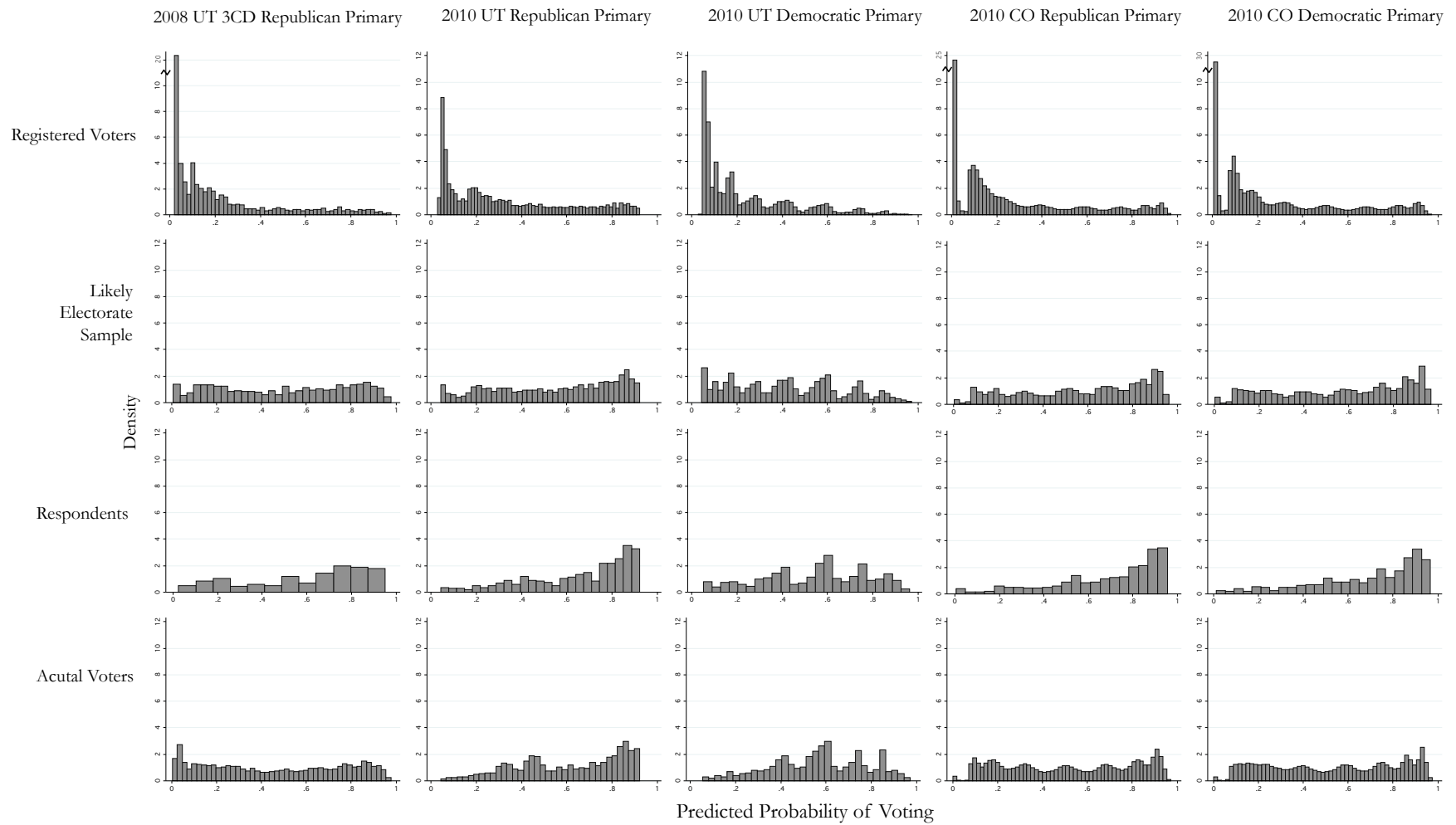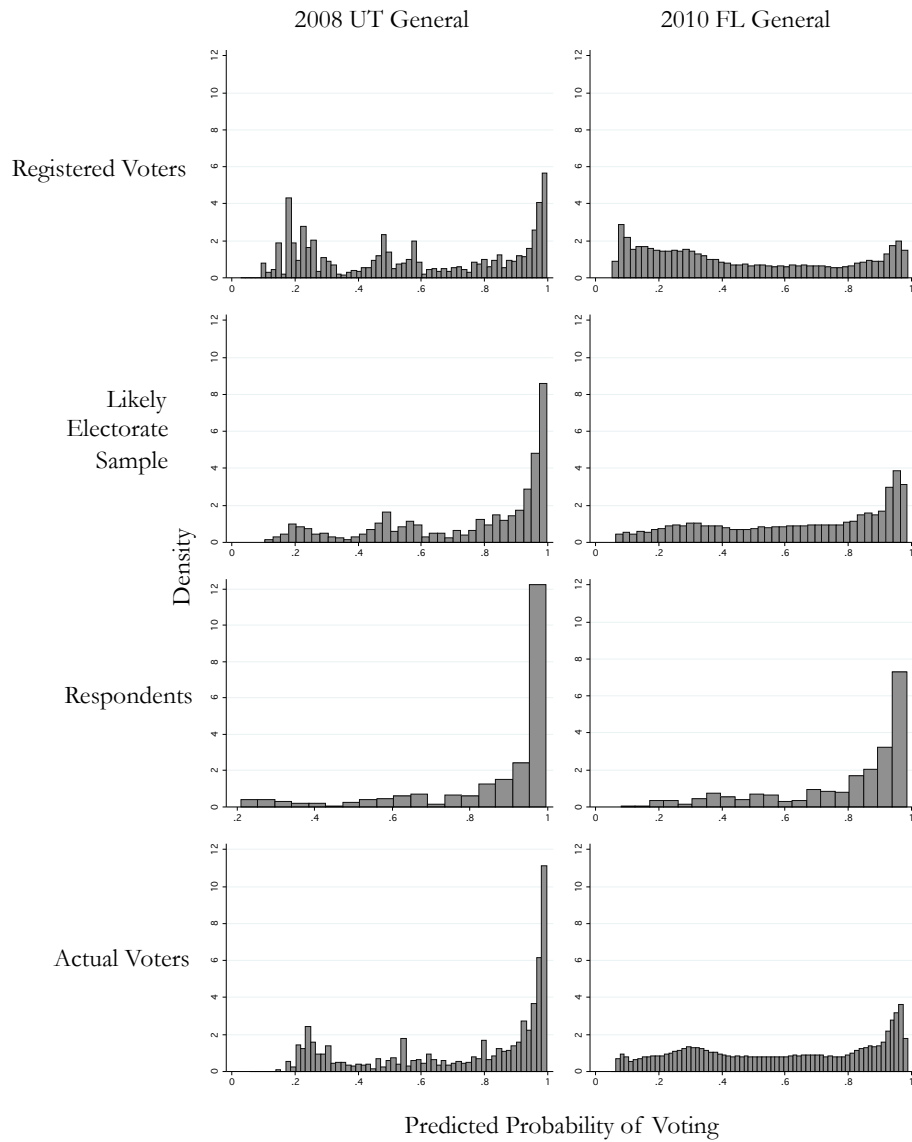## Supplementary Materials A: Figures for All 7 Surveys

Figure S1-A: Distribution of Predicted Probabilities of Voting in Primary Elections


(continued on next page)

