herron & Shotts 2004). These rates need not be estimated if one has access to ballot images; rather, they can be calculated directly from them.

⁵The archive is accessible at <http://www.umich.edu/~nes/florida2000/data/ballotimage.htm>.

⁶This caveat reflects that fact that we cannot exactly match vote totals from our ballot images to certified vote totals published by our ten counties. Given the frantic atmosphere that surrounded the Florida recount—see Merzer (2001) and Posner (2001) for details—this should not be considered overly surprising. Furthermore, as noted in fn. 8, the NES ballot archive contains test images or audits files that in some cases have replaced real images. While we can purge the archive of test images, we cannot restore images that were displaced.

⁷Our ballot images do not tell us anything about so-called hanging or pregnant chads, i.e., about punchcard chads that were not completely dislodged from their punchcards. If, say, a chad was pregnant, in the NES archive it would be indicated as not having been punched. One would have to consult the Florida ballot dataset assembled by the National Opinion Research Center to investigate chad-related issues. This dataset, accessible at

For instance, chads in positions one and two on the Palm Beach County ballot do not correspond to anything; the first meaningful punch on the Palm Beach ballot—which was for the presidential race given its presence at the top of the ballot—is position three. Despite this, 179 Palm Beach County voters punched position one and 297, position two. These seemingly errant punches were disregarded by Palm Beach County’s ballot counting mechanisms: they did not count for any candidate and they did not have any vote invalidating effects, either. There is no literature on what these punches mean, assuming that they mean anything at all. The presence of first and second chad punches in Palm Beach County, it is worthwhile pointing out, is not solely due to the county’s butterfly ballot, a subject on which we comment later. In fact, the first chad on Broward County’s 2000 general election ballot—this chad did not correspond to a presidential candidate—was punched by 794 voters.

Since our ten counties of interest had identical voting technology in 2000, our results cannot be confounded by variance in voting machine type across counties or across different types of voters. That is, like Sinclair & Alvarez’s (2004) study of invalid votes in Los Angeles County (only very recently has this county transitioned away from Votomatic machines), our analysis holds voting technology constant. This is a useful aspect of our research design, but one consequence is that we cannot know if our results reflect idiosyncrasies associated with Votomatic punchcards. Fortunately, at least from our perspective, Knack & Kropf (2002) conclude that there are not systematic socioeconomic differences across the types of voters who use different voting technologies. Nonetheless, punchcards are designed for large electorates and should be more common in highly populated counties and metropolitan areas. Punchcard voting technology produces disproportionately high invalid vote rates when compared to optical scan voting systems that alert voters to undervotes and overvotes and also to touchscreen voting systems, which can be programmed to prohibit overvotes (Brady, Buchler, Jarvis & McNulty 2001, The Caltech/MIT Voting Technology Project 2001, United States Commission on Civil Rights 2001, Tomz & van Houweling 2003).

<http://www.norc.uchicago.edu/fl>, only contains data on ballots with invalid presidential votes. In contrast, our research design requires analysis of ballots with invalid presidential votes as well as ballots with valid presidential votes.

Table 2: Valid Presidential Votes among Ten Florida Counties

County	<u>Election Day</u>				<u>Absentee</u>				Total
	Bush	Gore	Nader	Buchanan	Bush	Gore	Nader	Buchanan	
Broward	156876	359255	6512	706	20447	27306	589	82	571773
Highlands	12379	8709	359	84					21531
Hillsborough	157947	155577	6898	798	11161	7390	301	43	340115
Lee	92665	67188	3270	258	13462	6346	314	48	183551
Marion	47324	40652	1632	507	7329	3712	170	55	101381
Miami Dade	265211	311879	5054	516	20882	14138	251	41	617972
Pasco	59881	64096	3107	514	8701	5468	286	56	142109
Palm Beach	130688	241806	5014	3277	16819	19052	390	81	417127
Pinellas	158380	183138	9162	884	26445	17493	860	129	396491
Sarasota	70726	64363	3669	265	12374	8491	400	40	160328
Total	1152077	1496663	44677	7809	137620	109396	3561	575	2952378

Table 2 lists the ten counties whose ballots contribute to the NES ballot image archive, and the table also provides various summary statistics for each county based on presidential vote. In our forthcoming results in Section 4 we do not want to assume that election day and absentee voters have similar partisanship patterns. Table 2, therefore, breaks down each county’s vote totals into election day and absentee totals.

Our ten counties vary dramatically in number of ballots cast, Miami-Dade being the largest with 610,708 total ballots and Highlands the smallest with 22,237. The latter did in fact have absentee voters in 2000, but the NES ballot image archive does not contain any images from these voters. The across county size differences are noteworthy because they have implications for estimate precision. Namely, our statistical method can pin down features of Miami Dade voters in a more precise way than it can for Highlands votes. Nonetheless, the number of voters in Highlands County dwarfs the typical number of voters interviewed by the NES after surrounding general election.⁸

⁸The NES Florida ballot image archive is broken down by county, and each county has what is called a *balfile* for each of its precincts. For instance, Pasco County has 148 balfiles, one per election day precinct (total of 131) and one per absentee precinct (17). The balfiles for the ten counties in Table 2 were assembled by Dan Keating of *The Washington Post*, but some of the balfiles are audit files, i.e., represent a set of Votomatic punchcards that were counted in order to verify ballot counter accuracy. Audit files can override real election balfiles, and we do not want to include audit balfiles in our analysis. Keating’s documentation offers suggestions on how to do this for Palm Beach, Miami Dade, Hillsborough, Pasco, and Highlands Counties; the complement of this group is not known to have audit files in the archive. We followed Keating’s rules on eliminating audit balfiles except for those rules on Hillsborough County. For Hillsborough, and based on a phone conversation with an official in the Hillsborough Supervisor of Elections office, we treated as audit balfiles all precinct balfiles that had exactly 59 ballots in them. We also dropped all precincts to which

As the zeros and ones in each balfile correspond to a sequence of vote choices, we can compare a given ballot's presidential vote with a summary of the ballot's non-presidential votes. Consider, for example, a given ballot from Broward County, which had 58 contested races in November, 2000. Such a Broward ballot records a single voter's choice in the race for president, for Florida's open U.S. Senate seat, for Representative in Florida's 19th Congressional District if applicable, for Representative in Florida's 20th Congressional District if applicable, and so forth. We can ask, then, if a Broward County ballot with a valid vote for Gore, a Democrat, also contains valid Democratic votes in Congressional races, in Broward county races, and so forth. Note that these sorts of questions cannot be studied without ballot images.

Consider, in light of the example, Table 3, which describes the distribution of votes in Florida's U.S. Senate race by presidential vote choice. There were six Senate candidates on Florida's presidential ballot—their names, listed in official Florida order, are in the top row of the table. Beyond the ten official presidential candidates noted in Table 2, each ballot can contain a presidential undervote (a presidential vote that, for some reason, is missing) or an overvote (a vote for more than one presidential candidate). For the Senate race, and henceforth for all non-presidential races as well, Table 3 aggregates both undervotes and overvotes as abstentions.⁹

One can see from Table 3 that 83% of ballots with a valid Bush vote also had a valid McCollum vote, that 86% of the ballots with a valid Gore vote also had a valid Nelson vote, that 7% of valid Gore voters voted for the Republican Senate candidate McCollum, and so forth. We can assemble tables like Table 3 for all non-presidential races, and such an exercise shows that Gore voters tended to vote for Florida's Democratic candidate for U.S. Senate, that Bush voters supported the Republican candidate, and so forth.

