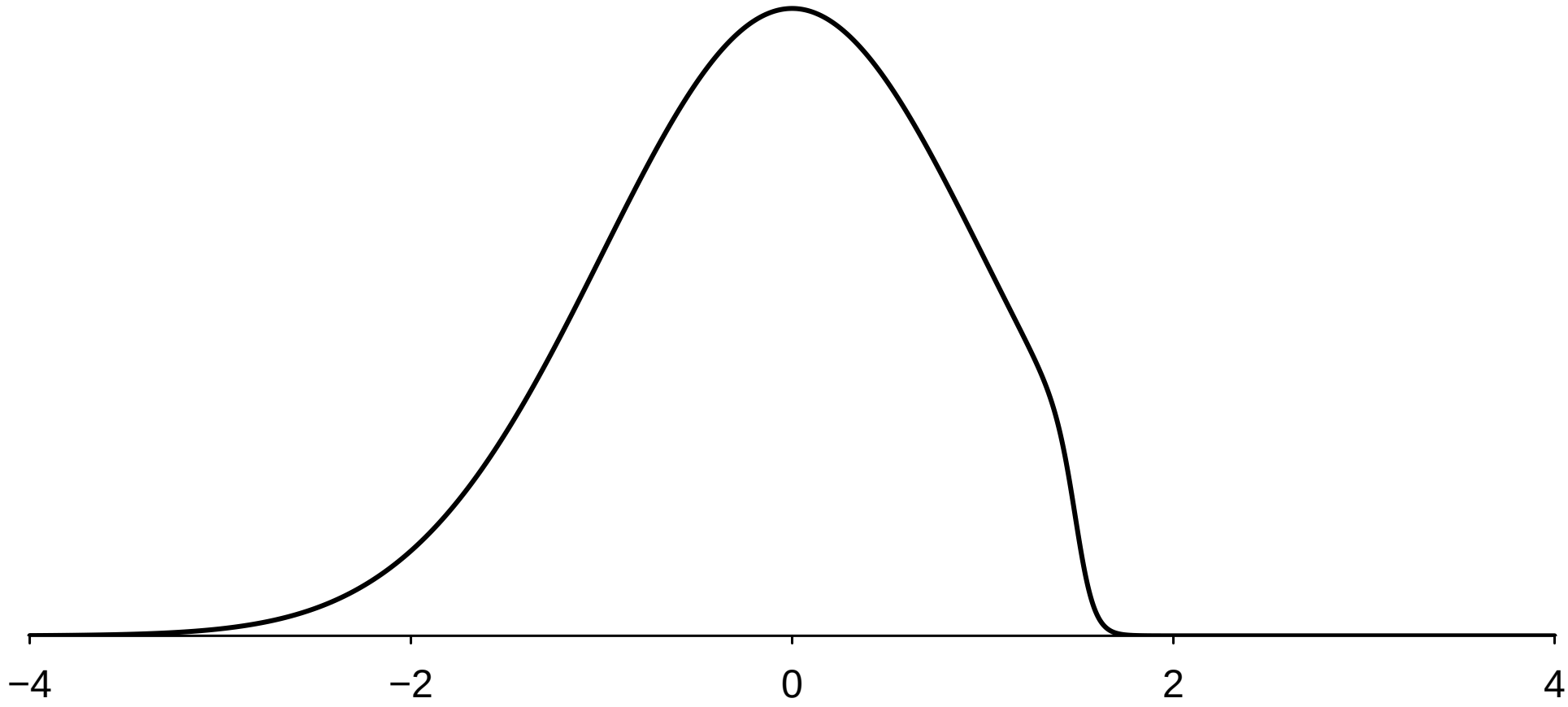
