Recounts From Undervotes: Evidence From the 2000 Presidential Election

Bruce E. HANSEN

The vote recount in the 2000 Presidential election (Broward, Miami-Dade and Palm Beach Counties, Florida) is examined for evidence of bias. A precinct-level dataset is constructed, incorporating the machine-vote tally, the recount vote tally, voter registration demographics, and the ballot review by media sources. A new multivariate beta-logit model is introduced that allows joint modeling of multivariate unobserved latent probabilities. A simple two-step estimator is proposed that approximates the joint maximum likelihood estimator. The estimates are consistent with a strong hypothesis: that the recount vote tally was unbiased. Specifically, it is found that the precinct-level machine-vote probability for a candidate is an unbiased predictor for the hand-recount undervote probability. There is no evidence of bias in the recount.

KEY WORDS: Beta-logit; Election; Latent variable; Overdispersion; Votes.

1. INTRODUCTION

In the State of Florida, the November 2000 Presidential election was decided in favor of Texas Governor George W. Bush by an official margin of 537 votes. The vote was bitterly contested, with the Democratic Party appealing for hand recounts in four counties. The core dispute concerned three large counties in Southern Florida (Broward, Miami-Dade, and Palm Beach) that used punch-card ballots, and had a large number of *undervotes*, ballots that registered no Presidential vote in the machine tally, and *overvotes*, ballots that registered multiple Presidential votes. Undervotes (rather than overvotes) were the primary focus of the recount effort. The debate centered on whether a voter's intent could be determined without bias from a hand examination of a punch-card undervote. This debate rests on two questions: (1) what is the underlying *cause* of undervotes, and (2) will a hand recount be *biased*.

A punch-card ballot contains small rectangular pieces called "chads." The voter inserts the punchcard into a voting machine, which lines up the chads with the candidate names, and uses a stylus to punch the ballot. According to design, the chad should fall away, leaving a clean hole, which should be read as a valid vote by the counting machine.

Many ballots were not machine readable, however, and hence were undervotes. Some of the undervotes were completely unmarked, and others had "marks" (a partially dislodged chad for a Presidential candidate), as observed in the ballot review. Some had chads with one, two, or three, corners detached (a "hanging chad"). Some chads were indented but with no corner detached (a "dimple" or "pregnant chad"). Others were punched through so light would pass, but the chad was not detached (a "pinprick"). To complicate matters, some ballots contained chads which were punched in locations which did not correspond to any Presidential candidate. Finally, some voters used a pen or pencil to indicate their votes, circling or underlying the appropriate number, or writing a candidate's name by hand.

Part of the political debate centered on the fundamental causes of the undervote marks. Predictably, the two political sides took opposing views. The Democrats argued that marked ballots were due to machine error and should be interpreted as legally valid votes. The Republicans presented alternative causes for partially dislodged chads, including mishandling, and the possibility that the marks were made by voters who contemplated making a vote, but decided to abstain from voting. The two political camps also had distinct positions on the bias of a hand recount. Most notably, the Republicans argued that a hand recount would be highly biased and prone to error.

Such claims have statistical implications. If undervote marks are caused by machine error, then the distribution of these marks across candidates (in an individual voting precinct) should follow the same probability distribution as machine-readable ballots. If the marks are due to mishandling, then the marks should be randomly distributed, independent of the machine-readable ballot distribution. If voter error (including aborted votes) is the cause, then the distribution of marks will be correlated with other factors such as precinct demographics, since voter error is likely to be systematic. If a hand recount is unbiased, then the hand recount will accurately reflect the mark distribution (regardless of the cause of the marks). Finally, if the hand recount is politically biased, then the hand recount distribution will be shifted to favor the preferred party.

This discussion leads naturally to a statistical investigation of the association between the machine-counted vote distribution and the hand-counted vote distribution. The tighter the link between these two distributions, the more solid the case that the hand-counted votes represented valid voter intent. Indeed, these two distributions coincide if the cause of undervote marks is random machine error and hand recounts are unbiased. If these distributions are indeed identical, then it is hard to conceive of an alternative mechanism through which this could occur. Somewhat more generally, an implication termed the *unbiased hypothesis* is that the latent machine-vote probability is an unbiased predictor of the recount vote probability. The present investigation focuses on this hypothesis.

This article reports on a statistical investigation of the vote and recount patterns in Broward, Miami-Dade, and Palm Beach Counties. The focus is on undervotes and the two-party Presidential vote, ignoring third-party votes and overvotes. The analysis is based on a collection of several data sources at the precinct level in the three counties: the official tally by the vote machines, the hand recounts conducted by the canvassing boards, the review of the undervote ballots by the Miami-Herald, and the precinct-level voter registration demographic aggregates.

Bruce E. Hansen is Professor, Department of Economics, University of Wisconsin, Madison, WI 53706 (E-mail: *bhansen@ssc.wisc.edu*). This research was supported by a grant from the National Science Foundation. The author thanks a referee, the editor, and Walter Mebane for careful and detailed comments on earlier drafts; Jasjeet Sekhon for providing demographic data for Broward county; and the Broward County Elections Board for their kind cooperation in obtaining the recount vote records.

Journal of the American Statistical Association
June 2003, Vol. 98, No. 462, Applications and Case Studies
DOI 10.1198/016214503000062

To investigate the association between the machine-readable votes and the hand-recount votes, a multivariate generalization of the beta-logit model is proposed. The precinct-level vote probabilities are modeled as unobservable latent variables, drawn from a beta distribution, with the mean a logit function of precinct-level observables. This latent variable model directly allows for overdispersion and clustering, in contrast to binomial-logit models, which require an ad hoc adjustment for overdispersion. The multivariate extension of the beta-logit model presented here is new and allows a study of the relationship between two unobserved probabilities. A very useful finding is that the joint likelihood function is well approximated by the sum of two likelihood components, and thus the maximum likelihood estimator (MLE) can be approximated by a two-step estimator in which each step is a univariate beta-logit estimator. Because the latter is computationally simple, estimation of the joint model is computationally easy.