Beyond the presidential race, our approximately three million ballot images have the following races in common: U.S. Senate, Florida State Treasurer, Florida State Commissioner of Education, three non-partisan retention votes for justices on the Florida Supreme Court, and three Florida Con-

Keating affixed an "X" and dropped all absentee precincts that did not correspond exactly with official Hillsborough results. Finally, we dropped Hillsborough precincts labeled R433C and R433C1.

⁹Write-in votes for all races are treated as undervotes. The NES ballot image archive does not keep track of write-in candidate names.

Table 3: Votes in Florida’s U.S. Senate Race by Presidential Vote Choice

	Abstain	McCollum (R)	Nelson (D)	Simonetta (Law)	Deckard (Ref)	Logan	Martin	McCormick	Total
Bush	0.04	0.83	0.11	0.00	0.00	0.01	0.00	0.00	1289697
Gore	0.04	0.07	0.86	0.00	0.00	0.02	0.00	0.00	1606059
Nader	0.06	0.27	0.47	0.05	0.04	0.08	0.01	0.02	48238
Buchanan	0.07	0.30	0.46	0.02	0.09	0.03	0.01	0.02	8384
Undervote	0.45	0.21	0.30	0.02	0.00	0.01	0.00	0.00	44969
Overvote	0.26	0.19	0.50	0.01	0.01	0.02	0.01	0.01	59870
Libertarian	0.08	0.30	0.30	0.10	0.06	0.07	0.02	0.07	6791
Socialist Workers	0.39	0.10	0.26	0.07	0.05	0.06	0.04	0.03	345
Natural Law	0.09	0.14	0.20	0.43	0.02	0.07	0.03	0.02	1355
Socialist	0.16	0.14	0.49	0.07	0.04	0.03	0.02	0.03	469
Constitution	0.14	0.39	0.16	0.04	0.08	0.07	0.05	0.06	559
Workers World	0.13	0.23	0.32	0.06	0.04	0.08	0.06	0.09	965

Note: Senate candidates Logan, Martin, and McCormick lacked party affiliations; presidential candidate order reflects official Florida order within the table’s two groupings.

stitutional Amendments. Moreover, many images share other common contests. All images from Palm Beach County, for instance, contain votes on the Palm Beach County sheriff race. Moreover, some ballots from Palm Beach County and some from Highlands County contain votes for Florida's 16th Congressional District seat. The point here is that our many images are linked through a common set of races, and there are many pairwise linkages as well across the images. This is very important for our statistical method—discussed shortly—as we wish to place scalar partisanship measures from our approximately three million ballots in a common policy space, i.e., on a single partisanship line. This would not be possible if the ballots we study lacked races in common.¹⁰

Table 4 describes the distribution of common partisan race votes among ballots with valid Bush, Gore, Nader, and Buchanan votes. A partisan race is one that features competing candidates, and all such common partisan races from our set of ten Florida counties include both Democratic and Republican candidates. The Florida U.S. Senate race included six candidates, the Treasurer race two candidates, and the Commissioner of Education race, three candidates.

Table 4 highlights a pattern discussed earlier, namely that Bush and Gore voters at the presidential level were strong partisans when considering non-presidential races. For instance, 68.62% of Bush voters voted for all three Republican candidates in our collection of common partisan races. Among Gore voters and Democratic candidates, the comparable figure is 57.59%. Note that these two numbers are straight ticket voting rates.

In contrast, and we will comment at length on the implications of this result later, Tables 4c and 4d show that Nader and Buchanan voters did not behave in a strict partisan sense at all. For instance, 18.36% of Nader supporters voted for all three Democratic candidates within our set of common partisan races. However, 14.57% voted for all three Republicans! If Nader voters were strong Democratic partisans, which is what one might think based on the Green Party's ostensible left wing bias, then we would expect these individuals to have supported Democratic candidates in non-presidential races. This did not happen, nor did Buchanan voters vote overwhelmingly for

¹⁰We could, in theory, place our ballot-level partisanship measures on a single line if there were as few as two common races on each of our approximately three million ballots. In our case, the partisanship line is identified by the presence of multiple races.

Distribution of State-wide Votes among Bush, Gore, Nader, and Buchanan Voters

		Votes for Republicans				
		0	1	2	3	Total
Votes for Democrats	0	2.51	2.53	8.20	58.62	81.86
	1	0.54	1.72	10.79		13.05
	2	0.48	3.24			3.72
	3	1.33				1.33
		4.86	7.49	18.99	68.62	100.00

(a) Bush voters ($N = 1289697$)

		Votes for Republicans				
		0	1	2	3	Total
Votes for Democrats	0	2.81	0.81	1.05	2.80	7.47
	1	3.06	2.84	8.32		14.22
	2	7.24	13.44			20.68
	3	57.59				57.59
		70.70	17.09	9.37	2.80	100.00

(b) Gore voters ($N = 1606059$)

		Votes for Republicans				
		0	1	2	3	Total
Votes for Democrats	0	5.53	4.00	6.78	14.57	30.88
	1	6.74	7.57	15.17		29.48
	2	8.67	12.54			21.21
	3	18.36				18.36
		39.30	24.11	21.95	14.57	100.00

(c) Nader voters ($N = 48238$)

		Votes for Republicans				
		0	1	2	3	Total
Votes for Democrats	0	5.13	5.43	12.99	25.44	48.99
	1	4.31	7.56	15.87		27.74
	2	5.05	9.31			14.36
	3	8.87				8.87
		23.36	22.30	28.86	25.44	100.00

(d) Buchanan voters ($N = 5206$)

Table 4: Each cell is the percentage of Florida voters who cast a given number of votes for Democrats and for Republicans across three partisan races: U.S. Senate (six candidates including one Democrat and one Republican), Florida Treasurer (one Democrat and one Republican), and Florida Commissioner of Education (one Democrat, one Republican, and one candidate with no party affiliation). Cells are shaded in proportion to the frequency with which voters fall into them.

Republican candidates in these races.

Henceforth we ignore uncontested races in our ten counties and we also ignore punched chads that do not correspond to any races.¹¹ We also ignore races that allowed voters to pick more than one candidate; these races had very small numbers of voters. Finally, of the three aforementioned Florida Supreme Court retention races (R. Fred Lewis, Barbara J. Pariente, and Peggy A. Quince) and of all other judge races, we retain for analysis only the Supreme Court Lewis retention vote. Pariente and Quince and other local judge races are dropped because they are highly correlated with one another and we do not want one particular set of races to dominate our collection of non-presidential races.

What we mean by highly correlated here is as follows. Consider Table 5, which describes patterns among nine vote choices made by 104,431 voters in Sarasota County; these voters had oppor-

¹¹This is a particularly noteworthy issue in Miami-Dade County, where the ten candidates running for president in Florida correspond to ballot punches 4, 6, 8, ..., 22 as opposed to a sequential sequence of punches (e.g., 4, 5, 6, ..., 13). Of our ten counties, Miami-Dade is unique in this regard. Merzer (2001, p. 179) claims that 1,667 ballots cast in Miami-Dade County contained punches in chads five and seven, i.e., directly below the chads associated with Bush and Gore, respectively.

tunities to vote for nine Democrats and nine Republicans across races contested by both Democrats and Republicans.¹² Of the 104,431 voters, approximately 21% voted for nine Republicans and approximately 14% for nine Democrats. This 35% group is highly partisan, and in contrast there is much variance across the remaining 65% of the 104,431 voters. One can see this in the outer diagonal on Table 5; this string of numbers describes voters who voted in all nine races; 5% of this group voted for eight Republicans and one Democrat, approximately 3% for seven Republicans and two Democrats, and so forth. The point here is that the nine races depicted in Table 5 appear to represent nine independent votes.