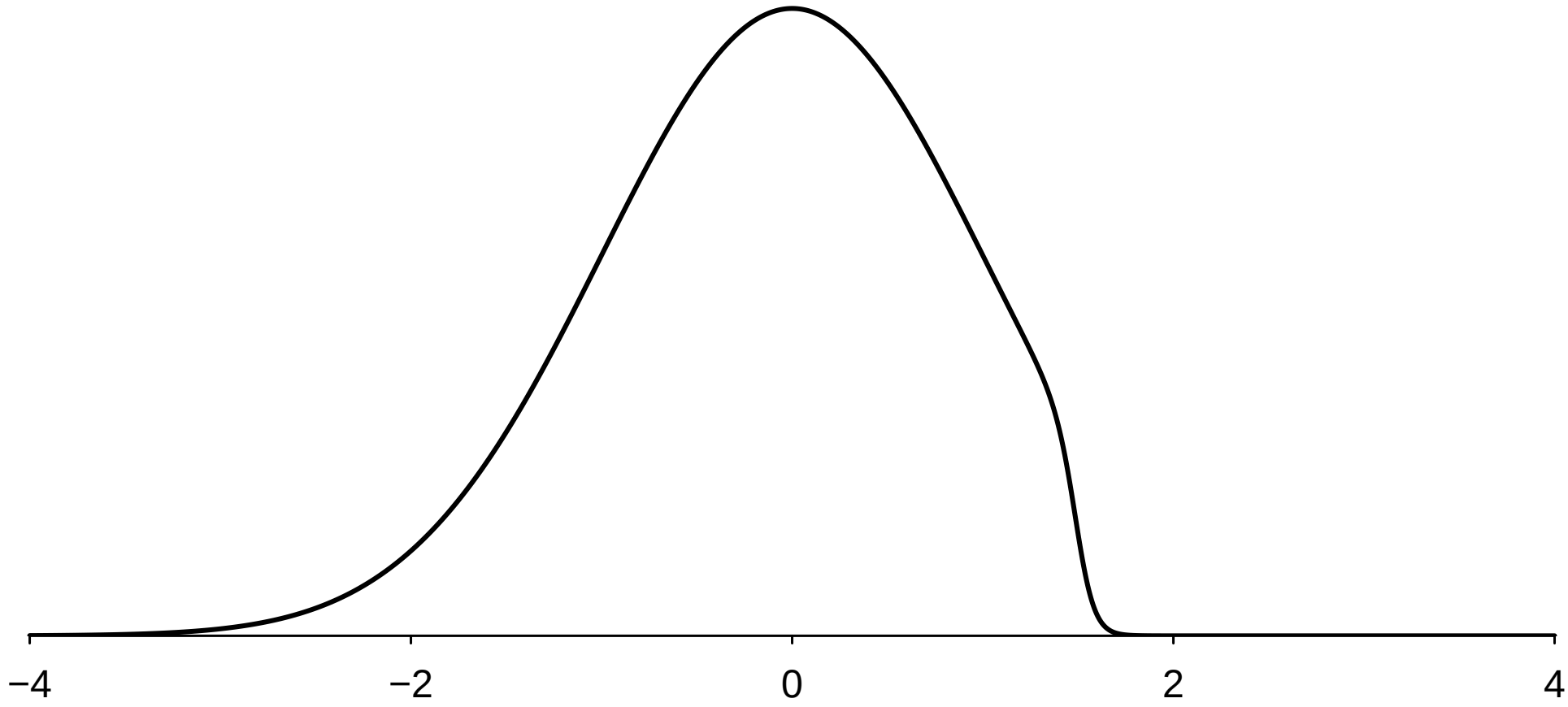