The univariate beta-logit model was proposed by Heckman and Willis (1977) to study female labor force participation. Other applications include strike duration (Kennan 1985), covariate measurement error (Prentice 1986), and ecological inference (King, Rosen, and Tanner 1999).

Our major question of interest is the determination of the recount vote-from-undervote probabilities. The empirical findings from the multivariate beta-logit model are clear, with strong support for the unbiased hypothesis and no statistical evidence for other claims. The estimated conditional expectation function for the recount vote probability is virtually identical to the machine vote probability, which is the precise implication of the unbiased hypothesis. There is no evidence of bias in the recount vote.

The article is organized as follows. Section 2 describes the univariate and bivariate beta-logit models. Section 3 describes the data and how they were collected. Section 4 describes the empirical model and reports the estimates. The Appendix presents explicit expressions for the first and second derivatives of the log-likelihood and constructs the approximate MLE for the bivariate beta-logit model. The complete dataset and Gauss programs used to compute the estimates are available on the author's webpage (*www.ssc.wisc.edu/~bhansen*).

2. MODEL

2.1 Univarate Beta-Logit

There are n precincts each with n_i voters. For simplicity, suppose that each voter either takes action A or B. (These actions can include voting for a particular candidate or specific "errors," such as undervoting or overvoting.) Because within a precinct the voters are undifferentiated, each voter within a precinct can be viewed as independent and identically distributed, with precinct-specific probability p_i of taking action A. Equivalently, in precinct i the number of A actions y_i out of n_i voters is binomial with parameter p_i. Because there is geographic clustering of voters with similar political preferences, p_i is a latent random variable whose distribution may depend on precinct-specific covariates x_i (average precinct characteristics). A simple parametric distribution is the binomial-logit, which sets $p_i = \lambda(\beta' x_i)$, where $\lambda(s) = (1 + \exp(-s))^{-1}$ is the logit function. Unfortunately, this deterministic specification is an unrealistic feature, because it is unlikely that the covariates can account for all variation in p_i, a situation known as *overdispersion*, or *extrabinomial variation*. One solution is to treat the binomial likelihood as a quasi-likelihood and adjust the standard errors to account for overdispersion, as was done by, for example, Wand et al. (2001). This is convenient for the study of a single set of ballots, but inappropriate for a study of two or more sets of ballots, as is reported in this article.

Our preferred approach is to explicitly model p_i as random using the beta distribution,

$$beta(p \mid \mu, \theta) = \frac{\Gamma(\mu)}{\Gamma(\theta\mu)\Gamma((1-\theta)\mu)} p^{\theta\mu - 1}(1-p)^{(1-\theta)\mu - 1}.$$

The conditional density of p_i given x_i is $beta(p_i \mid \mu_i, \theta)$, where $\mu_i = \lambda(\beta' x_i)$. Because $E(p_i) = \lambda(\beta' x_i)$, the beta-logit generalizes the binomial-logit, allowing for variation in p_i beyond the mean $\lambda(\beta' x_i)$. The parameter θ indexes the dispersion of p_i. As $\theta \to \infty$, the distribution collapses to the binomial-logit.

Let $\psi = (\beta, \theta)$. The beta-logit log-likelihood is

$$l_n(\psi) = \sum_{i=1}^{n} \log f(y_i \mid n_i, x_i, \psi),$$

$$f(y_i \mid n_i, x_i, \psi) = \binom{n_i}{y_i} \frac{\Gamma(y_i + \mu_i\theta)\Gamma(n_i - y_i + (1-\mu_i)\theta)\Gamma(\theta)}{\Gamma(n_i + \theta)\Gamma(\mu_i\theta)\Gamma((1-\mu_i)\theta)}. \quad (1)$$

The MLE, $\hat{\psi} = (\hat{\beta}, \hat{\theta})$, is found by numerical maximization of the log-likelihood function $l_n(\psi)$. Analytic first- and second-derivatives are given in the Appendix. The log-likelihood is quite well behaved and can be maximized using the Newton method in just a few seconds on a personal computer.

The conditional distribution of the latent p_i given y_i is beta and has the conditional mean $E(p_i \mid y_i) \equiv p_i^* = (y_i + \mu_i\theta)/(n_i + \theta)$ and variance $\operatorname{var}(p_i \mid y_i) = p_i^*(1 - p_i^*)/(1 + n_i + \theta)$. The conditional mean p_i^* is a weighted average of the precinct mean y_i/n_i and the unconditional mean μ_i, with the weights depending on θ and n_i. When θ is large (low overdispersion) or n_i is small, then $p_i^* \approx \mu_i$, and the realized data, y_i/n_i, has little impact on the precinct estimate. On the other hand, when θ is small (large dispersion) or n_i is large, then $p_i^* \approx y_i/n_i$, and the unconditional mean, μ_i, has little impact. It is also useful to observe that when $n_i + \theta$ is large, $\operatorname{var}(p_i \mid y_i)$ shrinks to 0, and the conditional distribution of p_i given y_i collapses to a point mass at p_i^*.

