Vote Estimation in Pre-election Polls: Who Are the Likely Voters?

Michael W. Traugott
Center for Political Studies
Institute for Social Research
University of Michigan

Transitioning from Compulsory Voting to Voluntary Voting

- **In a C V system, pre-election polls require a good sample of eligibles and good measurement of preferences**

- **In a V V system, beyond the sampling and measurement issues, the "likely" electorate has to be identified - sometimes described as the "secret sauce" of pre-election polling**

- **In the U.S., between 50 and 60% of voting age population does not vote**

Conceptualizing the Likely Electorate

A complicated problem because the "concept" does not exist in reality until Election Day

It requires a model consisting of multiple components that are translated into empirical measures and then combined into a single score
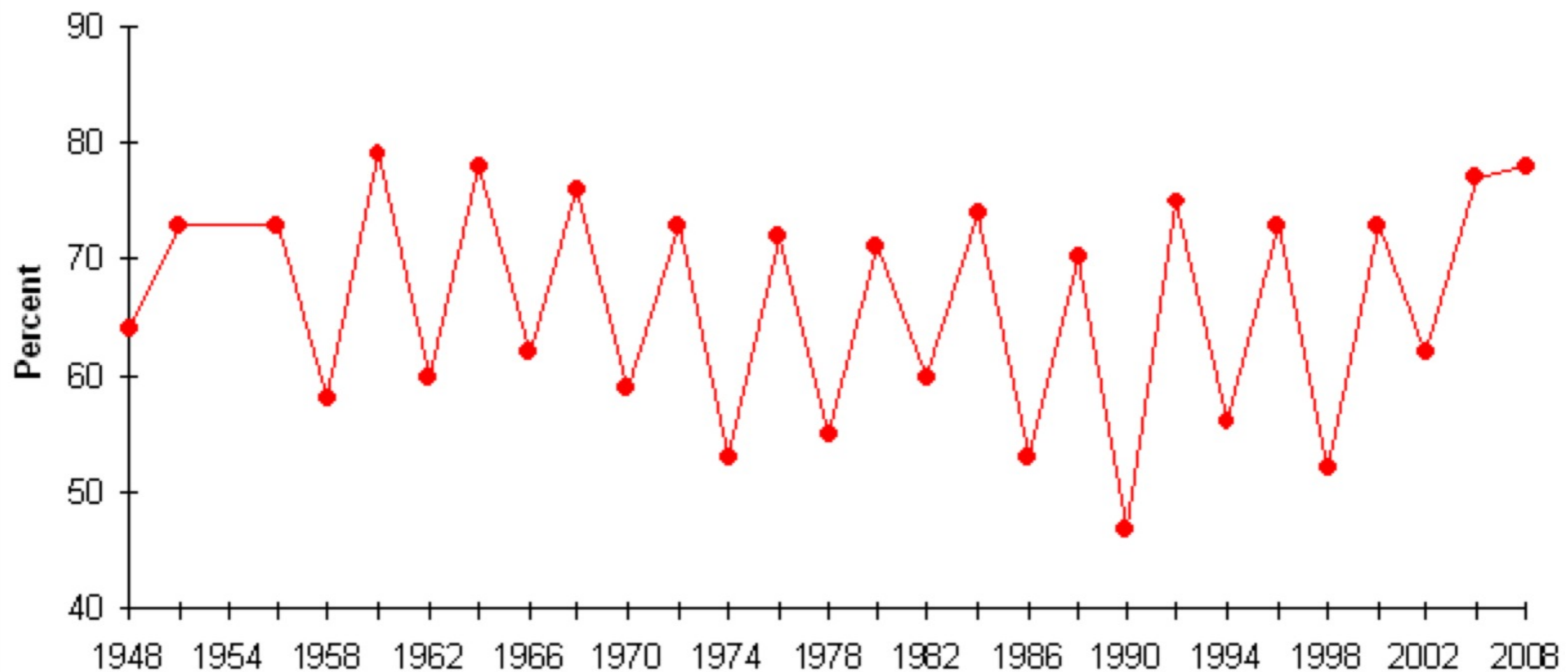
Screening, scaling, and propensity modeling

Relationship between samples of adults, registered voters and likely voters in the U.S.

Conceptualizing the Likely Electorate

Voter Turnout, Self-Report by Respondents 1948-2008

RESPONSE YES, VOTED

Graph 6A.2.2   Source: The American National Election Studies   Aug-23-2010

Conceptualizing the Likely Electorate

Presidential Turnout Rates 1948-2012

Issues in Estimating Likely Voters: Do past models hold?

In a first, black voter turnout rate passes whites

BY HOPE YEN

TOPICS: FROM THE WIRES, 4 NEWS, POLITICS NEWS

WASHINGTON (AP) — America's blacks voted at a higher rate than other minority groups in 2012 and by most measures surpassed the white turnout for the first time, reflecting a deeply polarized presidential election in which blacks strongly supported Barack Obama while many whites stayed home.

Had people voted last November at the same rates they did in 2004, when black turnout was below its current historic levels, Republican Mitt Romney would have won narrowly, according to an analysis conducted for The Associated Press.

Issues in Estimating Likely Voters: Do past models hold?

Presidential campaigns target just 10 states, ignore the other 40

- 96% of the spending on television ads between April 11th and November 6th by presidential campaigns and allied groups went to ten battleground states.
- Nearly six times as much ad money was spent in Florida alone than was spent in the 40 non-swing states and DC.
- 99% of campaigns stops by the presidential or vice presidential candidates were in these states.

Conceptualizing the Likely Electorate

What are the main components?

Eligibility

Social psychological attributes

Political attributes

Demographics

Past behavior

Conceptualizing the Likely Electorate

What are the main components?

Eligibility

Age

Citizenship

Is registration required?