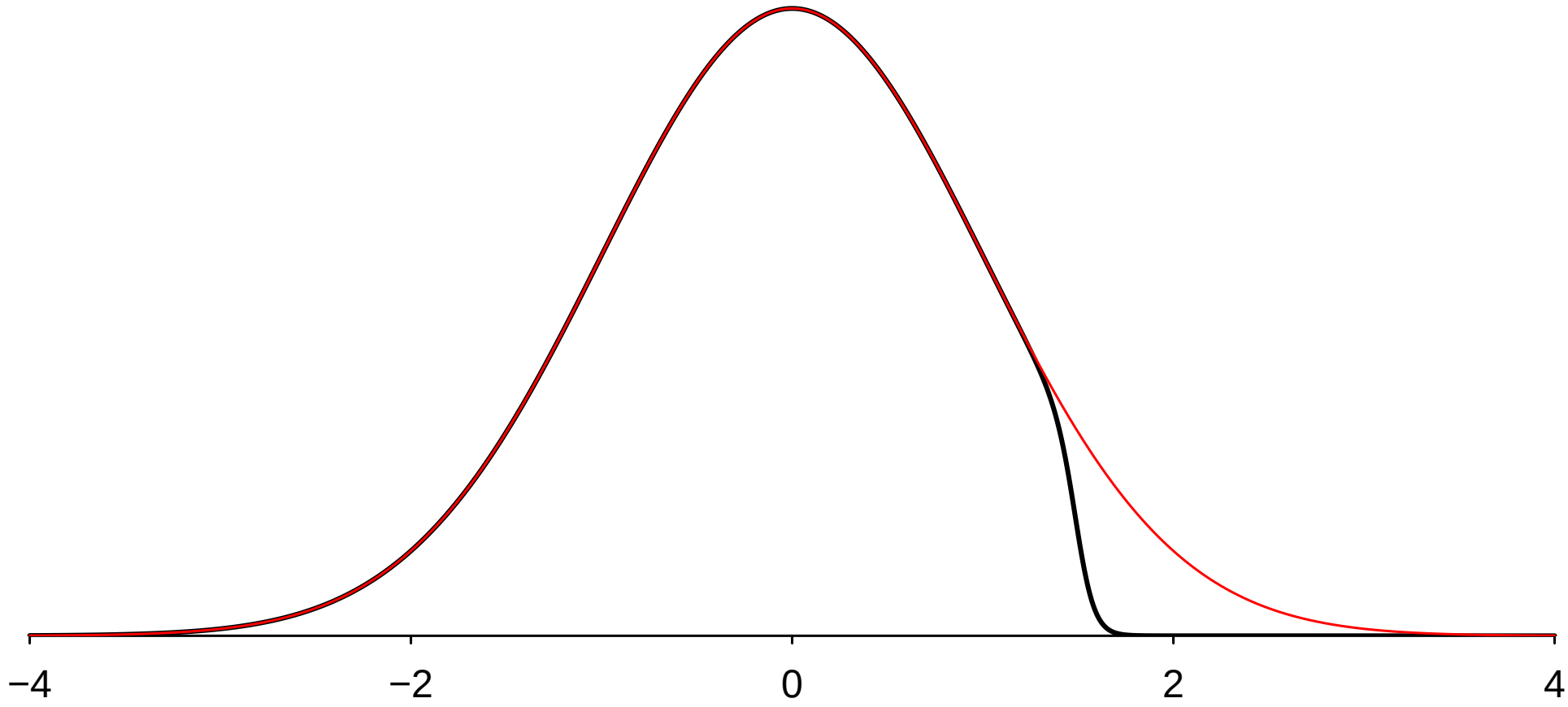
Observed data