Figure S1-B: Distribution of Predicted Probabilities for General Elections
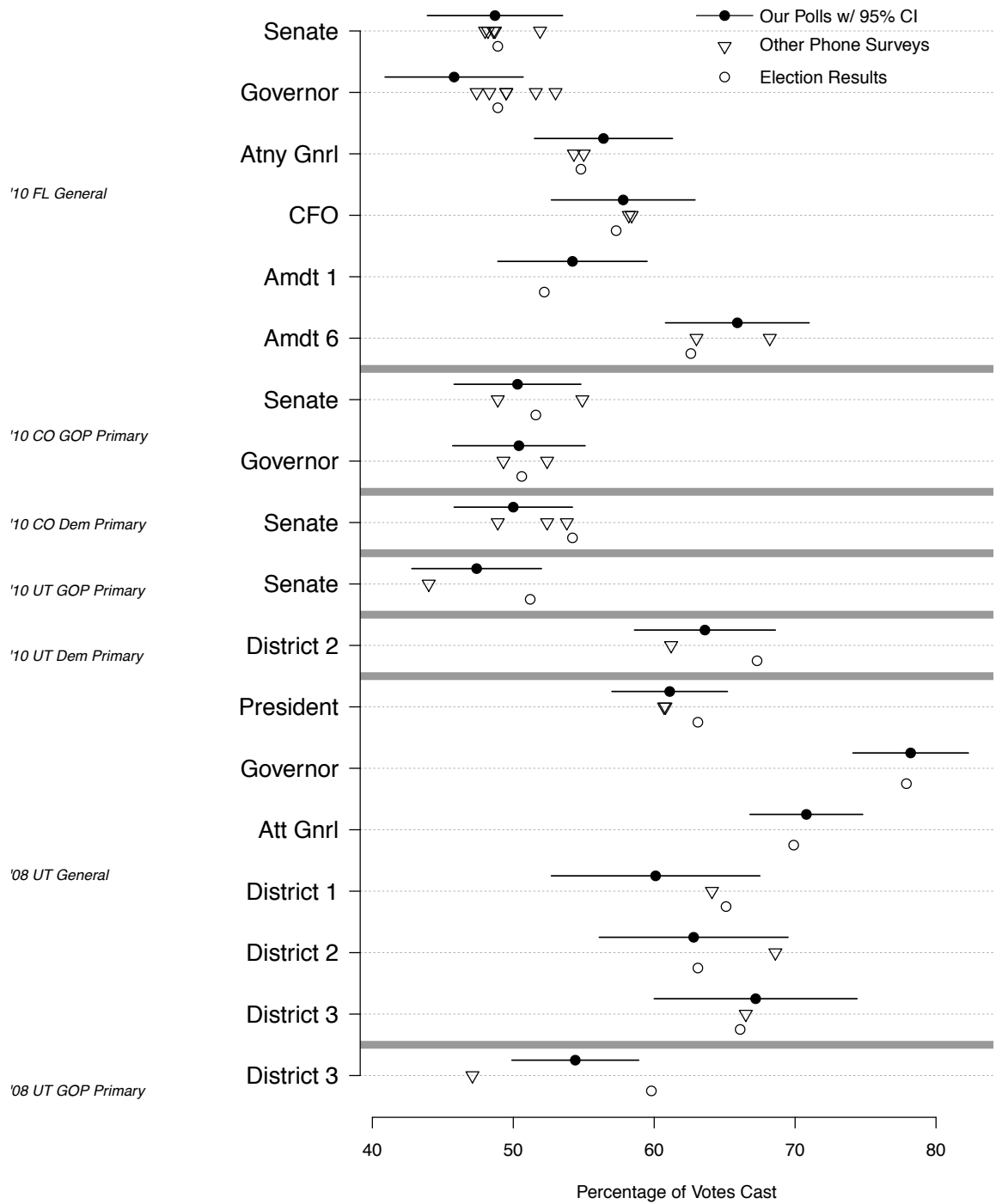

Predicted Probability of Voting

Each histogram shows the distribution across the predicted probabilities described in the paper. Each column displays a different survey. Each row displays the distribution of probabilities for different stages of the process. The first row displays the distribution all registered voters (limited to those eligible to vote in the primary election in Figure S1-A). The second row displays the distribution for the sample of the predicted likely electorate. The third row displays the probability distribution for those that responded to each survey. The fourth row displays the distribution of predicted probabilities for all people who actually voted in the election, based on public records of individual turnout.

Figure S2: ROC Curves for Each Model
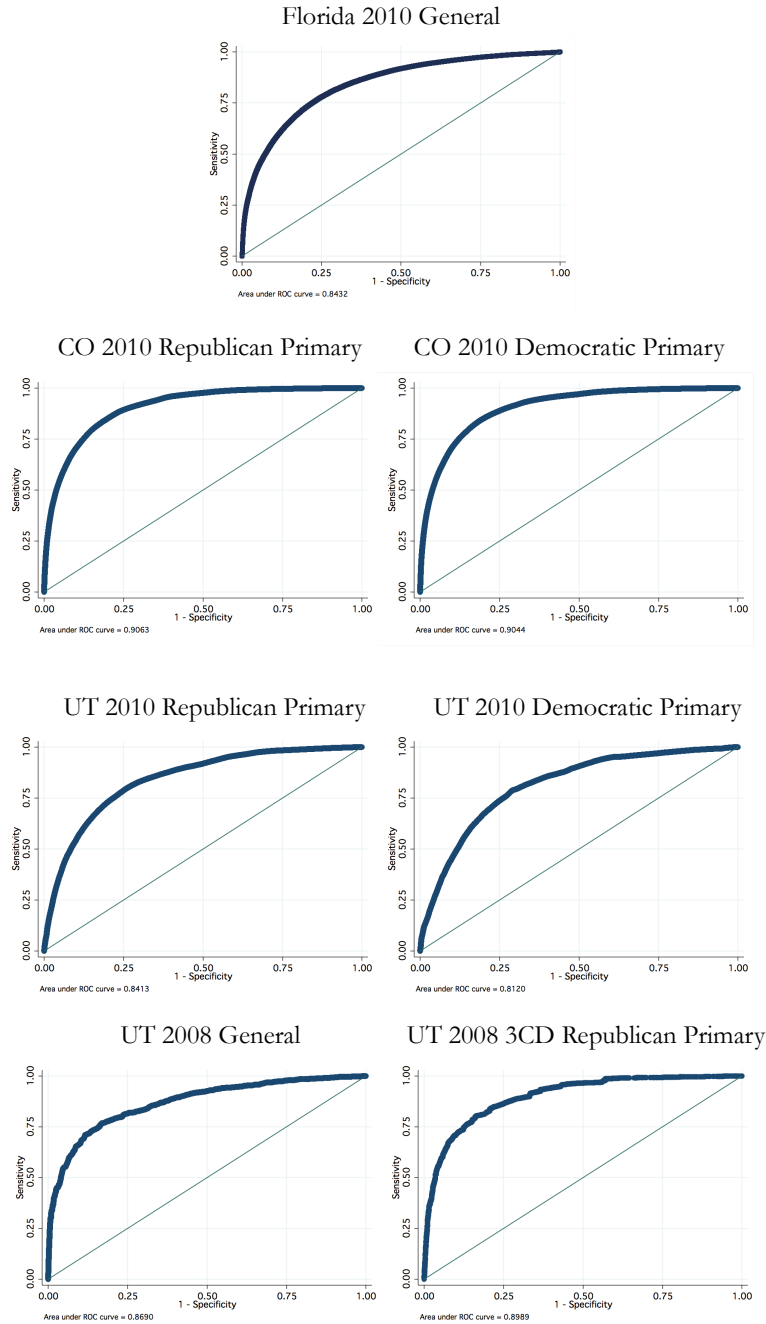
The area under each curve in Figure S2 is a measure of the accuracy of the model. A model that performed exactly randomly would follow the diagonal line from the bottom left to the top right corner (known as the "line of no discrimination") and have an area of 0.5. A model with a bias against predicting correct outcomes would have an area less than 0.5. A perfect model with no false negatives and no false positives would trace the y-axis and the x-axis and have an area of 1. Therefore, ROC curve areas close to 1 indicate a model that accurately predicts individual voter turnout for all potential predicted probabilities of voting.