2.2 Bivariate Beta-Logit

Consider a model of two jointly related counts (e.g., the machine-countable and hand-countable votes). Formally, in each precinct there are two sets of voters, which may be intersecting or disjoint. The first group has n_{1i} voters, of whom y_{1i} take action A_1 and $n_{1i} - y_{1i}$ take action B_1. The second group has n_{2i} voters, of whom y_{2i} take action A_2 and $n_{2i} - y_{2i}$ take action B_2. Assume that there are jointly dependent latent probabilities p_{1i} and p_{2i} such that conditional on (p_{1i}, p_{2i}), y_{1i} and y_{2i} are independent binomial random variables with binomial probabilities p_{1i} and p_{2i}. This means that all stochastic dependence between y_{1i} and y_{2i} is due to the joint dependence of p_{1i} and p_{2i}.

Assume that the marginal distribution of p_{1i} is beta, and that the conditional distribution of p_{2i} given p_{1i} is beta with p_{1i} only entering the mean function. Thus the pair (p_{1i}, p_{2i}) has the joint density $beta(p_{1i} \mid \mu_{1i}, \theta_1) beta(p_{2i} \mid \mu_{2i}(p_{1i}), \theta_2)$, where $\mu_{1i} = \lambda(\beta_1' x_{1i})$, $\mu_{2i}(p_{1i}) = \lambda(\beta_2' x_{2i}^*)$, and $x_{2i}^* = (x_{2i}' \quad t(p_{1i}))'$ is an augmented covariate vector, with $t(p_{1i})$ a transformation of p_{1i}. It will be useful to observe that the conditional mean functions of these variables are

$$E(p_{1i}) = \lambda(\beta_1' x_{1i}),$$

$$E(p_{2i} \mid p_{1i}) = \lambda(\beta_2' x_{2i}^*) = \lambda(\beta_{21}' x_{2i} + \beta_{22} t(p_{1i})),$$

where $\beta_2' = (\beta_{21}' \beta_{22})$ has been partitioned. Let $\psi_1 = (\beta_1, \theta_1)$ and $\psi_2 = (\beta_2, \theta_2)$.

One useful choice for $t(\cdot)$ is $t(p) = \log(p)$, for then the parameter β_{22} is the elasticity (i.e., elasticity of p with respect to x is the percentage change in p due to a 1% change in x) of p_{2i} with respect to p_{1i}. Another important choice is the inverse logit $t(p) = \lambda^{-1}(p) = \log(p/(1-p))$, for then the important restriction $E(p_{2i} \mid p_{1i}) = p_{1i}$ holds under the restrictions $\beta_{22} = 1$ and $\beta_{21} = 0$.

An approximate MLE for the bivariate beta-logit model is given in the Appendix.

3. DATA

Precinct-level data are available for the three punch-card counties in southern Florida that initiated hand recounts of the Presidential ballots: Broward, Miami-Dade, and Palm Beach. This analysis primarily considers the regular voting precincts, and treats absentee ballots separately.

In each county, there was an initial machine count of the ballots and a second machine count. A ballot is termed *machine-readable* if the second machine count determined there was a Presidential vote. For simplicity, we include votes only for Gore and Bush, excluding third-party votes. We use the second machine count as the recount changes were measured relative to it. The first section of Table 1 summaries the total number of ballots, the machine-count totals each for Gore and Bush, and the number of Undervotes. Also reported are the total net additional votes for Gore and Bush obtained from the hand recounts.

All three counties initiated a hand recount of the ballots. Broward and Palm Beach completed their recounts, whereas Miami-Dade recounted only 112 of 613 voting precincts. (The County board was recounting the precincts roughly in order by precinct number. This is not a random sample of precincts, because these precincts tended to be concentrated in older urban neighborhoods. This should not induce a selection bias, however, as it is not selecting on outcomes. To check robustness to this claim, all the analysis was repeated with the Miami-Dade precincts excluded, and none of the results changed meaningfully.) The data for Broward County were provided by the Broward County Elections Board in a format that needs to be explained in detail. Their only official precinct-level record is a sheet-by-sheet hand account of the recount process. These sheets contain marks indicating the changes in the vote count for each candidate. For many precincts, there are multiple marks indicating vote changes. Treating all of these markings as valid new votes, the markings indicate a total of 1,124 new votes for Gore and 582 new votes for Bush. (This total includes the absentee precincts and does not truncate negative precinct new vote totals.) This is close to (albeit slightly different from) the official gain of 1,142 for Gore and 579 for Bush. Despite this discrepancy, these data are used for the present analysis.

Table 1. Data Summary (Voting Precincts: Excludes Absentee Ballots)

	Broward	*Miami-Dade*	*Palm Beach*
Votes			
Voting precincts	609	613	516
Ballots	537,680	610,777	415,367
Machine Gore	359,255	311,943	244,892
Machine Bush	156,876	265,260	132,702
Undervotes	5,308	9,302	9,104
Recount Gore	877	305	507
Recount Bush	394	133	321
Undervote review			
No mark	2846	4633	2033
Punch in wrong hole	157	1948	123
Clean-punch	984	57	60
Dimple	1358	1564	5941
Pinprick	126	527	272
Hanging chad	206	52	71
Voter registration records			
Total	887,161	896,912	658,837
Democrat	456,617	396,518	296,122
Other party	163,925	161,520	131,089
Black	125,151	176,806	125,151
Hispanic	58,973	398,573	20,941
Female	478,693	498,335	358,714
Age 17–20	21,680	37,982	12,341
Age 21–29	96,066	121,855	58,687
Age 56–64	94,412	113,748	76,032
Age 65-up	214,093	212,590	228,073

The information reflects the changes in each candidate's vote totals due to the recount. In most precincts, the votes were unchanged or increased, but in some precincts there were slight vote decreases (nearly all just a single vote) for individual candidates. Because this is inconsistent with the present modeling strategy, the recount votes for each candidate are truncated at 0. A more sophisticated analysis would allow for vote decreases, but given the small number of these occurrences this does not seem important.