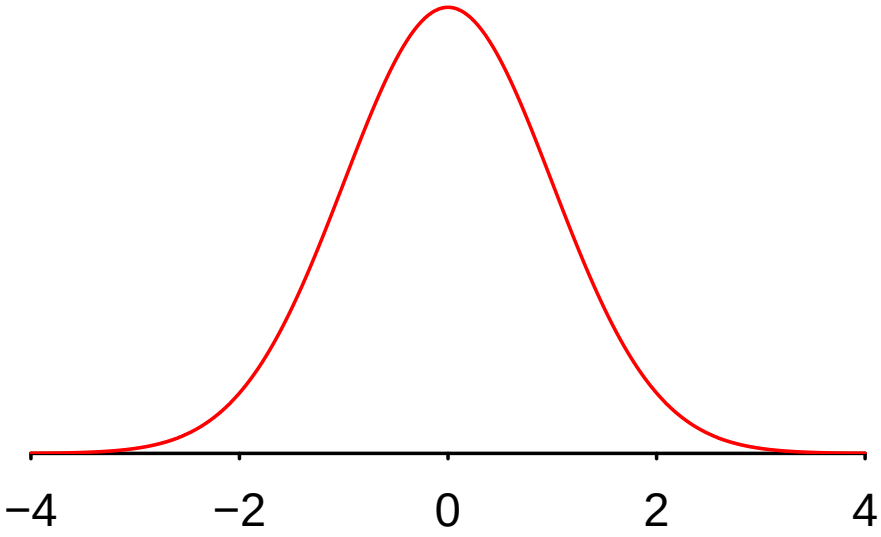
Imputed data−−Missing completely at random