Figure S3: Poll Forecasts Compared to Actual Election Results and Publicly Available Polls

Percentage of Votes Cast

Supplementary Materials B: Modeling Predicted Probability of Voting in Upcoming Election

Table S1 - Logit Models to Predict Probabilities of Voting

	2008		2010				
Survey:	UT GOP Primary†	UT General†	UT GOP Primary	UT Dem Primary	CO GOP Primary	CO Dem Primary	FL General
Dependent Variable	2006 GOP Primary	2004 General	2008 Pres. Primary	2008 Pres. Primary	2008 GOP Primary	2008 Dem Primary	2006 General
General Election Index	0.286*** (0.004)	1.292*** (0.003)	0.484*** (0.004)	0.547*** (0.006)	-0.120*** (0.015)	-0.127*** (0.015)	0.753*** (0.001)
Primary Election Index	0.608*** (0.006)	0.376*** (0.006)	0.322*** (0.010)	0.324*** (0.017)	1.356*** (0.011)	1.348*** (0.010)	1.019*** (0.001)
Off-Year Election Index			0.540*** (0.003)	0.704*** (0.007)	0.607*** (0.011)	0.633*** (0.011)	
Presidential Primary	0.634*** (0.004)	0.490*** (0.006)					
Republican	0.728*** (0.038)	0.457*** (0.005)	1.553*** (0.009)		1.920*** (0.034)		0.247*** (0.002)
Democrat		0.276*** (0.008)		1.111*** (0.019)		1.967*** (0.034)	-0.013*** (0.002)
Other Party		0.148*** (0.013)					
Years Since Original Reg			-0.026*** (0.001)	-0.043*** (0.003)	-0.014*** (0.001)	-0.016*** (0.001)	-0.019*** (0.001)
Years Since Last Reg Change					-0.022*** (0.006)	0.024*** (0.006)	
Age	0.013*** (0.000)	-0.004*** (0.000)	0.028*** (0.001)	0.027*** (0.002)	0.017*** (0.001)	0.032*** (0.001)	0.087*** (0.001)
Age Squared		0.000*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)	0.000*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)
Gender (Female)					0.087*** (0.006)	0.110*** (0.006)	-0.149*** (0.002)
Interactions							
Age * Years Since Original Registration			0.001*** (0.000)	0.001*** (0.000)			
General Index * Republican			-0.079*** (0.005)		0.047*** (0.016)		
Primary Index * Republican			-0.109*** (0.004)		-0.637*** (0.011)		
Off-Year Index * Republican					0.125*** (0.011)		
General Index * Democrat				0.092*** (0.011)		0.120*** (0.016)	
Primary Index * Democrat				0.050 (0.040)		-0.642*** (0.011)	
Off-Year Index * Democrat						0.152*** (0.011)	
Competative Dem District * Dem						0.902*** (0.008)	
Competative Dem District * Unaf						-.892*** (0.037)	
Competative GOP District * Rep					0.848*** (0.006)		
Competative GOP District * Unaf					-0.803*** (0.025)		
Constant	-4.582*** (0.037)	-0.931*** (0.005)	-3.572*** (0.022)	-3.229*** (0.045)	-5.478*** (0.043)	-6.025*** (0.044)	-4.602*** (0.007)
N	323,123	1,489,308	1,080,975	255,482	1,453,526	1,409,003	11,073,641

Significance levels : ∗ p<.10   ∗∗ p<.05   ∗∗∗ p<.01

Note: This table displays the logistic regression models used for each survey to generate predicted probability of voting in the upcoming election for each eligible registered voter. The predicted probability was then used to draw a probability proportionate to size [PPS] sample to reflect the likely electorate. Coefficients are displayed in logits. Standard errors are in parentheses. Statistical significance using two-tailed hypothesis testing: *<p<0.10, ** p<0.05, *** p<0.01. The previous similar election used to create each model is listed in Table 1of the paper. There are differences in the variables used in each model because of the differences in information available on the voter files in each state. Definitions of the variables used in each election are found below.

Description of Variables Used in Likely Voter Models

<u>**Survey: Florida 2010 General Election**</u>

Dependent Variable: The most recent mid-term general election, the 2006 Florida general election.

Independent Variables

- General Election Index: An index indicating the number of elections the individual voted in among the three most recent general elections prior to the 2006 general election: 2004, 2002, and 2000.

- Primary Election Index: An index indicating the number of elections each individual voted in among the three most recent primary elections prior to the 2006 general election: 2006, 2004, and 2002.

- Republican: A dummy variable taking a value of 1 if the individual is a registered Republican.

- Democrat: A dummy variable taking a value of 1 if the individual is a registered Democrat.

- Years Since Original Registration: The time measured in years between the date of the 2010 general election and the date the individual first registered to vote in Florida.

- Years Since Last Registration Change: The time measured in years between the date of the 2010 general election and the date of the last change in an individual's registration status. Changes could occur because the voter moved, changed party affiliation, or other reasons.

- Age: Measured in years from Election Day for the 2010 general election.

- Age Squared: Used to account for nonlinearities in the effect of age on voting probability.

- Gender: A dichotomous variable coded 1 for female.

<u>**Survey: Colorado 2010 Statewide Democratic Primary**</u>

Dependent Variable: The 2008 statewide primary election. See "Identification of Eligible Voters for Primary Elections" below.

Independent Variables

- General Election Index: An index indicating the number of elections the individual voted in using the three most recent general elections: 2008, 2006, and 2004.

- Primary Election Index: An index indicating the number of elections the individual voted in using the three most recent primary elections: 2006, 2004, and 2002.

- Municipal Election Index: An index indicating the number of elections that the individual voted in using the three most recent off-year elections: 2009, 2007, and 2005.

- Democrat: A dummy variable taking a value of 1 if the individual is a registered Democrat.

- Years Since Original Registration: The time measured in years between the date of the 2010 primary election and the date the individual first registered to vote in Florida.

- Years Since Last Registration Change: The time measured in years between the date of the 2010 primary election and the date of the last change in an individual's registration status. Changes could occur because the voter moved, changed party affiliation, or other reasons.

- Age: Measured in years from Election Day for the 2010 primary.

- Age Squared: Used to account for non-linearities in the effect of age on voting probability.

- Gender: A dichotomous variable coded 1 for female.

- General Election Index * Democrat: Used to account for a different effect among Democrats and Unaffiliated voters in general election voting. We interact the general election index with the Democrat dummy variable.

- Primary Election Index * Democrat: Used to account for a different effect among Democrats and Unaffiliated voters in primary election voting. We interact the primary election index with the Democrat dummy variable.

- Off-Year Election Index * Democrat: Used to account for a different effect among Democrats and Unaffiliated voters in off-year election voting. We interact the off-year, municipal election index with the Democrat dummy variable.

- Competitive Democratic District * Democrat: Used to account for different levels of salience in the elections geographically as well as between Democrats and unaffiliated voters. We interact the Democrat dummy variable with an indicator of competitiveness. The competitiveness variable takes a value of 1 if the individual lives in the 2nd Congressional District.