Conceptualizing the Likely Electorate

What are the main components?

Social psychological attributes: Long term

    Sense of citizen duty

    Personal efficacy

Social psychological attributes: Short term

    Interest in the campaign

    Sense the outcome will make a difference

    Self-reported likelihood of voting (Commitment to participate)

Conceptualizing the Likely Electorate

What are the main components?

Political attributes

Partisanship (strength and direction)

Contact by one or more campaigns

Political activity (past and present)

Attention to the media

Know where or how to vote

Conceptualizing the Likely Electorate

What are the main components?

Demographics

Socioeconomic status (Convenience)

Education / Political knowledge

Conceptualizing the Likely Electorate

What are the main components?

Past behavior

Voting in past similar elections

Frequency of past voting

Measuring the Likely Electorate: Validated Voting Rates by Question

Previously cast a vote in your precinct (65%)

Always or nearly always vote (63%)

Knowing where people in neighborhood vote (63%)

Follow what's going on in government and public affairs "most" or "some of the time" (63%)

Giving "some" or "a lot of" thought to the upcoming election (50%)

Source: Pew Philadelphia Validation Study, 2006

How to combine components?

A summative index with one or more cutoff points

A categorization with a probability of voting attached to each category

These cut points and probabilities are best determined with validation studies across several elections and can vary by election type (Reliability and Validity)

How to combine components? Pew

Pew Research Center:
1999 Philadelphia Validation

Chances of voting 10-pt scale	%		Voted
10	77	→	84%
9	6	→	71%
8	6	→	46%
7	3	→	33%
1-6	7	→	39%
DK/Refused	<u>1</u>		
	100		

How to combine components? Gallup

1. Thought given to election (quite a lot, some)
2. Know where people in neighborhood go to vote (yes)
3. Voted in election precinct before (yes)
4. How often vote (always, nearly always)
5. Plan to vote in 2010 election (yes)
6. How certain to vote (absolutely certain)
7. Voted in last midterm election (yes)

How to combine components? Gallup

The following adjustments are made to these raw scores:

- Respondents who are not registered to vote are assigned a score of 0.
- Respondents who do not say they plan to vote (see item 5 above) are assigned a score of 0.
- Younger respondents' scores are adjusted to account for their ineligibility to vote in some or all past elections. In other words, even though the model identifies voters on the basis of past voting history, younger voters are not penalized for not being of voting age in past election years.
  -- If aged 18-19, their scores are converted as follows: 1=2, 2=4, 3=5, 4+=7
  -- If aged 20-21, their scores are converted as follows: 1=1, 2=3, 3=4, 4=6, 5+=7

Probabilistic model: CBS News
(Adapted from Traugott and Tucker 1984)

[1] The question wordings of the items taken from the 1980 CPS Preelection Survey are:

Now, how about the election this November?
Are you registered so that you could vote in the November election if you wanted to?
    1. Yes
    5. No

Now, in 1976 you remember that Gerald Ford ran on the Republican ticket against Jimmy Carter for the Democrats.
Do you remember for sure whether or not you voted in that election?
    1. Yes, did vote   Past Voter
    5. No, didn't vote
    7. Don't remember if voted   } Nonvoter
    0. Inap, not of voting age in 1976

Some people don't pay much attention to political campaigns. How about you? Would you say that you have been very much interested, somewhat interested, or not much interested in following the political campaigns so far this year?
    1. Very much interested   High Interest
    3. Somewhat interested   } Low Interest
    5. Not much interested

The variables were combined in the following fashion to construct the eight categories of the index:

Unregistered	— Nonvoter	— Low Interest
Unregistered	— Past Voter	— Low Interest
Unregistered	— Nonvoter	— High Interest
Unregistered	— Past Voter	— High Interest
Registered	— Nonvoter	— Low Interest
Registered	— Nonvoter	— High Interest
Registered	— Past Voter	— Low Interest
Registered	— Past Voter	— High Interest

Probabilistic model: CBS News
(Adapted from Traugott and Tucker 1984)

Table 1. The Likelihood of Voting Index Applied to the 1980 CPS Preelection Survey

Likelihood of Voting	Proportion of Voting Age Population	Validated Voting Rate[a]	Proportion of Electorate in Group[b]
Low (1)	10.6%	3.6%	.004
(2)	1.5	12.5	.002
(3)	10.1	8.9	.009
(4)	3.8	29.7	.011
(5)	3.7	27.6	.010
(6)	9.0	58.2	.052
(7)	10.3	61.9	.063
High (8)	51.0	74.6	.380
	(1,560)		.531

Issues in Estimating Likely Voters

Do pollsters start with a preconceived notion of what turnout will be and construct a sample of that size?

OR

Do pollsters conceptualize their measure, combine the measurements, and see what turnout is likely to look like?

Problems in recent U.S. presidential elections and Barack Obama's get out the vote (GOTV) campaigns

How to combine components?

A summative index with one or more cutoff points

A categorization with a probability of voting attached to each category

These cut points and probabilities are best determined with validation studies across several elections and can vary by election type (Reliability and Validity)

What are the main issues pollsters face?

How "different" are voters from nonvoters?

Would it be better to start with a list of "likely" voters as a sample frame?

Does a public poll have a different purpose than a poll done for a candidate/party?

The complications of future pre-election polling: U.S. examples

The electoral system is changing all the time

Convenience as a factor: early, absentee ballots, voting by mail, super centers

Technology of the future affects polling AND voting

Hybrid designs and multi-mode methods

Polling will get more complicated: transparency is key

Conclusions

Pre-election polling is a difficult task in voluntary voting systems

When conducted using scientific methods, we can better understand when and how it works - and when it doesn't (more or less accurate estimation)

There is a shared interest in doing pre-election polling well, and information sharing is the key to common success