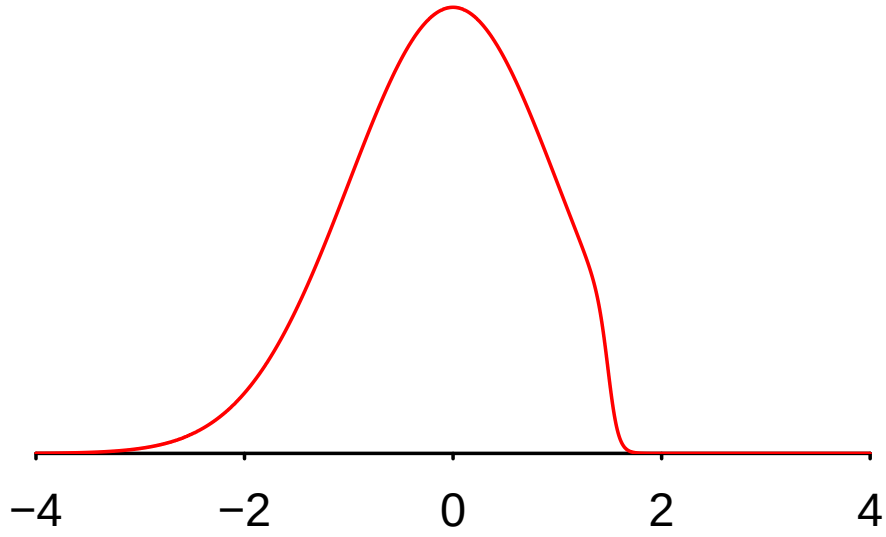


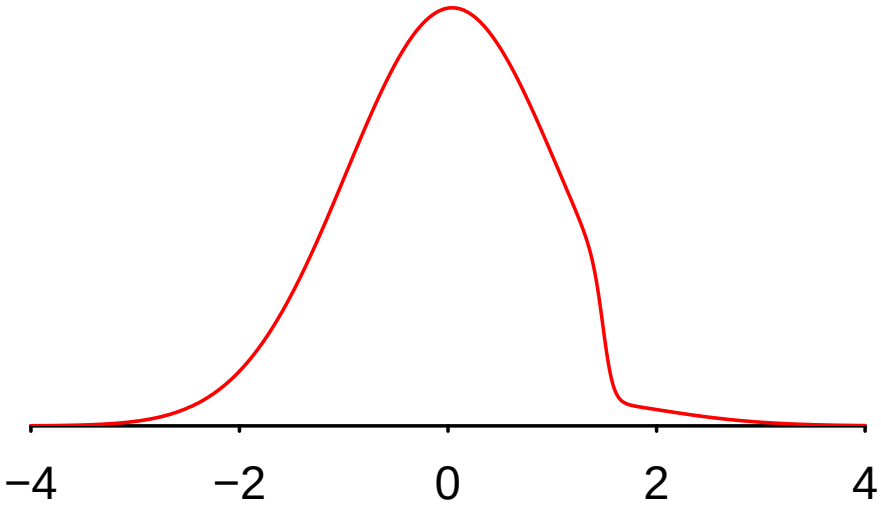
Imputed data−−Fitted normal model

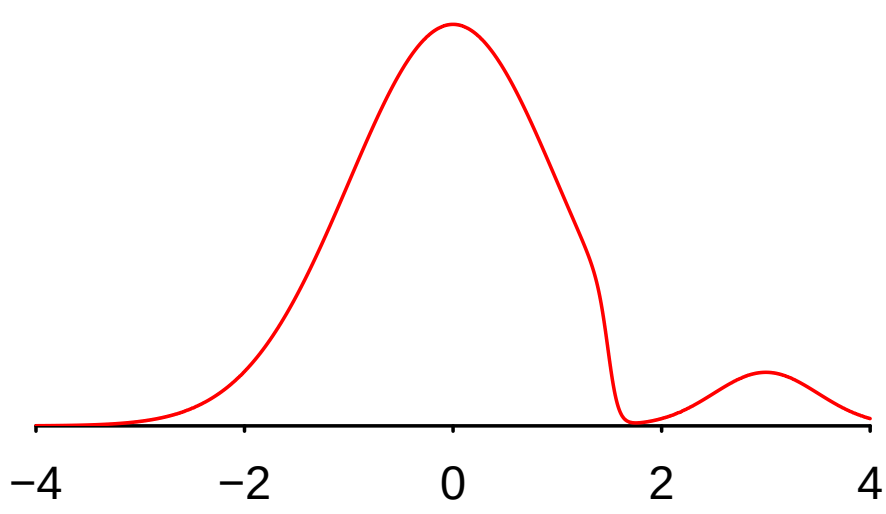


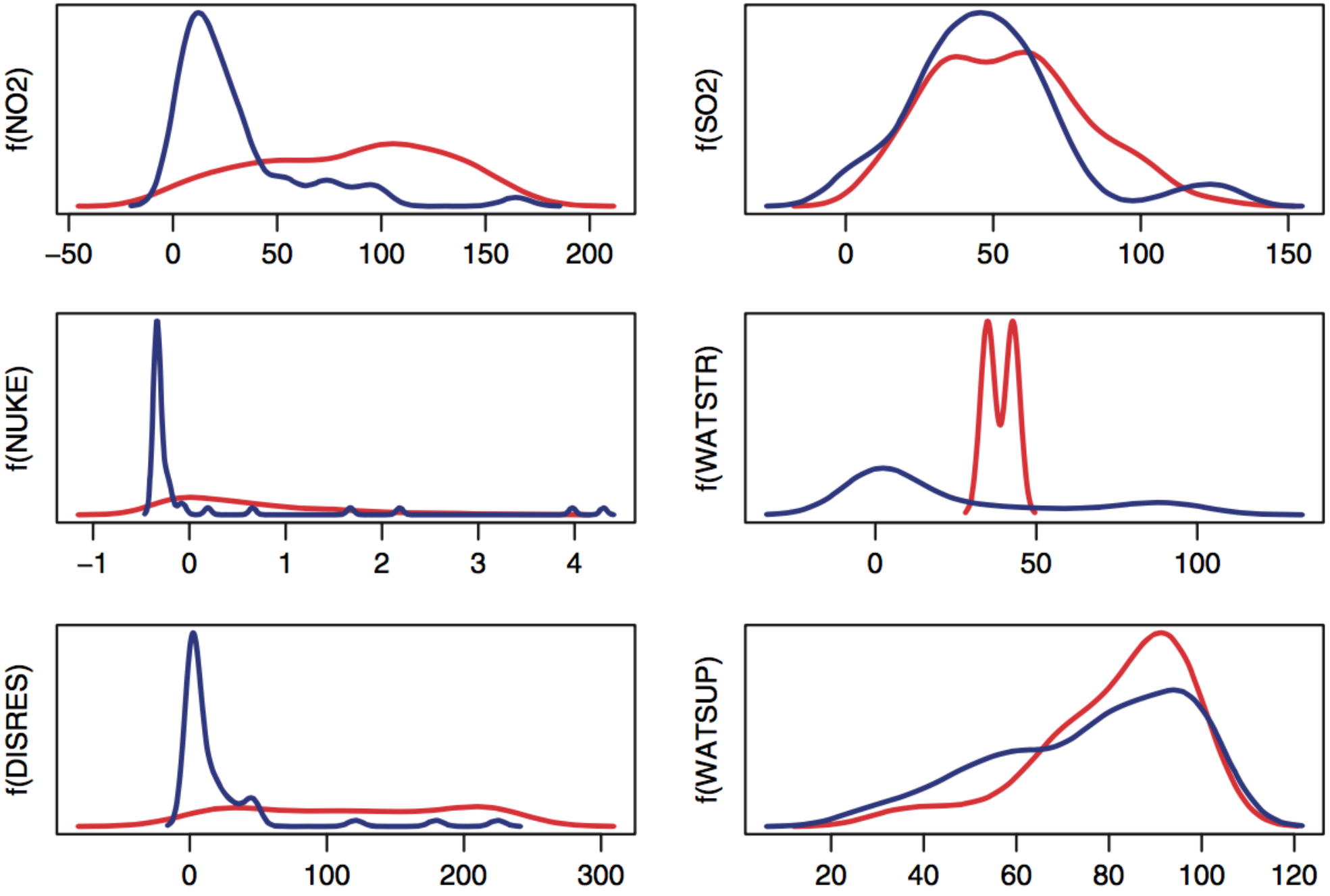
Imputed data−−Various assumptions










f(NO2)
−50
0
50
100
150
200
f(SO2)
0
50
100
150
f(NUKE)
−1
0
1
2
3
4
f(WATSTR)
0
50
100
f(DISRES)
0
100
200
300
f(WATSUP)
20
40
60
80
100
120

Should the problems with polls make us worry about the quality of health surveys?

Andrew Gelman

Department of Statistics and Department of Political Science
Columbia University, New York

Centers for Disease Control and Preventionnecticut, 21 Feb 2017

From Wikipedia:

> There is no satisfactory statistical solution to deal with missing data that may not be at random. Assuming an extreme bias in the responders is one suggested method of dealing with low survey response rates. Getting a high response rate (>80%) from a small, random sample is considered preferable to a low response rate from a large sample.[9]

Trend data were not readily available from the large number of private organizations who conduct survey research and polls. These organizations claim to be experiencing problems as the following excerpt from the 1973 Conference on Surveys of Human Populations (American Statistical Association, 1974) shows:

> "...spokesmen for a number of private survey organizations, large and small, who were queried by one of the conference participants, all report that their completion rates on general population samples now average approximately 60 to 65 percent, in spite of three or four callbacks. This recent experience is in contrast to a completion figure of 80 to 85 percent for the same firms in the decade of the sixties."