The *Miami-Herald* newspaper reviewed the undervote ballots, as described by Merzer (2001). For each ballot, they recorded whether the ballot contained no visible marks, whether their was a mark in a position that did not correspond to a valid vote, or whether there was a visible mark corresponding to a particular candidate. If there was such a mark, then they recorded the type of mark. Totals for the voting precincts are reported in the second section of Table 1.

Finally, precinct-level voter registration records were also used. The counties maintain records on voter registration applications and thus have information on voters' political affiliation, race, gender, and date of birth. For each precinct in all counties, available data include we have the total number of registered voters, political affiliation (one of three categories: Republican, Democrat, and "other," which includes third-parties and Independents), race and ethnicity (by three categories: "Black, not of Hispanic origin", "Hispanic", and "other," which includes non-Hispanic whites, Asians, and American Indians), and breakdowns by gender and the age categories 17–20, 21–29, 30–55, 56–64, and 65-up, as provided by the counties. Totals by county are reported in the third section of Table 1.

4. EMPIRICAL MODEL

4.1 Latent Probabilities

A joint distribution of four latent probabilities for the regular voting precincts is estimated. These probabilities are as follows:

- p, the probability that a machine-readable vote goes to Gore
- p_a, the probability that a ballot is unmarked (abstentions)
- p_m, the probability that a ballot is marked
- q, the probability that a recount vote goes to Gore.

Here a *recount vote* is an undervote that was determined to have a legal vote for a Presidential candidate by the county canvassing board, and an *undervote* is a ballot that the machine tally determined had no vote for a Presidential candidate (this excludes overvotes). A *marked* ballot is an undervote that had a mark (a partially dislodged chad) for a Presidential candidate, as observed in the ballot review. An *unmarked* ballot is an undervote that had no marks for a Presidential candidate, as observed in the ballot review.

It is important to note that these four probabilities are freely varying (i.e., there are no adding-up restrictions) with the exception that $p_a + p_m$ is the probability of an undervote, and thus must be less than one. Interest focuses mainly on $E(q \mid p)$.

Initially, consider an idealized environment in which the recount process was unbiased and nonpartisan, so that the votes counted in the hand recount are honest reflections of the ballot marks (regardless of the source of the marks). Let q^0 denote Gore's percentage of the recount ballots in this idealized situation, and consider the function $E(q^0 \mid p)$. Consider the implications for this function of the three potential causes for ballot marks. First, if machine error is the sole cause for marks, then marked ballots are random draws from the population of all intended votes in a precinct and hence $E(q^0 \mid p) = p$. Second, if voter error is the cause for marks, then q^0 will be determined by the political preferences of the subpopulation that is prone to make the ballot punching error (e.g., elderly, uneducated, and first-time voters). In this case $E(q^0 \mid p)$ can be a more complicated function, and in particular can vary with precinct-specific variables. Third, if mishandling is the sole cause for marks, then marked ballots are noise. Hence q^0 will be independent of precinct-specific factors, implying that $E(q^0 \mid p) = c$, a constant. If all three factors are relevant, then the function $E(q^0 \mid p)$ will be a weighted function of the three cases presented here. The weights may vary between precincts, so the weighted average $E(q^0 \mid p)$ may vary with these factor proportions.

Additionally, the recount process may have been subject to bias, taking the form of differing standards applied to ballots. Bias will have the effect of altering the probabilities that a marked ballot is counted as a valid vote, raising or lowering $E(q \mid p)$ relative to $E(q^0 \mid p)$. If the source of marks is pure machine error and the recount is unbiased, then $E(q \mid p) = E(q^0 \mid p) = p$. This is the leading scenario, and it is termed the *unbiased hypothesis*. Other causes for marks and/or recount bias will distort $E(q \mid p)$ from p; in particular, mishandling flattens the function, voter error introduces precinct-specific factors, and recount bias can shift the function up (pro-Gore) or down (pro-Bush).

4.2 Multivariate Beta-Logit

The multivariate beta-logit model as described in Section 2 is used with the ordering (p, p_a, p_m, q). (Thus p has a univariate beta-logit distribution, and p_a is conditionally beta-logit with a conditional mean depending on p, and soon.) This specification allows for full joint dependence between the latent probabilities. As a robustness check, if the ordering (p, p_m, p_a, q) is used instead, then the results do not change meaningfully.

In each equation the conditional mean includes dummy variables for each of the three counties, plus eight voter registration variables (the percentage of registered voters who are Democrats, other party, Black, Hispanic, Female, age 17–20, age 21–29, age 56–64, and age 65-up). The voter registration covariates are expressed as ratios to their sample averages. In all cases, the latent probabilities p, p_a, and p_m enter the conditional mean equations using a logarithmic transformation, with the important exception of the q equation, in which p enters using an inverse logit link as discussed in Section 2.2 and has a county-specific slope.

4.3 Machine Vote, Unmarked, and Marked Ballots

The first column of Table 2 contains the estimated equation for p—the probability that a machine-readable vote goes to