Votes for Democrats and Republicans cast across Nine Partisan Races, Sarasota County, Florida

		Votes for Democrats										
		0	1	2	3	4	5	6	7	8	9	Total
Votes for Republicans	0	0.51	0.81	0.38	0.31	0.61	0.36	0.63	1.32	4.34	14.14	23.41
	1	0.55	0.33	0.29	0.31	0.26	0.36	0.59	1.45	3.91		8.05
	2	0.35	0.29	0.34	0.26	0.37	0.51	1.18	2.60			5.90
	3	0.38	0.36	0.30	0.36	0.62	1.25	2.14				5.41
	4	0.88	0.29	0.40	0.65	1.32	2.08					5.62
	5	0.47	0.44	0.66	1.33	2.14						5.04
	6	0.86	0.81	1.49	2.27							5.43
	7	1.97	2.06	2.93								6.96
	8	7.51	5.00									12.51
	9	21.41										21.41
		34.89	10.39	6.79	5.49	5.32	4.56	4.54	5.37	8.25	14.14	100.00

Table 5: Each cell is the percentage of Sarasota County, Florida election-day voters who cast a given number of votes for Democrats and for Republicans across nine partisan races: President, U.S. Senate, Congress (Florida’s 13th district), State Assembly, State Treasurer and Commissioner, Commissioner of Education, Sarasota Supervisor of Elections, Board of County Commissioners, and Sarasota Charter Review Board. Cells are shaded in proportion to the frequency with which voters fall into them. $N = 104,431$.

Such independence is evident as well in Table 6, which depicts the results of 93,640 voters and nine ballot referenda from Tampa, a city in Hillsborough County. As in Table 5’s distribution of votes on partisan races, one can see much variance among the distribution of nine referenda votes.

¹²This means, for instance, that Table 5 ignores Florida State Representative Districts 71 and 74, both of which had uncontested races in 2000.

Approximately 5% of Tampa voters supported none of the referenda, approximately 6% supported all of them, and approximately 7% abstained on all the votes. Many Tampa voters chose to support a limited set of referenda, i.e., approximately 10% voted “yes” on five referenda and “no” on four.

Votes in Favor and Against across Nine Ballot Propositions, Tampa, Florida

		Votes for propositions										
		0	1	2	3	4	5	6	7	8	9	Total
Votes against propositions	0	7.02	0.80	0.44	0.51	0.27	0.31	0.36	0.25	0.53	6.33	16.82
	1	1.76	0.46	0.87	0.39	0.43	0.47	0.41	0.56	5.60		10.95
	2	0.60	0.46	0.37	0.39	0.43	0.54	0.87	6.91			10.57
	3	0.58	0.30	0.43	0.41	0.50	1.04	9.72				12.98
	4	0.36	0.24	0.38	0.50	0.90	10.46					12.84
	5	0.25	0.22	0.40	0.90	8.86						10.63
	6	0.25	0.26	0.66	8.19							9.36
	7	0.31	0.48	5.59								6.38
	8	0.59	3.94									4.53
	9	4.66										4.66
		16.38	7.16	9.14	11.29	11.39	12.82	11.36	7.72	6.13	6.33	100.00

Table 6: Each cell is the percentage of Tampa election-day voters who cast a given number of “yes” and “no” votes across nine ballot propositions. The propositions included three statewide measures (authorizing the construction of high-speed rail lines, changing manner of selecting circuit judges from direct election to retention election, and changing the manner of election county judges from direct election to retention election) and six citywide propositions (eliminating term limits for mayor, eliminating term limits for city council, expand the authority of the Director of Finance, time of the presentation of the city budget, timing of the drawing of new precinct and council district lines, and allowing the hiring of non-city resident election poll workers). Cells are shaded in proportion to the frequency with which voters fall into them. $N = 93,640$.

In contrast to the Sarasota and Tampa votes we have just reviewed, Table 7 displays the distribution of votes in ten judge retention cases from Pinellas County (out use of different counties is illustrative only and it vastly simplifies presentation as our different counties had many different types of non-presidential races). Note how there are three dominant choice patterns among the judge votes: vote in favor of all judges (approximately 43%), against all judges (approximately 13%), and abstain from (17%). In fact, over half of the voters that voted in favor of at least one judge voted to retain all ten.

When we say, then, that the various judge races voted on by our three million Florida voters

Voting on Ten Judge Retention Elections, Pinellas County, Florida

		Votes to retain											
		0	1	2	3	4	5	6	7	8	9	10	Total
Votes to not retain	0	17	1.29	0.97	1.5	1.03	0.3	0.46	0.28	0.45	0.85	42.95	67.08
	1	0.33	0.13	0.25	0.16	0.08	0.09	0.06	0.08	0.14	3.49		4.81
	2	0.09	0.11	0.15	0.05	0.07	0.05	0.07	0.11	2.47			3.17
	3	0.09	0.05	0.03	0.05	0.03	0.05	0.09	1.95				2.34
	4	0.14	0.02	0.03	0.03	0.05	0.09	1.98					2.34
	5	0.02	0.02	0.02	0.03	0.06	1.58						1.73
	6	0.07	0.02	0.03	0.05	1.27							1.44
	7	0.02	0.02	0.04	1.05								1.13
	8	0.06	0.03	1.43									1.52
	9	0.12	1.1										1.22
	10	12.89											12.89
		30.83	2.79	2.95	2.92	2.59	2.16	2.66	2.42	3.06	4.34	42.95	100.00

Table 7: Each cell is the percentage of Pinellas county election day voters who voted on a collection of ten judges facing retention. Cells are shaded in proportion to the frequency with which voters fall into them. $N = 360,266$.

were highly correlated, what we mean is that an affirmative vote in one of these races from a given ballot provides much information about votes on the other nine judge votes on the ballot. Since our statistical methodology, which we now explain, treats as independent all non-presidential votes on a ballot, we henceforth drop all judge races except for the Lewis Supreme Court retention vote.

Our methodology assigns a partisanship measure to each of our ballot images using a statistical procedure known as scaling. Details are in Lewis (2001) and they are briefly summarized in Appendix A. Here we provide intuition.

Consider a single ballot image; as noted this image contains a sequence of non-presidential votes, all of which can be treated as independent choices, the only exception being the judge votes we just discussed. If we knew the partisanship of the voter who produced the image, we could estimate the effect that this partisanship had on each of his or her votes. It would be natural to use logit models for these various estimation problems. For instance, we could use a logit model to analyze whether extreme Democratic partisans tended to vote Democratic in Florida's U.S. Senate race. Some races will split on the partisan dimension, meaning that voter partisanship predicts vote

choice very reliably, but some will not. Moreover, some races will involve incumbency or a valence advantages for one contestant, meaning that this individual will receive support from a large set of voters regardless of their underlying partisanship.