- Competitive Democratic District * Unaffiliated: Used to account for different levels of salience in the elections geographically as well as between Democrats and unaffiliated voters. We interact the unaffiliated dummy variable with an indicator of competitiveness. The competitiveness variable takes a value of 1 if the individual lives in the 2nd Congressional District. With this variable and the Competitive Democratic District * Democrat variable above, the comparison group is Democrats and unaffiliated voters who live in districts that are "uncompetitive" in the 2010 Democratic primary.

Survey: Colorado 2010 Statewide Republican Primary

Dependent Variable: The 2008 statewide primary election. See "Identification of Eligible Voters for Primary Elections" below.

Independent Variables

- General Election Index: An index indicating the number of elections the individual voted in using the three most recent general elections: 2008, 2006, and 2004.

- Primary Election Index: An index indicating the number of elections the individual voted in using the three most recent primary elections: 2006, 2004, and 2002.

- Municipal Election Index: An index indicating the number of elections the individual voted in using the three most recent off-year elections: 2009, 2007, and 2005.

- Republican: A dummy variable taking a value of 1 if individual is a registered Republican.

- Years Since Original Registration: The time measured in years between the date of the 2010 primary election and the date the individual first registered to vote in Florida.

- Years Since Last Registration Change: The time measured in years between the date of the 2010 primary election and the date of the last change in an individual's registration status. Changes could occur because the voter moved, changed party affiliation, or other reasons.

- Age: Measured in years from Election Day for the 2010 primary.

- Age Squared: Used to account for nonlinearities in the effect of age on voting probability.

- Gender: A dichotomous variable coded 1 for female.

- General Election Index * Republican: Used to account for a different effect among Republicans and Unaffiliated voters in general election voting. We interact the general election index with the Republican dummy variable.

- Primary Election Index * Republican: Used to account for a different effect among Republicans and Unaffiliated voters in primary election voting. We interact the primary election index with the Republican dummy variable.

- Off-Year Election Index * Republican: Used to account for a different effect among Republicans and Unaffiliated voters in off-year, municipal election voting. We interact the off-year election index with the Republican dummy variable.

- Competitive Republican District * Republican: Used to account for different levels of salience in the elections geographically as well as between Republican and unaffiliated voters. We interact the Republican dummy variable with an indicator of competitiveness. The competitiveness variable takes a value of 1 if the individual live in the 5th or 6th Congressional Districts.

- Competitive Republican District * Unaffiliated: Used to account for different levels of salience in the elections geographically as well as between Democrats and unaffiliated voters. We interact the unaffiliated dummy variable with an indicator of competitiveness. The competitiveness variable takes a value of 1 if the individual live in the 5th or 6th Congressional Districts. With this variable and the Competitive Republican District * Republican vraiable, the comparison group is Republicans and unaffiliated voters who live in districts that are "uncompetitive" in the 2010 Republican primary.

<u>**Survey: Utah 2010 2ⁿᵈ Congressional District Democratic Primary**</u>

Dependent Variable: The 2008 statewide presidential primary election. We use this election as the dependent variable because Utah has not had a Democratic primary election (statewide or in the 2ⁿᵈ district) for more than a decade. This leaves us with no election that closely mirrors the 2010 2ⁿᵈ Congressional District primary. Given this limitation, we select the 2008 Democratic Presidential primary election contested by Hillary Clinton and Barack Obama. This election has the advantage of being recent, potentially competitive, and salient. We felt that these characteristics most closely mirrored the 2010 election.

Independent Variables

- General Election Index: An index indicating the number of elections the individual voted in using the three most recent general elections: 2008, 2006, and 2004.

- Primary Election Index: An index indicating the number of elections the individual voted in using the three most recent primary elections: 2008, 2006, and 2004.

- Municipal Election Index: An index indicating the number of elections the individual voted in using the three most recent off-year elections: 2009, 2007, and 2005.

- Democrat: A dummy variable taking a value of 1 if the individual is a registered Democrat.

- Years Since Original Registration: The time measured in years between the date of the 2010 election and the date the individual first registered to vote in Utah.

- Age: Measured in years from Election Day for the 2010 primary.

- Age Squared: Used to account for nonlinearities in the effect of age on voting probability.

- Age * Years Since Original Registration: An interaction of the age variable and the years since original registration date variable.

- General Election Index * Democrat: Used to account for a different effect among Democrats and Unaffiliated voters in general election voting. We interact the general election index with the Democrat dummy variable.

- Primary Election Index * Democrat: Used to account for a different effect among Democrats and Unaffiliated voters in primary election voting. We interact the primary election index with the Democrat dummy variable.

<u>**Survey: Utah 2010 Statewide Republican Primary**</u>

Dependent Variable: The 2008 statewide presidential primary election. While not an exact match to other primaries, this was the only recent statewide primary election held in the state.

Independent Variables

- General Election Index: An index indicating the number of elections the individual voted in using the three most recent general elections: 2008, 2006, and 2004.

- Primary Election Index: An index indicating the number of elections the individual voted in using the three most recent primary elections: 2008, 2006, and 2004.

- Municipal Election Index: An index indicating the number of elections the individual voted in using the three most recent off-year elections: 2009, 2007, and 2005.

- Republican: A dummy variable taking a value of 1 if the individual is a registered Republican.

- Years Since Original Registration: The time measured in years between the date of the 2010 election and the date the individual first registered to vote in Utah.

- Age: Measured in years from Election Day for the 2010 primary.

- Age Squared: Used to account for nonlinearities in the effect of age on voting probability.

- Age * Years Since Original Registration: An interaction of the age variable and the years since the original registration date variable.

- General Election Index * Republican: Used to account for a different effect among Republicans and Unaffiliated voters in general election voting. We interact the general election index with the Republican dummy variable.

- Primary Election Index * Republican: Used to account for a different effect among Republicans and Unaffiliated voters in primary election voting. We interact the primary election index with the Republican dummy variable.

<u>**Survey: Utah 2008 General Election**</u>

Dependent Variable: The most recent presidential election, the 2004 Utah presidential election.

Independent Variables

- General Election Index: An index indicating the number of elections the individual voted in using the three most recent general elections: 2006, 2002, and 2000.

- Primary Election Index: An index indicating the number of elections the individual voted in using the four most recent primary elections: 2006, 2004, 2002, and 2000.

- Presidential Primary: A dichotomous variable taking the value of 1 if the individual voted in the 2008 presidential primary election in Utah.

- Republican: A dummy variable taking a value of 1 if the individual is a registered Republican.

- Democrat: A dummy variable taking a value of 1 if the individual is a registered Democrat.

- Other Party: A dummy variable taking a value of 1 if individual is registered with a party that is not the Republican or Democrat Parties and is not an unaffiliated voter.

- Years Since Original Registration * Age: This variable is a series of five dummy variables indicating the quintile of the continuous registration variable that the individual is a member of and then interacted with each of five quintiles of the continuous age variable.

- Age: Measured in years from Election Day for the 2008 general election.

- Age Squared: Used to account for nonlinearities in the effect of age on voting probability.

Survey: Utah 2008 3rd Congressional District Republican Primary

Dependent Variable: The 2006 3rd Congressional District Republican primary election.

Independent Variables
- General Election Index: An index indicating the number of elections the individual voted in using the six most recent general elections: 2006, 2004, 2002, 2000, 1998, and 1996.

- Primary Election Index: An index indicating the number of elections the individual voted in using the four most recent primary elections: 2002, 2000, 1998, and 1996.