Table 2. Ballots

	p	p_a	p_m	p_m–Dimples	p_m–Pinpricks	p_m–Chads
Palm Beach	$-3.83_{(.27)}$	$-5.01_{(1.01)}$	$-2.60_{(1.32)}$	$-2.79_{(1.42)}$	$-3.62_{(2.50)}$	$-12.86_{(3.89)}$
Miami-Dade	$-3.89_{(.27)}$	$-4.65_{(1.01)}$	$-3.72_{(1.31)}$	$-4.06_{(1.41)}$	$-3.30_{(2.49)}$	$-13.79_{(3.85)}$
Broward	$-3.83_{(.27)}$	$-4.92_{(1.01)}$	$-3.62_{(1.31)}$	$-3.90_{(1.40)}$	$-4.34_{(2.49)}$	$-12.17_{(3.85)}$
Democrat (%)	$2.34_{(.04)}$	$.45_{(.27)}$	$-.85_{(.35)}$	$-.73_{(.38)}$	$-1.63_{(.66)}$	$-.70_{(1.17)}$
Other party (%)	$.87_{(.04)}$	$.17_{(.17)}$	$-.28_{(.21)}$	$-.16_{(.23)}$	$-.79_{(.37)}$	$-.42_{(.68)}$
Black (%)	$.14_{(.01)}$	$.14_{(.02)}$	$.02_{(.03)}$	$.01_{(.03)}$	$.09_{(.07)}$	$-.13_{(.10)}$
Hispanic (%)	$-.11_{(.01)}$	$.10_{(.04)}$	$.21_{(.05)}$	$.14_{(.05)}$	$.33_{(.08)}$	$.44_{(.19)}$
Female (%)	$.19_{(.13)}$	$.19_{(.33)}$	$.91_{(.42)}$	$.98_{(.46)}$	$.21_{(.79)}$	$2.16_{(1.36)}$
Age 17–20 (%)	$.02_{(.02)}$	$-.16_{(.05)}$	$-.04_{(.07)}$	$-.03_{(.07)}$	$-.20_{(.13)}$	$.16_{(.22)}$
Age 21–29 (%)	$.54_{(.10)}$	$-.97_{(.33)}$	$-.19_{(.42)}$	$-.30_{(.45)}$	$-.54_{(.85)}$	$3.37_{(1.31)}$
Age 56–64 (%)	$.11_{(.03)}$	$-.29_{(.10)}$	$.28_{(.13)}$	$.25_{(.14)}$	$.20_{(.24)}$	$1.01_{(.38)}$
Age 65-up (%)	$.39_{(.05)}$	$-.09_{(.17)}$	$-.07_{(.21)}$	$-.13_{(.22)}$	$-.07_{(.42)}$	$1.25_{(.67)}$
log(p)		$-.13_{(.23)}$	$1.12_{(.32)}$	$.97_{(.36)}$	$1.68_{(.53)}$	$1.98_{(1.20)}$
log(p_a)			$.23_{(.07)}$	$.24_{(.08)}$	$.11_{(.13)}$	$.34_{(.15)}$
θ	$141_{(8)}$	$664_{(56)}$	$193_{(19)}$	$187_{(18)}$	$1,165_{(233)}$	$1,841_{(353)}$

Note: Generalized method-of-moments standard errors are in parentheses.

Gore. All of the included variables are statistically significant and fairly precisely estimated. The most important factor, as expected, is the percentage of registered Democrats. Because this variable (as well as the other covariates) are expressed as ratios to their sample averages, the elasticity of the probability of a vote for Gore is approximately the coefficient multiplied by $(1 - \overline{p}) \simeq .38$. So the estimated elasticity with respect to the percentage of registered Democrats is .89. The estimated scale parameter θ corresponds to a standard deviation in p of 4 percentage points. This means that the logit conditional mean cannot explain all of the variation in the Gore vote percentage. This variance includes all heterogeneity that is not explained by the limited set of covariates.

The second column of Table 2 contains the estimated equation for p_a, the probability that a ballot is unmarked. This is a reasonable proxy for abstentions—voters who intentionally did not cast a vote for President. It is useful to note that because the probability of an unmarked ballot is very small, the coefficients on the dummy variables are direct measures of the impact on the mean probability, and the coefficients on the other variables are elasticities. Abstentions are negatively associated with voters age 21–29 and positively associated with Democrats, Blacks, and Hispanics (although the last two effects are small in magnitude).

The third column of Table 2 contains the estimated equation for p_m, the probability that a ballot is marked. Marks are positively associated with females, Hispanics, and age 56–64. Marks are also positively associated with the abstention rate p_a and Gore's vote p. This means that Bush votes and marks were substitutes.

These results complement the regression analysis of Herron and Sekhon (2003), who showed that undervote rates in Broward and Miami-Dade Counties were positively associated with Hispanics and Blacks. The notable differences are that the present study (unlike Herron and Sekhon) has disaggregate undervotes into abstentions and marks, and conditions on political affiliation and the Gore vote probability.

As a robustness check, the model for p_m was estimated after marks were disaggregated into dimples, pinpricks, and hanging chads. The estimates are reported in the fourth, fifth, and sixth columns of Table 2. The results look qualitatively similar to the equation for p_m.

4.4 Vote Recount

Table 3 contains an estimated estimation for q. An equation was also estimated with the demographic variables included,

Table 3. Gore Versus Bush in Canvassing Board Recount

Palm Beach	$-.75_{(.56)}$
Miami-Dade	$-.56_{(.52)}$
Broward	$-.39_{(.57)}$
$\log(p_a)$	$.01_{(.09)}$
$\log(p_m)$	$-.09_{(.05)}$
$\lambda^{-1}(p) \cdot PBC$	$1.22_{(.15)}$
$\lambda^{-1}(p) \cdot MDC$	$.87_{(.14)}$
$\lambda^{-1}(p) \cdot BC$	$.81_{(.11)}$
θ	$34_{(22)}$

Note: Generalized method-of-moments standard errors are in parentheses.