Of course, we cannot observe voters' partisanship on account of ballot secrecy, and therefore logit models like the ones discussed above are not feasible. Nonetheless, the objective of scaling is to use choices on a sequence of votes (here, non-presidential races) to estimate *simultaneously* voter partisanship levels and the effect of voter partisanship on vote choice. This is akin to using vote choices of members of Congress to estimate ideology levels or ideal points (Poole & Rosenthal 1997) and votes of Justices to study partisanship on the Supreme Court (Martin & Quinn 2002).

We divide our voters into $10 \times 12 \times 2 = 240$ different groups based on county. This cross-product reflects the ten counties listed in Table 2, the type of presidential vote (undervote, valid vote for one of ten candidates, or overvote), and time of vote (election day or absentee). Then, for each type of our 240 types we scale the votes on all observed voting profiles for contests down the ballot, ignoring most judge races and so forth as discussed above. As noted in the introduction, this necessitates contending with 46,515,369 different vote choices. The result of our scaling algorithm is a partisanship measure for each of our 240 voter groups, and we can also generate estimated standard errors for these measures.

In this paper we consider only a subset of the 240 types of voters, and in particular we focus here on the ballots from our ten counties with valid votes for Bush, Gore, Nader, and Buchanan cast on election day and via absentee voting. This yields a total of $10 \times 4 \times 2 = 80$ different types of voters. In future papers we will describe the partisanship measures of presidential undervoters and overvoters and will also examine the partisanship of voters who supported very minor third party candidates in 2000, i.e., the Workers World candidate. Virtually nothing is known about these voters, due primarily to the fact that their numbers are so small. Yet, as the 2000 has shown, small groups of voters can be pivotal and hence deserve scrutiny.

As discussed in Appendix A we assume without loss of generality that low (i.e., left) partisanship levels are associated with Democratic preferences and large (i.e., right) values with Republican

partisanship. We also assume, again without loss of generality, that partisanship falls between negative and positive one. Finally, we assume that each of our 240 voter types has partisanship normally distributed with a mean to be estimated and variance one. The objective of our scaling algorithm, then, is estimation of 240 different means or locations where the mean for a given voter type—e.g., election day Gore from Lee County—describes the partisanship of the type.

4 Results

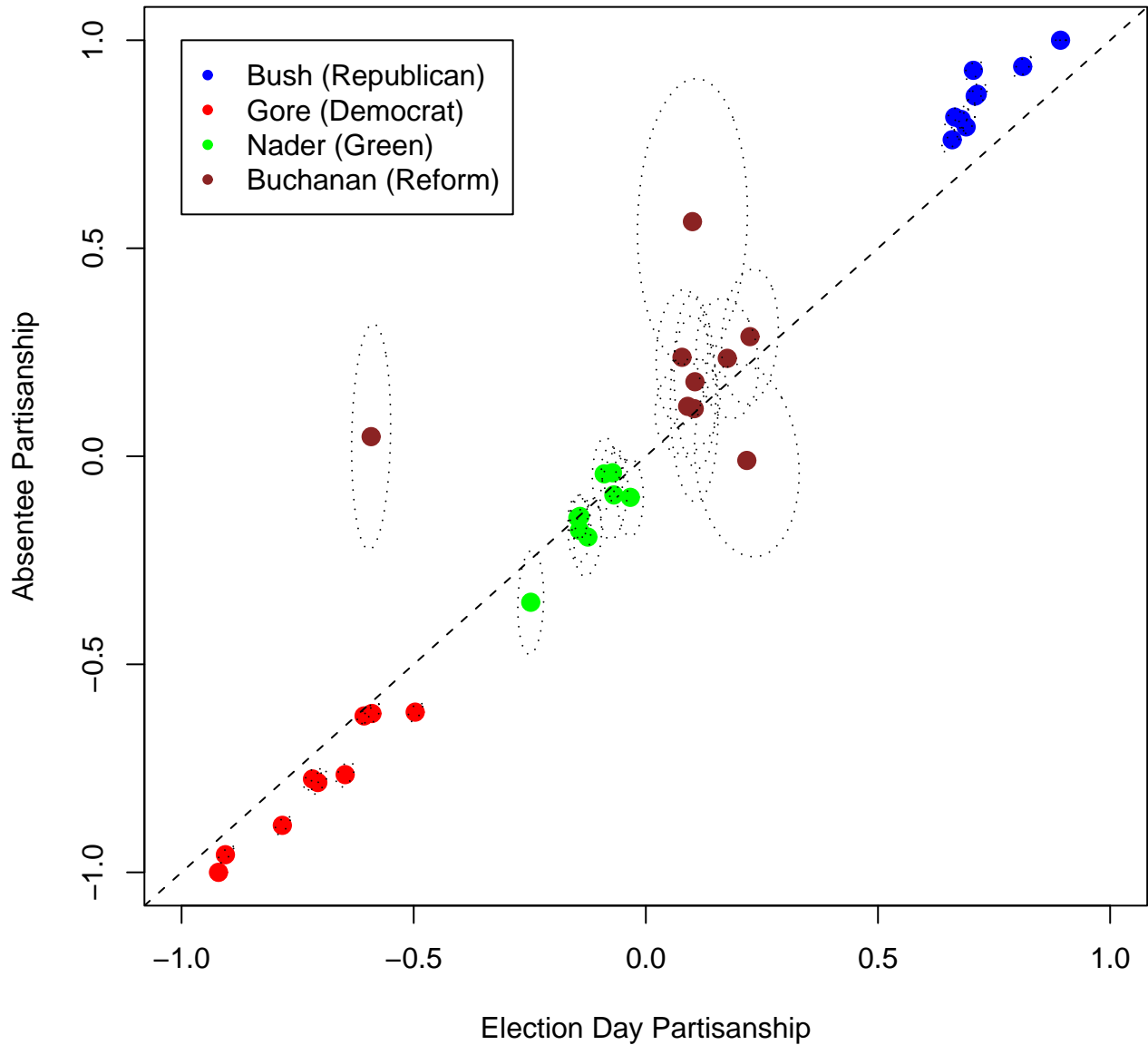
Our ballot scaling procedure produces many different estimates. In this section we describe some of them, beginning with estimated partisanship levels for 80 different voter types.

4.1 Partisanship Levels by Presidential Voter Type

Figure 1 displays estimated partisanship levels for Bush, Gore, Nader, and Buchanan voters. Each dot in the figure represents a county, and the location of a dot in the plot specifies partisanship levels for a pair of election day and absentee voters. Highlands County is not in this figure because, as already noted, we have no records on absentee voters from this county. The ellipses in Figure 1 are 95% confidence sets based on 70 bootstrap repetitions. The confidence sets are longer vertically than they are horizontally, and this is because there the number of election day voters exceeds that of absentee voters. It is easier, then, to estimate reliably the partisanship levels of the former. Note that some confidence sets in Figure 1 are so small that they are completely masked by a colored dot. This means that our scaling algorithm is pinning down these partisanship levels extremely well.

The red dots in Figure 1 are in the lower left corner of the figure, meaning that their election day and absentee partisanship levels are both close to negative one. This reflects the fact that voters who selected Al Gore, the Democratic presidential candidate in 2000, were very committed Democratic partisans insofar as voting Democratically in non-presidential contests. A related statement applies to Figure 1's blue dots, which represent Bush voter partisanship levels. The overall point here is that Gore and Bush voters displayed strong allegiances to non-presidential Democratic and Republican

Figure 1: Locations of Presidential Voters



candidates, respectively, as they considered races beyond the presidential contest.