- Presidential Primary: A dichotomous variable taking the value of 1 if the individual voted in the 2008 presidential primary election in Utah.

- Republican: A dummy variable taking a value of 1 if the individual is a registered Republican.

- Age: Measured in years from Election Day for the 2008 primary.

- Registration * Republican: The continuous registration variable is divided into quintiles and interacted with a dummy variable indicating whether the individual is a registered Republican.

Identification of Eligible Voters for Primary Elections

In Colorado, unaffiliated voters are allowed to declare their affiliation to either party to vote in either primary. Therefore, the models and samples for each party's 2010 Colorado primary election included unaffiliated as well as the registered partisans. Unaffiliated voters in Colorado were eligible for sampling in both surveys, although they generally have low probabilities of voting in either primary and accordingly had a low chance of selection for either PPS sample.

In order to avoid asking the same individual to respond to both surveys, we removed any individuals that were sampled for both surveys from one of the surveys. In the Colorado samples, 1 unaffiliated voter was removed from the Republican sample because she was also sampled for the Democratic survey. In Utah, only registered Republicans can vote in a Republican primary. However, unaffiliated voters can register with the Republican Party at the polling location on Election Day. The Utah Democratic Party allows registered Democrats and unaffiliated voters to vote in its primary. Again, because of the potential for sampling an unaffiliated voter in both surveys, 45 unaffiliated respondents were deleted from the 2010 Utah Republican primary sample to avoid double sampling of individuals.

Supplementary Materials C: Example Invitation Letters

CENTER FOR THE STUDY OF
ELECTIONS AND DEMOCRACY

Brigham Young University

«First_Name» «Last_Name»
«Address»
«City», CO «Zip»

June 12, 2010

Dear «First_Name»:

As the June 22, 2010 U.S. Senate primary election approaches, I invite you to participate in a special edition of the *Utah Voter Poll* conducted by the Center for the Study of Elections and Democracy at Brigham Young University.

You were selected at random from a list of all registered voters in Utah. Your participation is very important to us and will help make the survey accurate. This online survey takes about 10 minutes to complete and your answers are completely confidential.

To ensure that only voters who have been invited can participate in the survey, we have provided a unique access code. To begin the survey:

- Enter the following URL into any web browser: **http://utahvoterpoll.org/senate**
- Click on "**CLICK HERE TO BEGIN SURVEY**"
- Enter this five-digit "Access Code" in the space provided: **«Access_Code»**
-

If you have trouble accessing the survey, please email *utahvoterpoll@byu.edu* or call 801-422-3716. A list of "Frequently Asked Questions" is on our web site at *http://utahvoterpoll.org/faq.htm*

The survey is available now. Please begin and finish the survey before it closes at midnight on Monday, June 21st. Thank you very much for helping with this important study.

Sincerely,

J. Quin Monson, Ph.D.
Center for the Study of Elections and Democracy
Brigham Young University

P.S. Your participation is very important to us. Please take the survey before midnight on Monday, June 21st.

1142 Kimball Tower ■ Provo, UT 84602 ■ phone: (801) 422-3716

University of Colorado at Boulder
Department of Political Science

Ketchum 106
333 UCB
Boulder, Colorado 80309-0333

«First_Name» «Last_Name»
«Address»
«City», CO «Zip»

August 1, 2010

Dear «First_Name»:

As the August 10, 2010 Colorado primary elections approach, I invite you to participate in a special online poll conducted through the University of Colorado.

You were selected at random from a publicly available list of all registered voters in Colorado. Your participation is very important to us and will help make the survey accurate. This online survey takes about 10 minutes to complete, and your answers are completely confidential. Please note that your participation in this survey is voluntary, and you may stop at any time – none of your personal information will ever be shared with political organizations or the public.

To ensure that only voters who have been invited can participate in the survey, we have provided a unique access code. To begin the survey:

- Enter the following URL into any web browser: **http://coloradopoll.org/**
- Click on **"CLICK HERE TO BEGIN SURVEY"**
- Enter this six-digit "Access Code" in the space provided: **«Access_Code»**

If you have trouble accessing the survey, please email me at anand.sokhey@colorado.edu, or call the poll helpline toll free at 888-207-7317. A list of "Frequently Asked Questions" is available on our web site at http://coloradopoll.org/faq.htm

The survey is <u>available now</u>. Please begin and finish the survey before it closes at midnight on Monday, August 9th. Thank you very much for helping with this important study.

Sincerely,

Anand Edward Sokhey, Ph. D.
Department of Political Science
The University of Colorado at Boulder
email: anand.sokhey@colorado.edu

P.S. Your participation is important to us. Please take the survey before midnight on Monday, August 9th!

U University of Miami
Department of Political Science

5250 University Drive
PO Box 248047
Coral Gables, FL 33124-6534

<<First Name>> <<Last Name>>
<<Address>>
<<City>>, FL <<Zip Code>>

October 22, 2010

Dear <<First Name>>:

As the November 2, 2010 Florida general election approaches, I invite you to participate in a special online poll conducted through the University of Miami.

You were selected at random from a publicly available list of all registered voters in Florida. Your participation is very important to us and will help make the survey accurate. This online survey takes about 10 minutes to complete, and your answers are completely confidential. Please note that your participation in this survey is voluntary, and you may stop at any time — none of your personal information will ever be shared with political organizations or the public.

To ensure that only voters who have been invited participate in the survey, we have provided a unique access code. To begin the survey:

- Enter the following URL into any web browser: **http://umfloridapoll.org/**
- Click on "**CLICK HERE TO BEGIN SURVEY**"
- Enter this five-digit "Access Code" in the space provided: **<<Access Code>>**

If you have trouble accessing the survey, please email me at cmann@miami.edu, or call the poll helpline at 888-207-7317. A list of "Frequently Asked Questions" is available on our web site at *http://umfloridapoll.org/*

The survey is available now. Please begin and finish the survey before it closes at midnight on Monday, November 1st. Thank you very much for helping with this important study.

Sincerely,

Christopher B. Mann

Christopher B Mann

Christopher B. Mann
Assistant Professor
Department of Political Science
University of Miami
email: cmann@miami.edu

P.S. Your participation is important to us. Please take the survey before midnight on Monday, November 1!

Supplementary Materials D: Public Polling Data
Table S2: Public Polling Data for Figure S3

Election	State	Election Type	Office	Polling Firm	Field Dates	Sample Size	Sample Type	Winner Public Poll Forecast	2nd Place Public Poll Forecast	3rd Place Public Poll Forecast	Winner Vote Share Forecast Closer to Actual
2008	Utah	Primary - Republican	3rd CD	Deseret News	6/18-6/20	312	RV	47.6	52.4	-	Our Poll
2008	Utah	General	1st CD	Deseret News	10/24 - 10/30	1205	RV	69.0	31.0	-	Public Poll
2008	Utah	General	2nd CD	Deseret News	10/24 - 10/30	1205	RV	71.4	28.6	-	Our Poll
2008	Utah	General	3rd CD	Deseret News	10/24 - 10/30	1205	RV	70.7	29.3	-	Our Poll
2008	Utah	General	Governor	Mason-Dixon	08/13 - 08/15	400	LV	89.0	11.0	-	Date Not Comparable
2008	Utah	General	President	Deseret News	10/24 - 10/30	1205	RV	64.0	36.0	-	Public Poll
2008	Utah	General	President	Mason-Dixon	10/23 - 10/25	625	LV	63.2	36.8	-	Public Poll
2010	Colorado	Primary - Republican	Governor	PPP (D)	8/7-8/8	767	LV	49.4	50.6	-	Our Poll
2010	Colorado	Primary - Republican	Governor	Denver Post/Survey USA	7/27-7/29	588	LV	52.4	47.6	-	Our Poll
2010	Colorado	Primary - Republican	Senate	PPP (D)	8/7-8/8	767	LV	48.9	51.1	-	Our Poll
2010	Colorado	Primary - Republican	Senate	Denver Post/Survey USA	7/27-7/29	588	LV	54.9	45.1	-	Our Poll
2010	Colorado	Primary - Democrat	Senate	PPP (D)	8/7-8/8	448	LV	53.3	46.7	-	Public Poll
2010	Colorado	Primary - Democrat	Senate	Denver Post/Survey USA	7/27-7/29	536	LV	48.4	51.6	-	Our Poll
2010	Utah	Primary - Democrat	2nd CD	Deseret News	6/12-6/17	409	LV	61.2	38.8	-	Our Poll
2010	Utah	Primary - Republican	Senate	Deseret News	6/12-6/17	581	LV	44.0	56.0	-	Our Poll
2010	Florida	General	Governor	Sunshine State News/VSS	10/31 - 11/1	1526	LV	52.1	47.9	-	Our Poll
2010	Florida	General	Governor	PPP (D)	10/30 - 10/31	773	LV	49.5	50.5	-	Public Poll