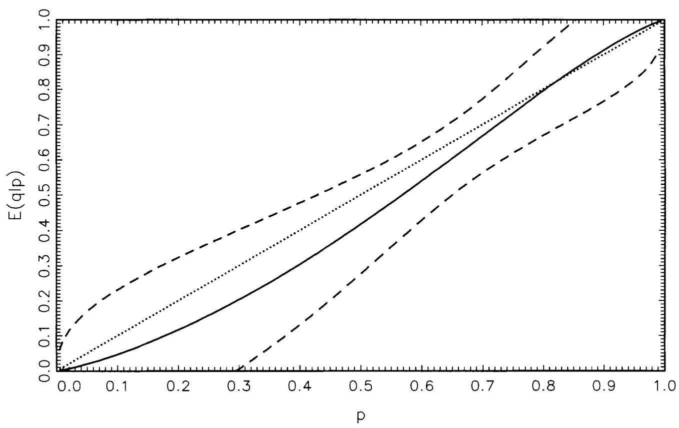

Figure 1. Gore Percentage in Recount, Palm Beach County.

but none of the demographic coefficients were significant, and so this equation is not presented here to conserve space. None of the individual coefficient estimates is statistically different than 0 other than the coefficients on $\lambda^{-1}(p)$, which are close to unity. The unbiased hypothesis cannot be rejected and in fact the point estimates are extremely close to this case. There is no strong evidence of any systematic variation in recount vote beyond that predicted by the unbiased hypothesis. In particular, observe that the coefficients on the county dummy indicators are insignificantly different from 0. This means that there is no statistical evidence of bias specific to a particular county canvassing board.

The unbiased hypothesis states that the function $E(q_i \mid p_i = p) = p$ should be the 45-degree line, so this can be qualitatively assessed by plotting the estimated function

$$\hat{E}(q_i \mid p_i = p) = \lambda\Big(\hat{\beta}_{21}'\overline{x}_2 + \hat{\beta}_{22}\lambda^{-1}(p)\Big).$$

This is done for the three counties in Figures 1–3. The machine vote probability p is on the x-axis, and the recount vote probability q is on the y-axis. The solid line represents the estimated relationship, the dotted line, the 45-degree line (the prediction of the unbiased hypothesis), and the dashes, pointwise 90% confidence intervals. It is fairly clear that the estimated relationships are very close to the unbiased hypothesis. The estimates are not consistent with a substantial amount of ballot mishandling (which would have flattened the functions) or bias (which would have shrunk the functions toward the favored party).

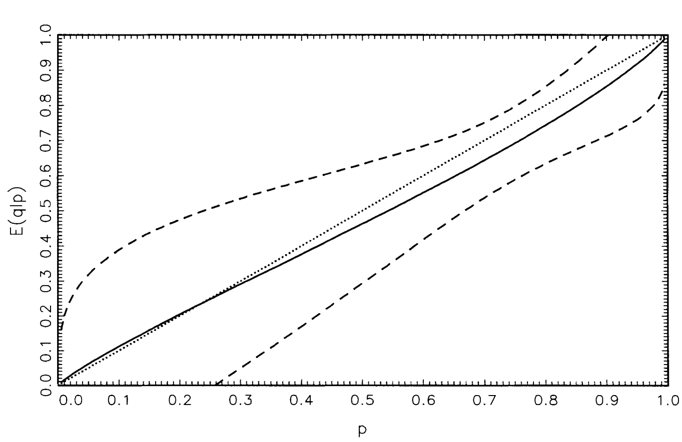

Figure 2. Gore Percentage in Recount, Miami-Dade County.

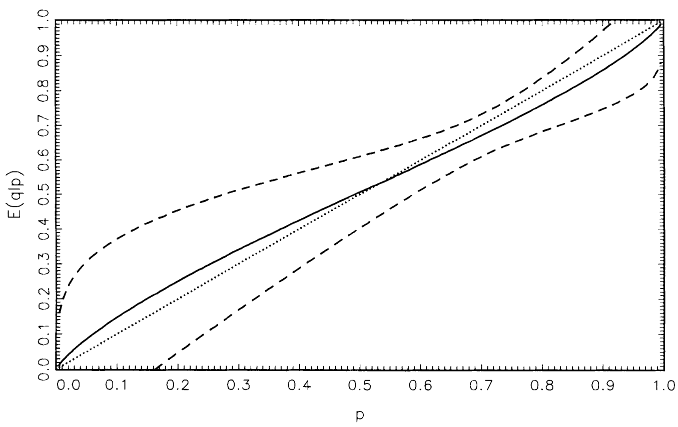

Figure 3. Gore Percentage in Recount, Broward County.

The only evidence pointing to the presence of voter error is the coefficient on $\log(p_m)$. In the constrained equation, this is borderline statistically significant. However, the point estimate of $-.09$ is very small, indicating an elasticity of the Gore recount percentage with respect to the marked vote percentage of $-.034$. That is, the point estimate implies that if the percentage of marked votes doubled, then Gore's percentage in the recount votes would decline by 3.4% (e.g., from 50% to 48.3%). Thus even if the effect is valid, it is very small. On balance, it can be concluded that the evidence points to the inconclusive possibility of a slight Bush leaning among the undervoters.

The statistical analysis described so far has excluded the absentee precincts, because they have several fundamental differences with regular voting precincts, including that there is no demographic measurement for absentee precincts. A robustness check involved separately examining the recount vote in the absentee ballots. Absentee ballots were recounted in Broward and Palm Beach Counties only, for a total of 282 precincts. A bivariate beta-logit model was estimated for the probabilities (p, q) with no covariates; Figure 4 plots the estimated relationship $E(q \mid p)$. The point estimate lies nearly on the 45 degree line, consistent with the unbiased hypothesis.

To summarize, the statistical evidence points very strongly to the unbiased hypothesis: the machine-vote percentage is an unbiased predictor of the hand-count undervote percentage. There is no evidence of any bias in the hand recount, and only slight evidence of a possible Bush preference tilt among the undervoters.