Note that the red dots in Figure 1 fall under the figure's dashed 45-degree line yet the blue dots are over the line. This means that absentee voters who supported Gore and Bush voted in a more partisan way than their election day counterparts. There are a number of potential explanations for

this result. Absentee voters may in general tend to be more committed partisans than election day voters. Or, absentee voters may have similar partisanship levels as election day voters, all of whom may want to vote straight party, yet the former may commit fewer voting errors due to a lack of time pressure.

Figure 1 also displays the partisanship of Nader voters (green dots) and Buchanan voters (brown dots), and green and brown dot locations in Figure 1 show that Green party voters were slightly pro-Democratic and Reform party voters, slightly pro-Republican. One clear exception to this rule is the brown dot with an election day mean of approximately -0.5 and an absentee mean of slightly greater than zero. This dot's election day coordinate is relatively Democratic and in fact is more Democratic than Lee County's Democratic location on election day and equivalent to Pasco County's election day level. The Democratic-looking Buchanan dot is from Palm Beach County, and its location reflects the county's butterfly ballot. We discuss this issue shortly.

What is perhaps most striking about the estimated Nader and Buchanan locations in Figure 1 is how non-partisan they are. Ralph Nader's platform in 2000 was pro-environment and anti-free trade, and both of these policy positions place Nader close to stereotypical Democratic preferences. Despite this, Nader supporters were clearly mavericks in the poll booth, and they appear to have had no loyalty to Democratic candidates. Note that the cloud of green points in Figure 1 is clearly distinct from the corresponding cloud of red points (and that the standard error ellipses surrounding the green dots do not intersect the standard error ellipses from the red dots).

Similarly, Figure 1 shows that Buchanan voters were only mildly pro-Republican, despite the fact that Pat Buchanan was perceived as being much closer to standard Republican positions than to Democrats. If Reform party supporters had voted spatially, then their partisanship levels should have been much closer to those of Republican supporters.

One possible explanation for the apparently non-partisan locations of Nader and Buchanan voters is that those voters might tend to abstain from voting in non-presidential contests. Our scaling algorithm treats abstentions as missing at random, i.e., independent of the underlying Democratic-Republican partisanship dimension. So, if abstention were more likely among those at the ends of

Rate of down ballot participation by Presidential vote choice and county

President	Broward	Highlands	Hillsborough	Lee	Marion	Miami-Dade	Palm Beach	Pinellas	Sarasota
Republican	0.77	0.75	0.79	0.82	0.77	0.48	0.78	0.83	0.81
Democrat	0.75	0.71	0.77	0.79	0.74	0.49	0.77	0.80	0.79
Libertarian	0.65	0.77	0.75	0.77	0.77	0.49	0.75	0.80	0.77
Green	0.79	0.75	0.78	0.81	0.75	0.56	0.77	0.81	0.79
Socialist Workers	0.44	0.98	0.45	0.73	0.65	0.22	0.43	0.58	0.63
Natural Law	0.72	0.78	0.69	0.78	0.62	0.40	0.69	0.79	0.70
Reform	0.78	0.80	0.78	0.80	0.76	0.53	0.69	0.83	0.79
Socialist	0.60	0.76	0.46	0.97	0.64	0.38	0.59	0.79	0.53
Constitution	0.77	0.83	0.73	0.82	0.75	0.38	0.60	0.83	0.85
Worker World	0.77	0.80	0.72	0.84	0.75	0.40	0.55	0.86	0.84
Undervote	0.39	0.53	0.46	0.44	0.57	0.20	0.52	0.40	0.39
Overvote	0.58	0.61	0.57	0.69	0.55	0.39	0.69	0.67	0.64

Table 8: Shows the percent of down-ballot races in which election-day voters for each presidential candidate in each county participated.

the partisan dimension and those voting for minor party candidates were more likely to abstain on non-presidential races, then our estimates of third party voter partisanship would be biased toward the center. However, as shown in Table 8, Nader and Buchanan voters were not systematically more likely to abstain in down-ballot contests compared to Gore and Buchanan voters. While the presence of Nader and Buchanan on the general election ballot may have drawn new voters to the polls, these voters did not head home after casting a presidential vote. Rather, they voted in other elections at relatively high rates.

4.2 Posterior Distributions of Voter Partisanship

A key assumption in our scaling algorithm is that, unconditionally, the partisanship for a given voter type (e.g., Bush election day voters from Broward County) is normally distributed with a mean to be estimated and unit variance. We have estimated group means for our 80 different voter types (actually 240 types, but here we are commenting only on the former) and can now address using

Bayes Rule, say, the posterior distribution of voter partisanship conditional on a vote for Bush.

Figure 2 describes posterior distributions for our 80 voter types. It divides the posteriors in several ways: by presidential vote type (vertically in the figure), by election day versus absentee voting (horizontally), and within county (by square).

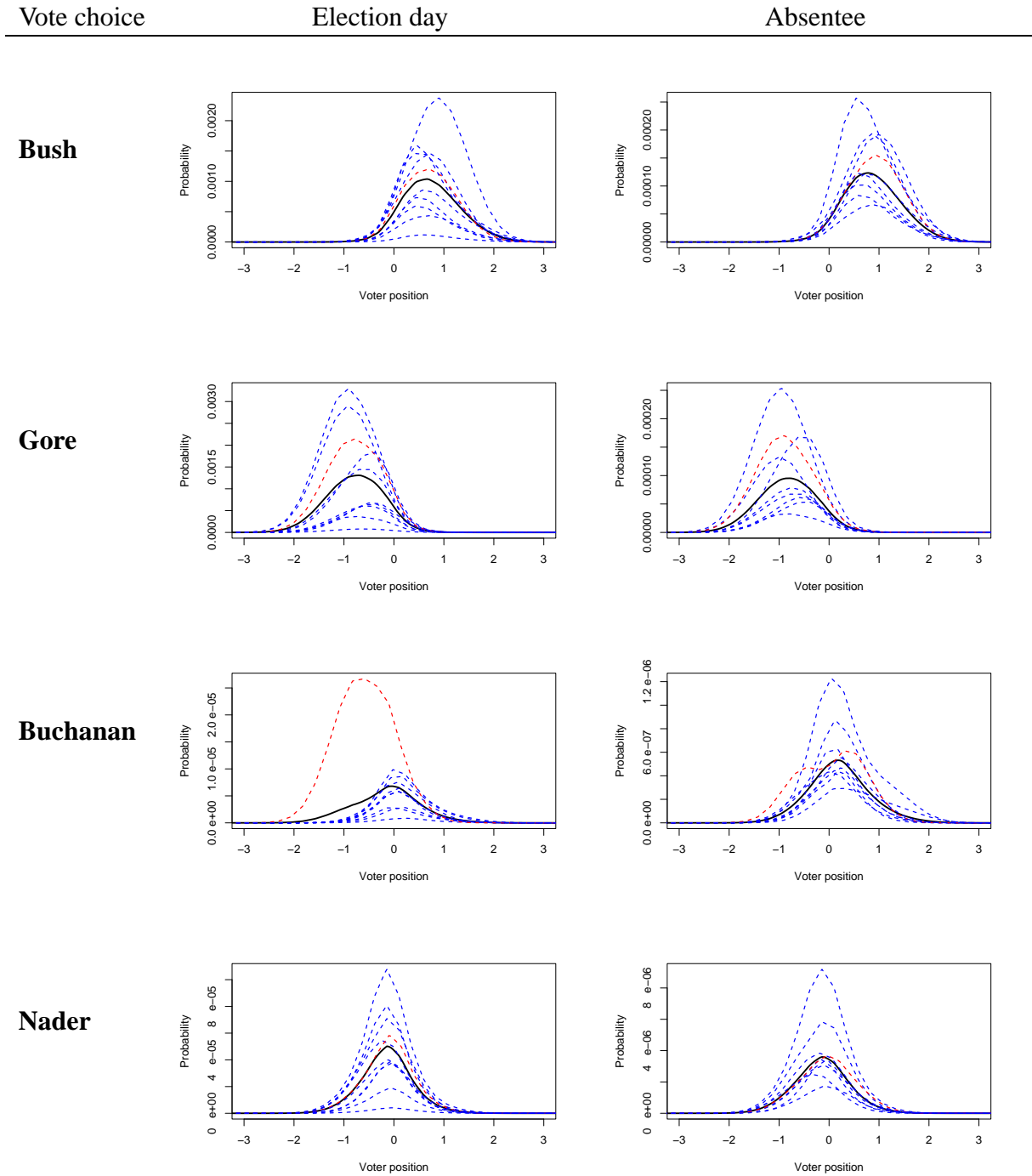
Recall that we characterized partisanship so that small values reflect Democratic voting patterns and large values, Republican voting. With this in mind, the Bush and Gore posteriors in Figure 2 are highly intuitive and, it could be argued, function as a consistency check on our overall results. The Bush posteriors in the figure, both election day and absentee, are located toward the right hand side of our unidimensional partisanship space, and there does not appear to be any dramatic election day versus absentee differences. Similar comments apply to Gore voter posterior distributions: they are toward the left of our partisanship line, which is intuitive, and do not manifest large election day versus absentee differences.