AD A05
AD No.
DDC FILE COPY

(6) Survey Response Rates: Some Trends, Causes and Correlates,

(10) Kent H./Marquis

DDC
MAY 8 1978
F

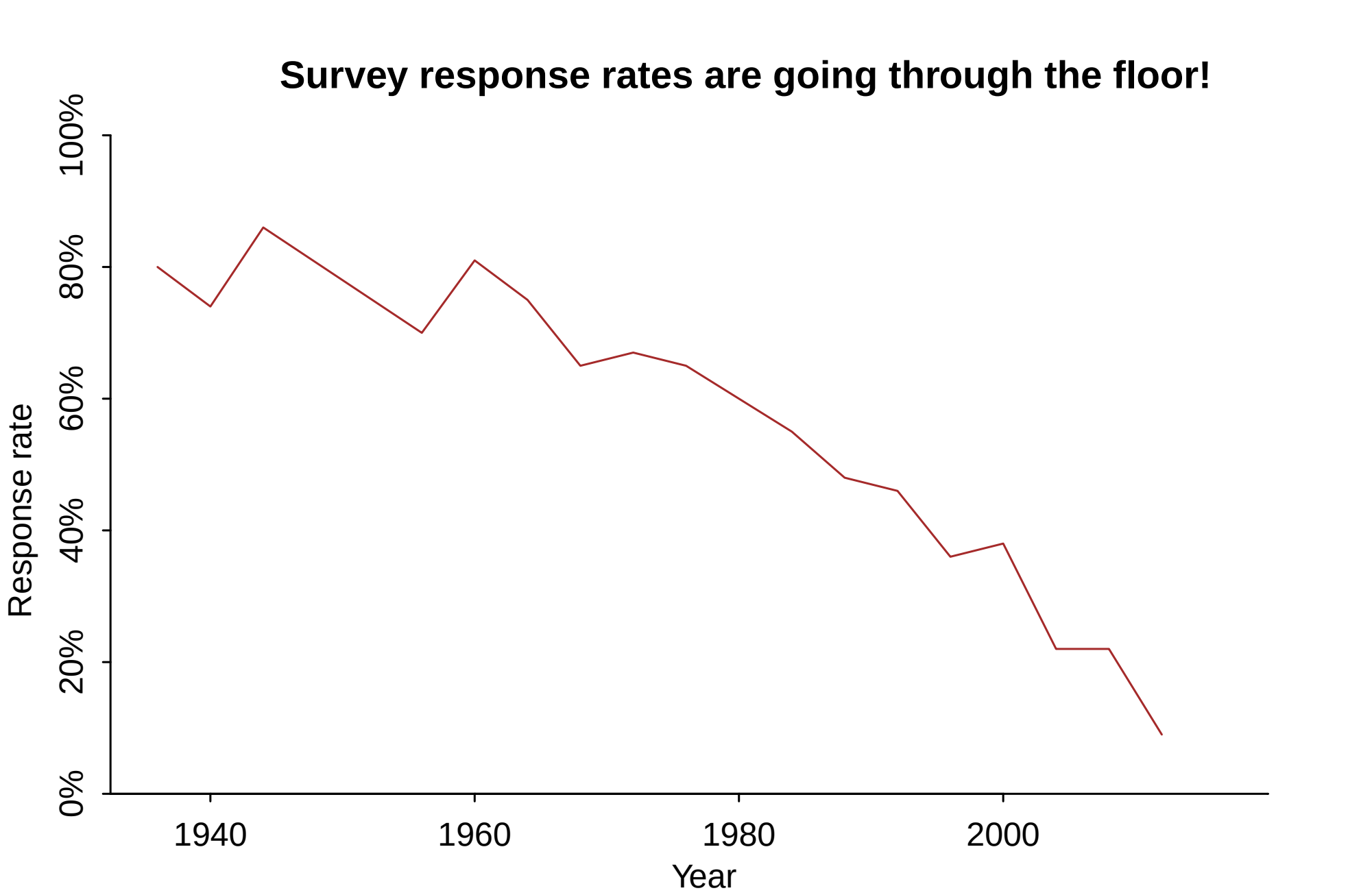
(11) 1 Apr[redacted]77

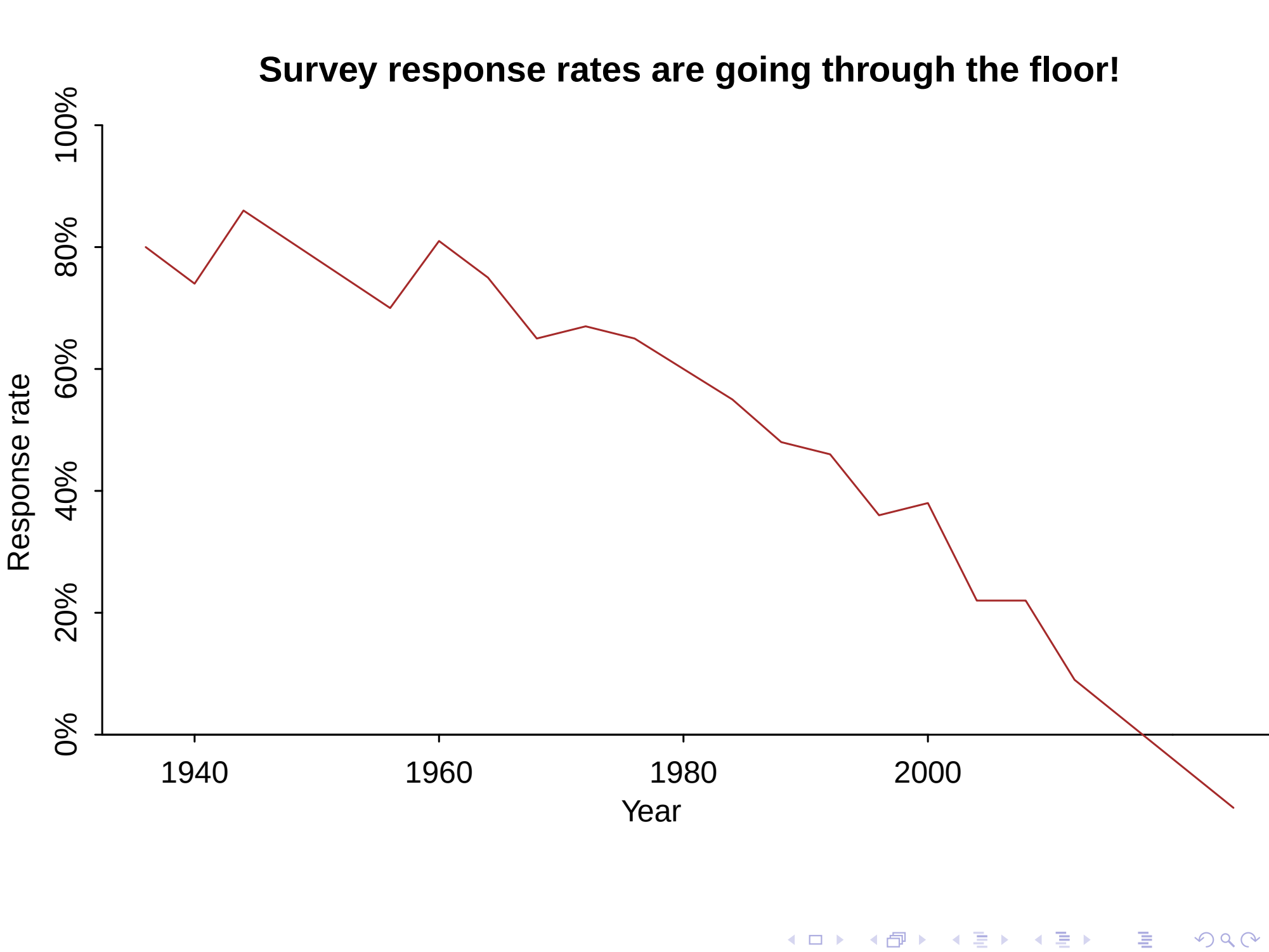
(12) 29p.

Surveys Face Growing Difficulty Reaching, Persuading Potential Respondents

	1997	2000	2003	2006	2009	2012
	%	%	%	%	%	%
Contact rate (percent of households in which an adult was reached)	90	77	79	73	72	62
Cooperation rate (percent of households contacted that yielded an interview)	43	40	34	31	21	14
Response rate (percent of households sampled that yielded an interview)	36	28	25	21	15	9