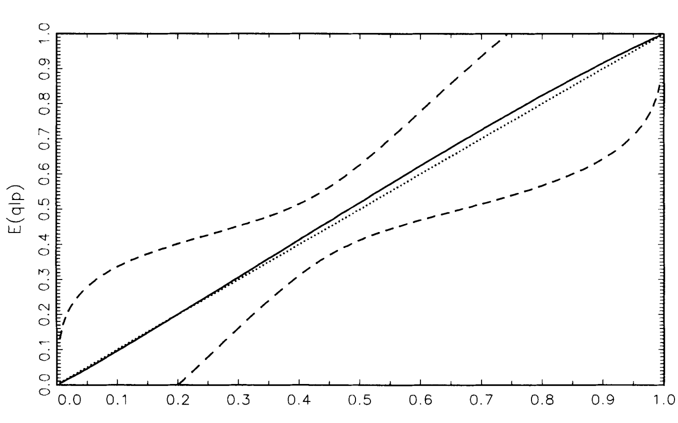

Figure 4. Gore Percentage in Recount, Absentee Precincts.

APPENDIX: BETA-LOGIT LIKELIHOOD

A.1 Likelihood Derivatives

This section presents analytic formulas for the derivatives of the univariate beta-logit model (1). Define $G(s) = \frac{d}{ds}\log(\Gamma(s))$, $H(s) = \frac{d^2}{ds^2}\log(\Gamma(s))$, and

$$\begin{aligned}
\lambda_i &= \lambda(x_i'\beta),\\
\Delta G_{1i} &= G(y_i + \lambda_i\theta) - G(\lambda_i\theta),\\
\Delta G_{2i} &= G(n_i - y_i + (1-\lambda_i)\theta) - G((1-\lambda_i)\theta),\\
\Delta H_{1i} &= H(y_i + \lambda_i\theta) - H(\lambda_i\theta),
\end{aligned}$$

and

$$\Delta H_{2i} = H(n_i - y_i + (1-\lambda_i)\theta) - H((1-\lambda_i)\theta).$$

The first and second derivatives of the beta-logit log-likelihood are

$$\frac{\partial}{\partial\beta} l_n(n_i, x_i) = \theta\sum_{i=1}^{n} x_i\lambda_i(1-\lambda_i)(\Delta G_{1i} - \Delta G_{2i}),$$

$$\frac{\partial}{\partial\theta} l_n(n_i, x_i) = \sum_{i=1}^{n}(\lambda_i\Delta G_{1i} + (1-\lambda_i)\Delta G_{2i} - G(n_i+\theta) + G(\theta)),$$

and

$$\begin{aligned}
\frac{\partial^2}{\partial\beta^2} l_n(n_i, x_i\theta) = \sum_{i=1}^{n} x_i x_i'\big[&\theta^2\lambda_i^2(1-\lambda_i)^2(\Delta H_{1i} + \Delta H_{2i})\\
&+ \theta\lambda_i(1-\lambda_i)(1-2\lambda_i)(\Delta G_{1i} - \Delta G_{2i})\big],
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2}{\partial\beta\partial\theta} l_n(n_i, x_i) = \sum_{i=1}^{n} x_i\lambda_i(1-\lambda_i)[&\theta(\lambda_i\Delta H_{1i} - (1-\lambda_i)\Delta H_{2i})\\
&+ \Delta G_{1i} - \Delta G_{2i}],
\end{aligned}$$

and

$$\frac{\partial^2}{\partial\theta^2} l_n(n_i, x_i) = \sum_{i=1}^{n}\big(\lambda_i^2\Delta H_{1i} + (1-\lambda_i)^2\Delta H_{2i} - H(n_i+\theta) + H(\theta)\big).$$

A.2 Approximate Maximum Likelihood Estimator

This section presents an approximate MLE for the bivariate beta-logit model of Section 2.2. Let $b(\cdot \mid \cdot)$ denote the binomial distribution and let

$$p_{1i}^*(\psi_1) = \frac{y_{1i} + \lambda(\beta_1' x_{1i})\theta_1}{n_{1i} + \theta_1}$$

denote the estimate of the conditional mean of p_{1i} given y_{1i}. By a first-order Taylor series expansion, the distribution of y_{2i} conditional on p_{1i} is approximately

$$\begin{aligned}
f_2(y_{2i} \mid p_{1i}) &= \int_0^1 b(y_{2i} \mid p_{2i})\,beta(p_{2i} \mid \mu_{2i}(p_{1i}), \theta_2)\,dp_{2i}\\
&\simeq f_2(y_{2i} \mid p_{1i}^*) + \frac{d}{dp_{1i}} f_2(y_{2i} \mid p_{1i}^*)(p_{1i} - p_{1i}^*),
\end{aligned}$$

where the dependence on parameters is suppressed. Letting $f_1(y_{1i})$ denote the marginal distribution of y_{1i} and $\pi_1(p_{1i} \mid y_{1i})$ denote the conditional density of p_{1i} given y_{1i}, the joint distribution of y_{1i} and y_{2i} is approximately

$$\begin{aligned}
f(y_{1i}, y_{2i}) = \int_0^1\int_0^1 &b(y_{1i} \mid p_{1i})b(y_{2i} \mid p_{2i})\,beta(p_{1i} \mid \mu_{1i}, \theta_1)\\
&\times beta(p_{2i} \mid \mu_{2i}(p_{1i}), \theta_2)\,dp_{1i}\,dp_{2i}
\end{aligned}$$

$$= f_1(y_{1i}) \int_0^1 \pi_1(p_{1i} \mid y_{1i}) f_2(y_{2i} \mid p_{1i}) dp_{1i}$$

$$\simeq f_1(y_{1i}) f_2(y_{2i} \mid p_{1i}^*) + f_1(y_{1i}) \frac{d}{dp_i} f_2(y_{2i} \mid p_{1i}^*)$$