With respect to Buchanan and Nader posteriors, Figure 2 displays strong evidence of partisan centrism (we comment on the red dashed Palm Beach posteriors shortly). Note that the Buchanan posteriors are marginally on the right hand side of our unidimensional partisanship policy space, and Nader posteriors are to the left. It follows, then, that Buchanan and Nader voters were not committed Republicans and Democrats, respectively, on non-presidential races, and this is directly contrary to the notion of Nader as a Gore spoiler in November, 2000.

The absentee Buchanan posteriors have greater variances than comparable election day posteriors. This could mean that absentee Buchanan voters were more ideologically mixed compared to their election day counterparts. Alternatively, the former are prone to excessive voting errors. Since, though, voting errors are presumably more likely in a time-pressured environment like an election day precinct, this second explanation is not particularly plausible.

With respect to Palm Beach County, consider the dotted red posteriors in the Buchanan row of Figure 2. Among election day posteriors for Buchanan voters, the Palm Beach County posterior is an extreme oddity: it is shifted way to the political left compared to the other nine election day posteriors with which it is logically compared. Moreover, and still focusing solely on election day

Figure 2: Posterior Distributions of Voter Partisanship



Note: Dashed red lines represent posterior distributions from Palm Beach County, dashed blue lines are posteriors from other counties, and dashed black lines are posteriors from all ten counties combined.

posteriors, the Palm Beach County Buchanan posterior does not appear appreciably different than the various Gore posteriors depicted in the row of distributions immediately above it. With respect to partisanship, then, Buchanan voters on election day were barely distinguishable from election day Gore voters. This reflects Palm Beach County's butterfly ballot.

Absentee voters in Palm Beach County did not use a butterfly ballot, and this means that it is logical to compare absentee and election day posteriors for Buchanan voters. Note that the absentee Buchanan posterior for Palm Beach County is not very different than comparable absentee posteriors for other counties. This is, of course, quite a contrast with the election day Buchanan situation. And, the absentee Palm Beach posterior is very different than the election day posterior. Note that the absentee Palm Beach posterior for Buchanan voters is bimodal. This is somewhat puzzling, and we suspect that it might reflect data contamination, i.e., some of our ballot images that are labeled as absentee by the NES ballot archive may in fact be election day images.

Overall, the features we have noted in various Buchanan posteriors speak very strongly to the theory that the butterfly ballot used in Palm Beach County led Democrats to vote accidentally for Buchanan. Among our ten counties we do not observe any that contain Buchanan posteriors that approximate what is evident in Palm Beach.

4.3 Counterfactual Analysis: Allocating Votes to Major Party Candidates

Ultimately, the matter of whether Nader was a spoiler for Gore depends on how Nader voters would have voted had they treated the 2000 general election as a Bush versus Gore contest. In light of this, we now consider how to reallocate Buchanan and Nader voters in a way that is consistent with the partisanship measures we have estimated and presented in earlier sections. We first discuss vote reallocations based on our estimated voter partisanship and then we reallocate Nader and Buchanan votes in a non-parametric way based on Table 4.

Table 9: Reallocating Buchanan and Nader Voters to Gore

County	Absentee				Election day			
	Nader		Buchanan		Nader		Buchanan	
	Percent	Swing	Percent	Swing	Percent	Swing	Percent	Swing
Broward	0.63	155	0.42	-13	0.64	1826	0.48	-35
Highlands					0.51	10	0.34	-27
Hillsborough	0.59	57	0.34	-14	0.59	1202	0.37	-214
Lee	0.60	60	0.54	4	0.57	437	0.41	-49
Marion	0.52	8	0.32	-20	0.53	98	0.41	-88
Miami Dade	0.71	107	0.42	-6	0.66	1663	0.43	-73
Palm Beach	0.57	58	0.51	1	0.62	1233	0.83	2176
Pasco	0.56	35	0.42	-9	0.58	485	0.46	-43
Pinellas	0.63	218	0.39	-27	0.63	2431	0.42	-145
Sarasota	0.62	96	0.25	-20	0.60	733	0.44	-30
Total	0.61	794	0.41	-103	0.61	10117	0.59	1471

Note: Percent refers to the fraction of a given voter type that is allocated to Gore. Swing is the number of votes Gore received from this reallocation minus the number of votes Bush received.

4.3.1 Reallocations from Estimated Voter Partisanship

Our reallocation scheme that uses estimated voter partisanship levels calls for reallocating Buchanan and Nader voters at a given location in our unidimensional partisanship space to Bush and Gore in proportion to the shares that these two candidates were estimated to have received among other voters at the position. Reallocation results are displayed by county and by time of voting in Table 9. In addition, the bottom row of the table show the total number of votes that Nader voters and Buchanan voters would have contributed to Bush and Gore had they voted for one of these two candidates.

The various “Percent” columns in Table 9 indicate the fraction of a county’s election day or absentee ballots that would have been cast for Gore rather than Bush, and the swing is the difference between reallocated Gore votes and reallocated Bush votes. Positive swing numbers, then, highlight gains for Gore.

The Gore swing numbers in Table 9 are quite small, and this is true for both election day and absentee reallocations. Note that the absentee swings are smaller than the election day swings due

to the relative paucity of absentee voters. The largest Gore swing is found in Pinellas County where, we find, Gore lost 2,431 on account of Nader's presidential candidacy.

In Broward County, for instance, we estimate that Nader and Buchanan voters combined would have contributed 1,933 votes to Gore had the 2000 presidential election been a two candidate race (this is the sum of the Gore swings in the top row of Table 9). This number is tiny compared to the number of ballots we have considered. Indeed, 1,933 can only be thought of as large because the Bush-Gore margin in Florida was so incredibly tight. In fact, the total Nader swing away from Gore is 10,911 votes (this combines election day and absentee allocations), meaning that Gore lost only 10,911 votes in our ten counties due to Nader's candidacy.