2010	Florida	General	Governor	Quinnipiac	10/25 - 10/31	925	LV	49.4	50.6	-	Public Poll
2010	Florida	General	Governor	Rasmussen Reports	10/27 - 10/27	750	LV	51.6	48.4	-	Public Poll
2010	Florida	General	Governor	Mason-Dixon	10/25 - 10/27	625	LV	48.3	51.7	-	Public Poll
2010	Florida	General	Governor	Florida Poll/NYT-USF	10/23 - 10/27	696	LV	53.0	47.0	-	Our Poll
2010	Florida	General	Senate	PPP (D)	10/30 - 10/31	773	LV	48.0	30.6	21.4	Our Poll
2010	Florida	General	Senate	Sunshine State News/VSS	10/29 - 10/31	1527	LV	48.5	31.3	20.2	Our Poll
2010	Florida	General	Senate	Quinnipiac	10/25 - 10/31	925	LV	47.9	33.0	19.1	Our Poll
2010	Florida	General	Senate	Rasmussen Reports	10/27 - 10/27	750	LV	52.1	31.3	16.7	Our Poll
2010	Florida	General	Senate	Mason-Dixon	10/25 - 10/27	625	LV	47.9	29.8	22.3	Our Poll
2010	Florida	General	AG	Mason-Dixon	10/25 - 10/27	625	LV	54.3	45.7	-	Public Poll
2010	Florida	General	AG	Ipsos Public Affairs	10/15-10/19	577	LV	55.0	45.0	-	Public Poll
2010	Florida	General	CFO	Mason-Dixon	10/25 - 10/27	625	LV	58.2	41.8	-	Our Poll
2010	Florida	General	CFO	Ipsos Public Affairs	10/15-10/19	577	LV	58.4	41.6	-	Our Poll
2010	Florida	General	Amdt 6	Mason-Dixon	10/25 - 10/27	625	LV	63.0	37.0	-	Public Poll
2010	Florida	General	Amdt 6	Ipsos Public Affairs	10/15-10/19	577	LV	68.2	31.8	-	Our Poll

Note: Undecided voters in public polls are allocated proportionally for comparison with our forecasts, since our surveys did not allow an undecided response option.

Table S3: Pre-Election Forecasts and Actual Election Outcomes for Figure S3

		Actual Results	Forecast Results	95% M.E.	N
2010 Florida General Election					
US Senate:	Rubio	48.9	48.7	± 4.8	199
	Meek	20.2	19.1		78
	Crist	29.7	32.1		131
Governor:	Scott	48.9	45.8	± 4.9	183
	Sink	47.7	51.0		204
Attny General:	Bondi	54.8	56.4	± 4.9	219
	Gelber	41.4	42.8		166
CFO:	Atwater	57.3	57.8	± 5.1	215
	Ausley	38.9	40.3		150
Amendment 1:	Yes	52.2	54.2	± 5.3	187
	No	47.5	45.8		158
Amendment 6:	Yes	62.2	65.9	± 5.1	245
	No	37.1	34.1		127
2010 Colorado GOP Primary					
US Senate:	Buck	51.6	50.3	± 4.5	245
	Norton	48.4	47.6		232
Governor:	Maes	50.6	50.4	± 4.7	202
	McInnis	49.3	44.1		231
2010 Colorado Dem Primary					
US Senate:	Bennett	54.2	50.0	± 4.2	273
	Ranomoff	48.9	48.9		267
2010 Utah Statewide GOP Primary					
US Senate:	Lee	51.2	47.4	± 4.6	219
	Bridgewater	48.8	50.4		233
2010 Utah 2CD Dem Primary					
US District 2:	Matheson	67.3	63.6	± 5.0	250
	Wright	32.7	34.4		135
2008 Utah General					
President:	McCain	63.1	61.1	± 4.1	377
	Obama	33.9	31.4		194
Governor:	Huntsman	77.9	78.2	± 4.1	474
	Springmeyer	19.5	17.5		106
Attny General:	Shurtleff	69.9	70.8	± 4.0	425
	Hill	26.4	27.2		163
US District 1:	Bishop	65.1	60.1	± 7.4	110
	Bowen	30.1	34.4		63
US District 2:	Matheson	63.1	62.8	± 6.7	140
	Dew	34.7	33.6		75
US Distsrict 3:	Chaffetz	66.1	67.2	± 7.2	131
	Bennion	27.8	27.2		53
2008 3CD Utah GOP Primary					
US District 3:	Chaffetz	59.8	54.4	±4.5	243
	Cannon	40.2	41.4		185

This table shows the actual results of each race within each poll as well as the predicted result, the 95% margin of error for each question and the number of people responding in each race. We see that in every election the actual result falls within the poll margin of error.

Supplementary Materials E: Discussion of Sampling Methods

Probability Proportionate to Prediction vs. Probability Proportionate to Size Sampling

It is important to note that in PPS sampling the probability of selection is known before sampling begins and the total sample size is also determined before sampling begins. Another method of unequal probability sampling is known as Probability Proportionate to Prediction sampling, or PPP sampling. In PPP sampling, unlike PPS sampling, the probability of inclusion in the sample is unknown before the sample is drawn. When drawing the sample, the researcher estimates an upper bound on the size of all units in the population and then chooses a value, L, larger than that estimate. As each observation is encountered, the size of the observation is observed, and a random number, x_i, is drawn from the interval [0,L]. If x_i is smaller than the measured size of the unit, then the unit is included in the sample. Thus, larger units have higher probability of being included in the sample. Note that the total size of the sample is unknown until all units have been observed, and the probability of selection is not known before sampling begins. While we do use a predicted probability, our sampling method is closer to PPS than PPP sampling since we can calculate the probability of selection beforehand *and* the total sample size is determined before sampling begins.

Simple Random Sampling vs. Our PPS Sampling Method

In the 2008 Utah primary and general elections, we compared the performance of surveys using a simple random sample of registered voters to our approach to using PPS to sample the likely electorate.[1] The SRS sample and PPS sample distributions in the 2008 Utah primary and general elections have typical distributions of likelihood of voting in these types of elections. Figures S4 (Primary) and S5 (General) display the distribution of predicted probability of voting for the all eligible registered voters, the distributions in the PPS and SRS samples, the distribution of respondents in each sample, and the distribution of actual voters according to individual turnout records from election officials (Figures S4 & S5 are similar to Figure 1 in the text).