$$\times \int_0^1 \pi_1(p_{1i} \mid y_{1i})(p_{1i} - p_{1i}^*) dp_{1i}$$

$$= f_1(y_{1i}) f_2(y_{2i} \mid p_{1i}^*), \tag{A.1}$$

the product of two beta-logit densities. Hence the joint log-likelihood is approximately

$$l_n(\psi_1, \psi_2) \simeq l_{1n}(\psi_1) + l_{2n}(\psi_1, \psi_2),$$

where

$$l_{1n}(\psi_1) = \sum_{i=1}^n \log f_1(y_{1i} \mid n_{1i}, x_{1i}, \psi_1),$$

$$l_{2n}(\psi_1, \psi_2) = \sum_{i=1}^n \log f_2(y_{2i} \mid n_{2i}, x_{2i}^*(\psi_1), \psi_2),$$

and

$$x_{2i}^*(\psi_1) = \begin{pmatrix} x_{2i} \\ t(p_{1i}^*(\psi_1)) \end{pmatrix}.$$

The functions f_1 and f_2 are the beta-logit density functions (1).

The approximation due to the Taylor expansion merits comment. As an alternative, numerical integration of the integral in (A.1) is possible. This is a difficult numerical integral, because the conditional density $\pi_1(p_{1i} \mid y_{1i})$ is close to a point mass at p_{1i}^* when $n_{1i} + \theta_1$ is large, which is true for most observations. Furthermore, the Taylor approximation is likely to be highly accurate, because the fact that $\pi_1(p_{1i} \mid y_{1i})$ is close to a point mass means that the approximation error will be quite small. In sum, the potential gains from full-fledged numerical integration are small and the costs are large.

Joint estimation over (ψ_1, ψ_2) is numerically difficult; thus the following two-step estimator is proposed. In the first step, $\hat{\psi}_1$ maximizes $l_{1n}(\psi_1)$. In the second step, $\hat{\psi}_2$ maximizes

$$l_{2n}(\hat{\psi}_1, \psi_2) = \sum_{i=1}^n \log f_2(y_{2i} \mid n_{2i}, \hat{x}_{2i}^*, \psi_2),$$

where $\hat{x}_{2i}^* = x_{2i}^*(\hat{\psi}_1) = (x_{2i}' \quad t(\hat{p}_{1i}^*))'$. This estimator is computationally easy to obtain, because each step is numerically identical to the estimation of a beta-logit model. In the second step, the distinctive feature is that one "regressor" is $\hat{p}_i^*$ obtained from the first-step estimates.

Because $l_{1n}(\psi_1)$ is a valid likelihood function, the first-step estimator $\hat{\psi}_1$ is consistent for ψ_1. Because ψ_2 only enters the joint likelihood through $l_{2n}(\psi_1, \psi_2)$ and $\hat{\psi}_1$ is consistent for ψ_1, it follows that $\hat{\psi}_2$ is consistent for ψ_2. However, $\hat{\psi} = (\hat{\psi}_1, \hat{\psi}_2)$ is not the MLE; thus calculation of the covariance matrix must take both steps into account. Based on the generalized method-of-moments principle (see Newey and McFadden 1994), an estimator of the asymptotic covariance matrix is $\hat{V} = (\widehat{M}'\widehat{\Omega}^{-1}\widehat{M})^{-1}$, where $\widehat{M} = \sum_{i=1}^n \frac{\partial}{\partial \psi'} m_i(\hat{\psi})$, $\widehat{\Omega} = \sum_{i=1}^n m_i(\hat{\psi}) m_i(\hat{\psi})'$, and

$$m_i(\psi) = \begin{pmatrix} \log f_1(y_{1i} \mid n_{1i}, x_{1i}, \psi_1) \\ \log f_2(y_{2i} \mid n_{2i}, x_{2i}^*(\psi_1), \psi_2) \end{pmatrix}.$$

Because the cross-derivative

$$\frac{\partial}{\partial \psi_1'} \log f_2(y_{2i} \mid n_{2i}, x_{2i}^*(\psi_1), \psi_2)$$

is complicated and not available in closed form, it must be evaluated numerically.

Using sequential conditioning, this model of two counts can be easily generalized to multiple counts.

[Received August 2002. Revised January 2003.]

REFERENCES

Heckman, J. J., and Willis, R. J. (1977), "A Beta-Logistic Model for the Analysis of Sequential Labor Force Participation by Married Women," *Journal of Political Economy*, 85, 27–58.

Herron, M. C., and Sekhon, J. S. (2003), "Overvoting and Representation: An Examination of Overvoted Presidential Ballots in Broward and Miami-Dade Counties," *Electoral Studies*, 22, 21–47.

Kennan, J. (1985), "The Duration of Contract Strikes in U.S. Manufacturing," *Journal of Econometrics*, 28, 5–28.

King, G., Rosen, O., and Tanner, M. A. (1999), "Binomial-Beta Hierarchical Models for Ecological Inference," *Sociological Methods and Research*, 28, 61–90.

Merzer, M. (2001), *The Miami Herald Report: Democracy Held Hostage*, New York: St. Martin's Press.

Newey, W. K., and McFadden, D. L. (1994), "Large Sample Estimation and Hypothesis Testing," *Handbook of Econometrics, Vol. IV*, eds. R. F. Engle and D. L. McFadden, New York: Elsevier.

Prentice, R. L. (1986), "Binary Regression Using an Extended Beta-Binomial Distribution, With Discussion of Correlation Induced by Covariate Measurement Error," *Journal of the American Statistical Association*, 81, 321–327.

Wand, J. H., Shotts, K. W., Sekhon, J. S., Mebane, W. R., Herron, M. C., and Brady, H. E. (2001), "The Butterfly Did It: The Aberrant Vote for Buchanan in Palm Beach County, Florida," *American Political Science Review*, 95, 793–810.