Still, the Nader swing figures in Table 9 are all positive, and this implies that Gore lost relative to Bush as a consequence of Nader. With respect to Buchanan, the swing figures are mostly negative and smaller than comparable Nader swings. Hence, Bush lost votes to Gore thanks to Buchanan's candidacy.

Without Palm Beach County, on which we comment below, Buchanan voters supported Gore at a rate of approximately 42%. So, 58% of such voters would have voted for Bush had neither Nader nor Buchanan run for president in 2000, and this is remarkably close to the 61% figure we calculate for Gore support among Nader voters.

An exception to pro-Bush nature of Buchanan voters is the positive election day Gore swing in Palm Beach County due to Buchanan. The positive sign of this swing in contrast with the fact that the other nine Buchanan-related election day swings in Table 9 are negative point directly to the Palm Beach County's butterfly ballot. To be precise, we estimate that Gore's net loss of votes to Bush as a consequence of Buchanan's candidacy in conjunction with the butterfly ballot was at least 2,176 votes. It is certainly greater than this number insofar as, according to our data, Palm Beach County's true Gore swing due to Buchanan should be negative and not simply zero. Our estimate of the butterfly ballot effect for Gore is roughly comparable to the estimated butterfly effects (between approximate 2,456 and 2,973 votes lost to Gore) described in Wand et al. (2001)

4.3.2 Reallocations based on Aggregate Democratic and Republican Votes

We now return to Table 4 and consider Nader and Buchanan vote reallocations based on number of Democratic and Republican votes cast across the common partisan races that existed in Florida. Namely, we first divide our ballots into groups based on county, time of vote (election day versus absentee), and number of Democratic and Republican votes cast among these three races. Then, based on frequencies in which Gore and Bush voters existed in these groups, we reassign Nader and Buchanan voters to Gore and Bush.

For example, if in one such group, e.g., election day voters in Pinellas county who voted for two Democrats and one Republican among the common partisan races, 60% of voters supported Gore, then we assume that 60% of election day Nader voters in Pinellas County who voted for two Democrats and one Republican in the common races would also have voted for Gore. This non-parametric reallocation scheme is independent of our estimated partisanship levels and functions as a consistency check on our scaling results. If our non-parametric reallocations are dramatically different than our reallocations based on scaling, then this would cast doubt on our scaling approach in general.

In fact, according to our non-parametric reallocations, 61% of Nader and 45% of Buchanan supporters would have voted for Gore had the 2000 presidential contest been a two-candidate race. These numbers, particularly the Nader figure, are very close to the partisanship reallocations figures we have discussed before.

5 Conclusion

Still to come...

A The voting model and estimator

Consider an electorate comprised of voters with quadratic preferences or ideal points over a single partisan dimension. In an election each voter casts a vote in contests $k = 1, 2, \dots, K$ where each contest includes J_k candidates. Index the candidates in each contest by $j = 1, 2, \dots, J_k$ and let c_{jk} be the location on the partisan dimension of candidate j in race k . The utility of candidate j to voter i with ideal point θ_i in race k is

$$U(\theta_i, c) = v_{jk} - (\theta_i - c_{jk})^2 + \epsilon_{ijk}$$

where v_{jk} reflects non-policy/partisan utility provided by candidate j (the value of incumbency, for example) and ϵ_{ijk} is an idiosyncratic utility shock which follow a type I extreme value distribution. Voter i selects candidate j in race k if

$$U(\theta_i, c_{jk}) \geq U(\theta_i, c_{j'k}) \text{ for } j' = 1, 2, \dots, J_k.$$

Note that the continuity of the distribution of idiosyncratic shocks (ϵ) insures that $U(\theta_i, c_{jk}) = U(\theta_i, c_{j'k})$ for $j \neq j'$ is a zero probability event and thus the possibility that a voter is indifferent between two candidates can be ignored.

As shown by McFadden (1974), the assumption of independent type I extreme value shocks means that the probability a voter with ideal point θ supports candidate j in race k is

$$\Pr(Y_k = j|\theta) = \frac{\exp(v_{jk} - (c_{jk} - \theta)^2)}{\sum_{j'=1}^J \exp(v_{j'k} - (c_{j'k} - \theta)^2)} \text{ for } j = 1, 2, \dots, J.$$

Rearranging the above we find that

$$\Pr(Y_k = j|\theta) = \frac{\exp(v_{jk} - c_{jk}^2 + 2c_{jk}\theta)\exp(-\theta^2)}{\exp(-\theta^2) \sum_{j'=1}^J \exp(v_{j'k} - c_{j'k}^2 + 2c_{j'k}\theta)} = \frac{\exp(v_{jk} - c_{jk}^2 + 2c_{jk}\theta)}{\sum_{j'=1}^J \exp(v_{j'k} - c_{j'k}^2 + 2c_{j'k}\theta)}.$$

Letting $\alpha_{jk} = v_{jk} - c_{jk}^2$ and $\beta = 2c_{jk}$ produces the familiar multinomial logit model,

$$\Pr(Y_k = j|\theta, \boldsymbol{\alpha}_k, \boldsymbol{\beta}_k) = \frac{\exp(\alpha_{jk} + \beta_{jk}\theta)}{\sum_{j'=1}^J \exp(\alpha_{j'k} + \beta_{j'k}\theta)}.$$

Conditional on θ , votes for candidates across races are independent. Thus, the probability of voting for a particular set of candidates can be written as

$$P(j_1, j_2, \dots, j_K|\theta, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \Pr(Y_1 = j_1, Y_2 = j_2, \dots, Y_K = j_K) = \prod_{k=1}^K \Pr(Y_k = j|\theta, \boldsymbol{\alpha}_k, \boldsymbol{\beta}_k).$$

These probabilities could be evaluated if θ were observed. However, we treat θ as a random variable. In particular, we partition our approximately three million voters into 240 county, presidential-vote, voting-place (election day or absentee) triples. For example, one such group is (Gore, Broward, Election Day). The distribution of ideal points θ within each of these groups $g = 1, 2, \dots, 240$ is assumed *a priori* to be normal with mean μ_g and standard deviation one. Thus, *a priori* we assume the distribution of preferences within each voter group differs only by a shift parameter μ . Fixing the standard deviation to one results in no further loss of generality as this restriction only serves to identify the otherwise unidentified the units of the underlying dimension.

We marginalize the distribution of vote choices of voters from each group g with respect to θ so that

$$P(j_1, j_2, \dots, j_K|\boldsymbol{\alpha}, \boldsymbol{\beta}, \mu_g) = \int P(j_1, j_2, \dots, j_K|\theta, \boldsymbol{\alpha}, \boldsymbol{\beta})\phi(\theta|\mu_g)d\theta.$$

The parameter matrices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ and the parameter vector μ are estimated by marginal maximum likelihood. We group voters into common patterns of votes cast across the k offices, $p = 1, 2, \dots, J_1 J_2 \dots J_K$ within each group, let n_{pg} be the number of voters casting vote pattern p in group g , and let $p(g, k)$ represent the candidate in the k th race chosen by a voter in the g th group casting the p th voter pattern. Then, the loglikelihood is

$$L(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\mu}) = \sum_g \sum_p n_{pg} \ln P(p(g, 1), p(g, 2), \dots, p(g, K)|\boldsymbol{\alpha}, \boldsymbol{\beta}).$$

This likelihood can be maximized by standard numerical techniques (using Gauss-Hermite quadrature to approximate the normal integral) or by an EM approach.