[Figures S4 and S5 About Here]

We begin by comparing all eligible registered voters (first row) to the samples (second row). On the left, the distribution of the SRS sample mirrors the distribution for all registered voters in the top row, as expected. On the right, the PPS sample distributions are skewed towards registered voters more likely to vote – and closely resemble the actual voters in the bottom row. The PPS sample distributions are different in the primary and general elections due to the differences between the underlying distributions from which the PPS samples were drawn. In the general election, the differences are quite small because the probability of turning out in the 2008 general election resembles the uniform distribution of the SRS. The main difference between the SRS and PPS samples is low turnout probability voters on the left side of the histogram are less likely to be included in the PPS sample. The impact of the PPS sample is more dramatic in low turnout elections like the 2008 Utah primary. For the primary, the SRS sample mirrors registered voters with a strong skew to the left because many registered voters have a low individual likelihood of voting in the

---

[1] In the 2008 3rd Congressional District Primary, 2000 voters were sampled using PPS sampling and 8000 were sampled using simple random sampling. In the 2008 Utah general election, 5000 voters were sampled using each method.

primary. In the PPS sample, the large share of voters on the left is discounted by their low probability of voting, while the small share of voters on the right is inflated because of their high probability of voting. Balancing of the density of distribution of voters and intensity from the predicted likelihood of voting creates a PPS sample that closely resembles the actual electorate in the last row.

Before examining the distribution of the respective survey respondents in the third row of Figures S4 & S5, we examine the response rates for each type of sample. Voting and participating in surveys correlate with levels of interest in, engagement with, and knowledge about elections. Therefore, we expected people with a higher probability of voting are also more likely to complete the survey. Since higher probability voters make up a larger portion of the PPS samples than the SRS samples in both types of elections, we expected higher response rates from the PPS samples. This hypothesis proved true in both 2008 Utah elections. In the primary, the completion rate in the SRS sample was 5.61 percent and 10.10 percent in the PPS sample. In the general election, the response rate for the SRS sample was 5.56 percentwhile the PPS sampling response rate was 6.96 percent. The narrower gap in the general election is consistent with the smaller difference in the distributions of the general election PPS and SRS samples.

In the third row of Figures S4 & S5, we see the respondents in both columns reflect the respective SRS and PPS samples (second row) as expected. Therefore, the SRS sample respondents continue to reflect the distribution of all eligible registered voters (top row) while the PPS sample respondents closely resemble the distribution of the actual turnout (bottom row). These results further support the idea of drawing likely voter samples by considering those observable characteristics that correlate with voting. A Kolmogorov-Smirnov test comparing distributions confirms the PPS outperforms the SRS in both the primary and general elections in 2008. In both cases, the distributions of predicted probabilities among those who were *sampled* using PPS and

those who *responded* from that sample are closer to the distribution of predicted probabilities of those who actually *voted* than the SRS sample and respondents.

The distribution of respondents in the SRS samples illuminate why pre-election polling in low turnout elections is so difficult and costly when using simple random samples of registered voters (and more so when starting with an SRS of the general population). SRS respondents are biased away from the sampling frame of voters in the upcoming election because low probability voters are substantially over-represented compared to the actual electorate. As the proportion of voters with a low individual probability of voting increases, as in primaries and local elections, the gap between an SRS sample and the likely electorate grows. In conventional pre-election polls for general elections that start with an SRS sample, techniques such as screens relying on self-reported vote intention only need to do a small amount of work to refine the sample of responses to be representative of the likely electorate. In a primary election, a survey using an SRS sample relies heavily on likely voter screens and other techniques for selecting or weighting responses to compensate for the gap between the sample and the likely electorate.

Using conventional deterministic approaches to screening SRS samples to identify "likely voters" can cause bias in either direction. When voters with low individual probability of turning out in the upcoming election are screened out, the survey respondents are biased by the exclusion of many people who will actually vote – particularly in low turnout elections like primaries. Although individual respondents have a low probability of voting, the people they represent in a sample may make up a significant share of the actual electorate. However, an SRS sample may also be biased by including too many people drawn from the low-turnout probability end of the distribution. The skew of initial respondents towards voters with low individual probabilities of turning out, especially in primaries (and other low-turnout elections), makes it possible that the screening questions will select a disproportionate number of individuals who over-report being personally likely to vote (due

to social desirability and other biases in self-reporting). Thus, the available techniques for closing the gap between an SRS sample and the intended likely electorate sampling frame can fail in both directions. At best, voters across the distribution of turnout probability have similar preferences so the sample distribution does not bias pre-election forecasts. However, in this best case, the use of these techniques is costly, due to longer screening batteries, discarded responses from unlikely voters, and other factors.

Table S3 displays the pre-election forecasts in the Utah 2008 surveys from the PPS samples in comparison to the SRS samples. The point estimates of the PPS sample are more accurate than the SRS sample in five of seven races.

[Table S4 About Here]

We note the PPS sample was not more accurate than the SRS sample in every race in Table S4. Therefore, future research should investigate a possible connection between forecast accuracy and level of interest, information, and/or contestation. Perhaps sampling based on previous voting behavior may be superior in cases when likely voters are also highly informed about the candidates, but if habitual voters are as uninformed or uninterested as unlikely voters, the difference between sampling methods may disappear.

Figure S4: Distribution of Predicted Probabilities for 2008 Primary Election by Sampling Method
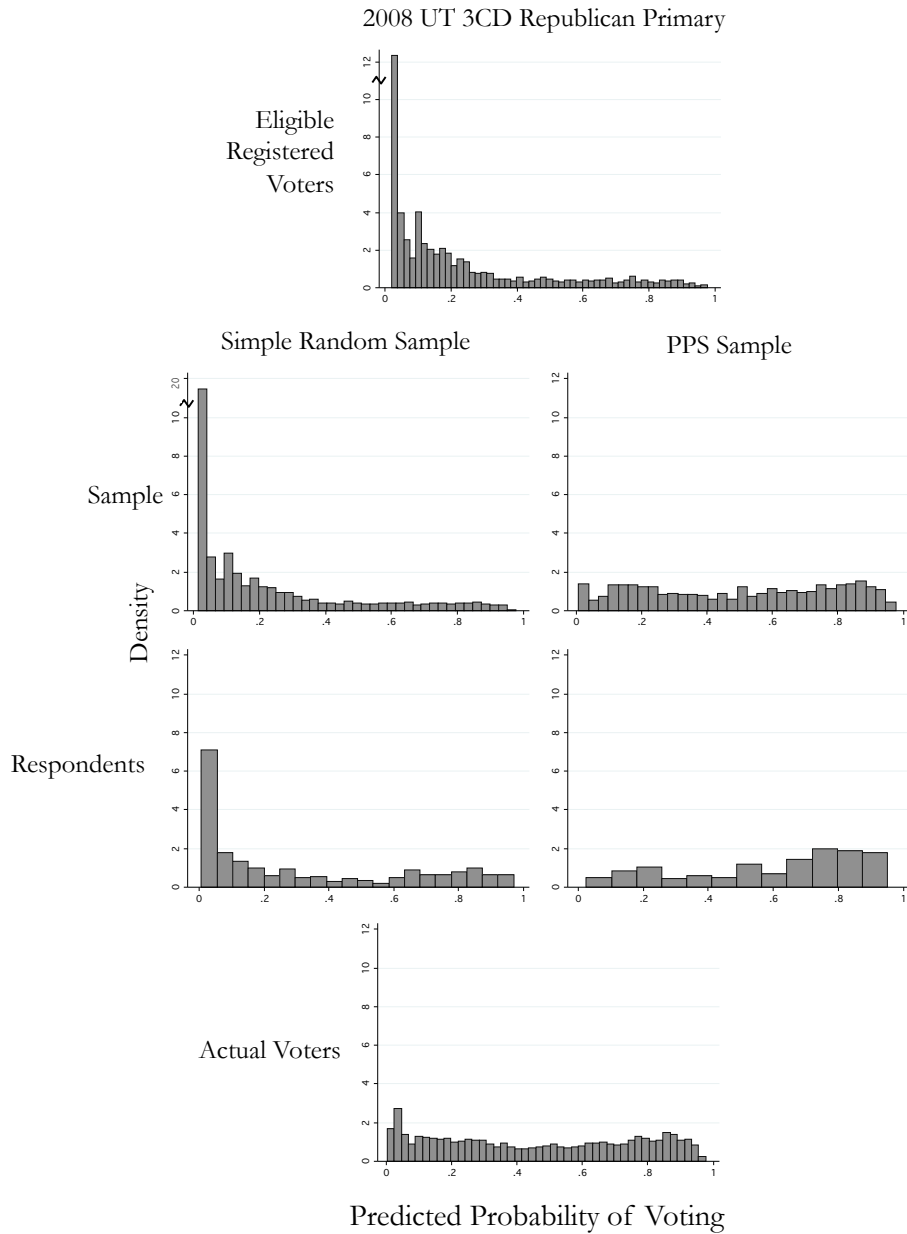
This figure shows the different distributions that arise from the different sampling methods used in the 2008 Utah 3rd Congressional Republican primary. While the simple random sample closely mirrors the distribution of the voting population (row 1), the PPS sample closely mirrors the distribution of voters. Since our poll is concerned with the opinions of voters, not eligible voters, these distributions suggest the PPS method is superior to the simple random sample.

Figure S5: Distribution of Predicted Probabilities for 2008 General Election by Sampling Method
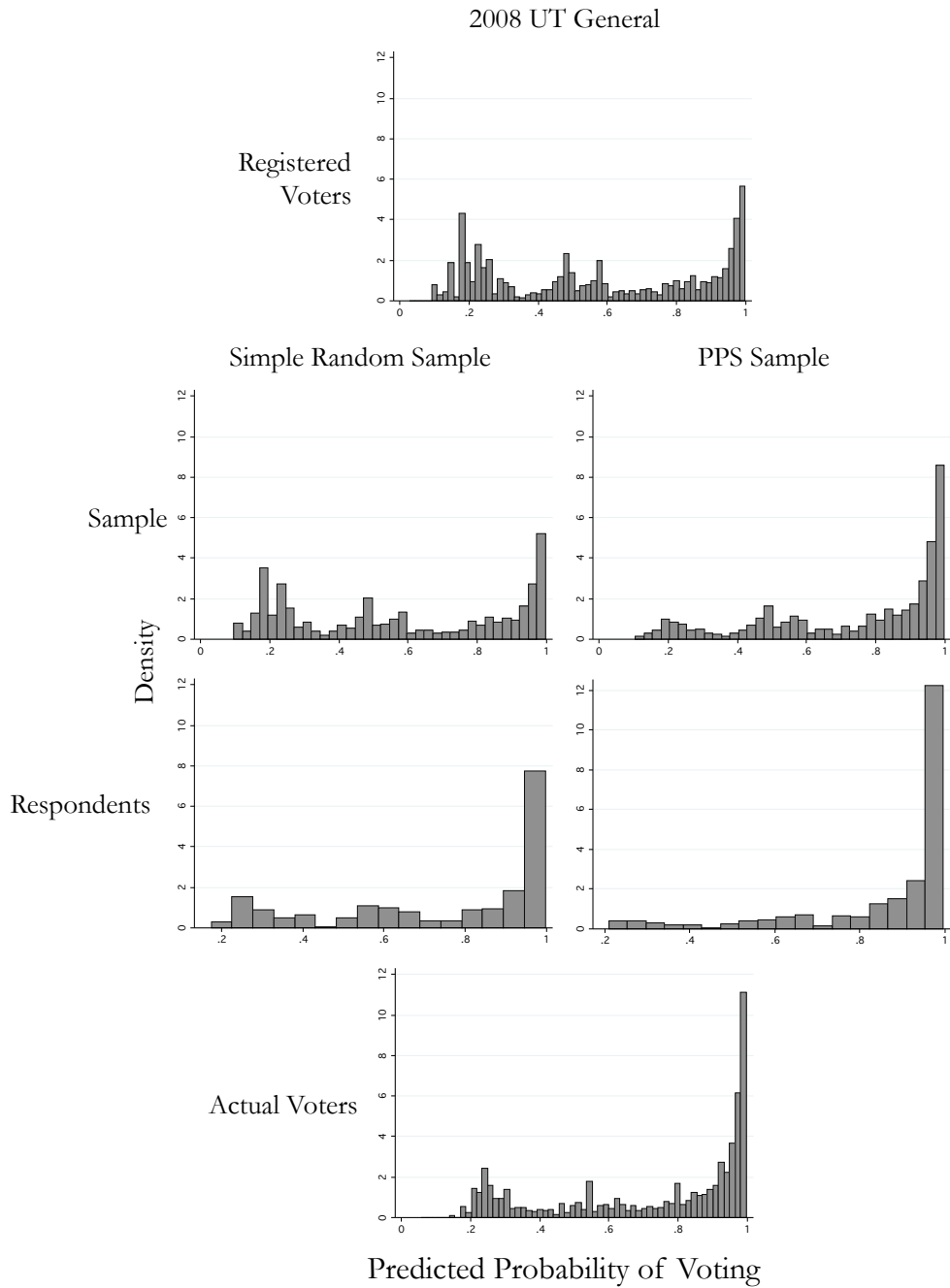
This figure shows the different distributions that arise from the different sampling methods used in the 2008 Utah general election. While the simple random sample closely mirrors the distribution of the voting population (row 1), the PPS sample closely mirrors the distribution of voters. In this case, the distribution of eligible voters is closer to the distribution of voters. However, the PPS distributions are closer to the distribution of voters than the SRS distributions.

Table S4: Comparison of Pre-Election Forecasts from PPS Samples and SRS Samples

	PPS Sample		Random Sample			
	Forecast	N	Forecast	N	Actual Results	Winner
2008 3CD Utah GOP Primary						
US District 3:						
Chaffetz	52.71	68	52.88	147	59.8	
Cannon	41.86	54	43.53	121	40.2	PPS
2008 Utah General						
President:						
McCain	57.85	199	65.20	178	63.1	
Obama	33.72	116	28.57	78	33.9	SRS
Governor:						
Huntsman	77.65	264	78.95	210	77.9	
Springmeyer	17.94	61	16.92	45	19.5	PPS
Attny General:						
Surtleff	68.25	230	74.14	195	69.9	
Hill	29.67	100	23.95	63	26.4	PPS
US District 1:						
Bishop	51.46	53	71.25	57	65.1	
Bowen	42.72	44	23.75	19	30.1	SRS
US District 2:						
Matheson	69.35	86	54.55	54	63.1	
Dew	26.61	33	42.42	42	34.7	PPS
US District 3:						
Chaffetz	64.91	74	70.37	57	66.1	
Bennion	28.07	32	25.93	21	27.8	PPS

This table displays the comparison between the PPS samples and the simple random samples in the 2008 Utah survyes where the two methods were used. We see that the PPS sample was more accurate than the simple random sample in five of the seven races.