To identify the model, the valance and position of the first candidate in each race is normalized to zero. That is, we assume $v_{1k} = 0$ and $c_{1k} = 0$ (and consequently $\alpha_{1k} = 0$ and $\beta_{1k} = 0$) for $k = 1, 2, \dots, K$. This is the standard normalization used in multinomial logit models. The candidate valance and locations can then be thought of as relative to the valance and location of the first candidate in each race. It should be noted to that this precludes the possibility of directly comparing valances and locations of candidates across races. For such comparisons additional identifying restricts are required. For example, the average valance might be assumed to be zero in each race. Finally, we linearly transform our (without loss of generality) so that the estimates of μ s range from -1 to 1 where $\mu = -1$ is the most Democratic group and $\mu = 1$ for the most Republican group.

Standard errors are arrived at through a non-parametric bootstrap in which individual voters are sampled with replacement from the full dataset. The sampling is stratified by voter type so that each voter type appears as many times in each bootstrap sample as is the original data. Standard error estimates are based on 70 bootstrap samples. Confidence ellipses presented in Figures 1 and 1 are based the assumption of normal sampling distributions and the covariance matrix of the estimates derived from the bootstrap.

Given estimates of α , β , and μ , the posterior distribution of θ for a given voter group conditional on observed vote choices is calculated by Bayes rule. The allocation of voters to Bush and Gore (see Table 9) are made by finding the estimated probability that a voter located at each position voted for Gore conditional on voting for either Bush or Gore and then taking the expectation of these conditional probabilities over various posterior distributions.

References

- Achen, Christopher H. & W. Phillips Shively. 1995. *Cross-Level Inference*. Chicago: University of Chicago Press.
- Alvarez, R. Michael & Jonathan Nagler. 1995. "Economics, Issues, and the Perot Candidacy: Voter Choice in the 1992 Presidential Election." *American Journal of Political Science* 39(3):714–44.
- Brady, Henry E., Justin Buchler, Matt Jarvis & John McNulty. 2001. "Counting All the Votes: The Performance of Voting Technology in the United States." Report published by the Survey Research Center and Institute of Governmental Studies, University of California, Berkeley.
- Burden, Barry C. & David C. Kimball. 1998. "A New Approach to the Study of Ticket Splitting." *American Political Science Review* 92(3):533–44.
- Cho, Wendy K. Tam & Brian J. Gaines. 2004. "The Limits of Ecological Inference: The Case of Split Ticket Voting." *American Journal of Political Science* 48(1):152–171.
- Freedman, D.A., S.P. Klein, M. Ostland & M.R. Roberts. 1998. "On 'Solutions' to the Ecological Inference Problem." *Journal of the American Statistical Association* 93(444):1518–22.
- Herron, Michael C. & Jasjeet S. Sekhon. 2003. "Overvoting and Representation: An examination of overvoted presidential ballots in Broward and Miami-Dade counties." *Electoral Studies* 22(1):21–47.
- Herron, Michael C. & Kenneth W. Shotts. 2004. "Logical Inconsistency in EI-based Second Stage Regressions." *American Journal of Political Science* 48(1):172–183.
- King, G. 1999. "The Future of Ecological Inference Research: A Reply to Freedman et al." *Journal of the American Statistical Association* 94(445):352–55.
- King, Gary. 1997. *A Solution to the Ecological Inference Problem*. Princeton, NJ: Princeton University Press.

- Knack, Stephen & Martha Kropf. 2002. "Who uses inferior voting technology?" *PS: Political Science and Politics* 35(3):541–8.
- Lacy, Dean & Barry C. Burden. 1999. "The Vote–Stealing and Turnout Effects of Ross Perot in the 1992 U.S. Presidential Election." *American Journal of Political Science* 43(1):233–55.
- Lewis, Jeffrey B. 2001. "Estimating Voter Preference Distributions from Individual-Level Voting Data." *Political Analysis* 9(3):275–297.
- Martin, Andrew D. & Kevin M. Quinn. 2002. "Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the U.S. Supreme Court, 1953-1999." *Political Analysis* 10(2):134–153.
- McFadden, Daniel. 1974. "The Measurement of Urban Travel Demand." *Journal of Public Economics* 3:308–328.
- Mebane, Jr., Walter R. 2003. "The Wrong Man is President! Overvotes in the 2000 Presidential Election in Florida." Forthcoming, *Perspectives on Politics*.
- Merzer, Martin. 2001. *The Miami Herald Report: Democracy Held Hostage*. New York: St. Martin's Press.
- Poole, Keith T. & Howard Rosenthal. 1997. *Congress: A Political–Economic History of Roll Call Voting*. New York: Oxford University Press.
- Posner, Richard A. 2001. *Breaking the Deadlock: The 2000 Election, the Constitution, and the Courts*. Princeton, NJ: Princeton University Press.
- Rosenstone, Steven J., Roy L. Behr & Edward H. Lazarus. 1996. *Third Parties in America: Citizen Response to Major Party Failure*. 2 ed. Princeton, NJ: Princeton University Press.
- Sinclair, D. E. & R. Michael Alvarez. 2004. "Who Overvotes, Who Undervotes, using Punchcards? Evidence from Los Angeles County." Forthcoming, *Political Research Quarterly*.

- Sinclair, Robert C., Melvin M. Mark, Sean E. Moore, Carrie A. Lavis & Alexander S. Soldat. 2000. "An electoral butterfly effect." *Nature* 408:665–666.
- Smith, Richard L. 2002. "A Statistical Assessment of the Buchanan's Vote in Palm Beach County." *Statistical Science* 17(4):441–457.
- Tam, Wendy K. 1998. "If the Assumption Fits...: A Comment on the King Ecological Inference Solution." *Political Analysis* 7:143–63.
- The Caltech/MIT Voting Technology Project. 2001. "Residual Votes Attributable to Technology: An Assessment of the Reliability of Existing Voting Equipment." Working draft available from <http://www.vote.caltech.edu>.
- Tomz, Michael & Robert P. van Houweling. 2003. "How Does Voting Equipment Affect the Racial Gap in Voided Ballots?" *American Journal of Political Science* 47(1):46–60.
- United States Commission on Civil Rights. 2001. "Election Reform: An Analysis of Proposals and the Commission's Recommendations for Improving America's Election System."
- Wand, Jonathan N., Kenneth W. Shotts, Jasjeet S. Sekhon, Walter R. Mebane, Jr., Michael C. Herron & Henry E. Brady. 2001. "The Butterfly Did It: The Aberrant Vote for Buchanan in Palm Beach County, Florida." *American Political Science Review* 95(4):793–810.
- Wright, Gerald. 1993. "Errors in Measuring Vote Choice in the National Election Studies, 1952–1988." *American Journal of Political Science* 37(1):543–563.
- Wright, Gerald C. 1990. "Misreports of Vote Choice in the 1988 NES Senate Election Study." *Legislative Studies Quarterly* 15(4):543–563.
- Wright, Gerald C. 1992. "Reported versus Actual Vote: There is a Difference and It Matters." *Legislative Studies Quarterly* 17(1):131–143.