PEW RESEARCH CENTER 2012 Methodology Study. Rates computed according to American Association for Public Opinion Research (AAPOR) standard definitions for CON2, COOP3 and RR3. Rates are typical for surveys conducted in each year.


Survey response rates are going through the floor!
Response rate
0%
20%
40%
60%
80%
100%
1940
1960
1980
2000
Year


Survey response rates are going through the floor!
Response rate
0%
20%
40%
60%
80%
100%
1940
1960
1980
2000
Year

The poststratification identity

$$\theta = \frac{\sum_{j=1}^{J} N_j \theta_j}{\sum_{j=1}^{J} N_j}$$

The poststratified estimate

$$\hat{\theta} = \frac{\sum_{j=1}^{J} N_j \hat{\theta}_j}{\sum_{j=1}^{J} N_j}$$

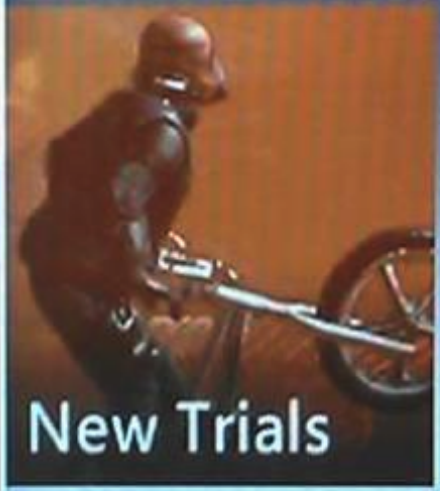
mister P

me social tv video games music apps

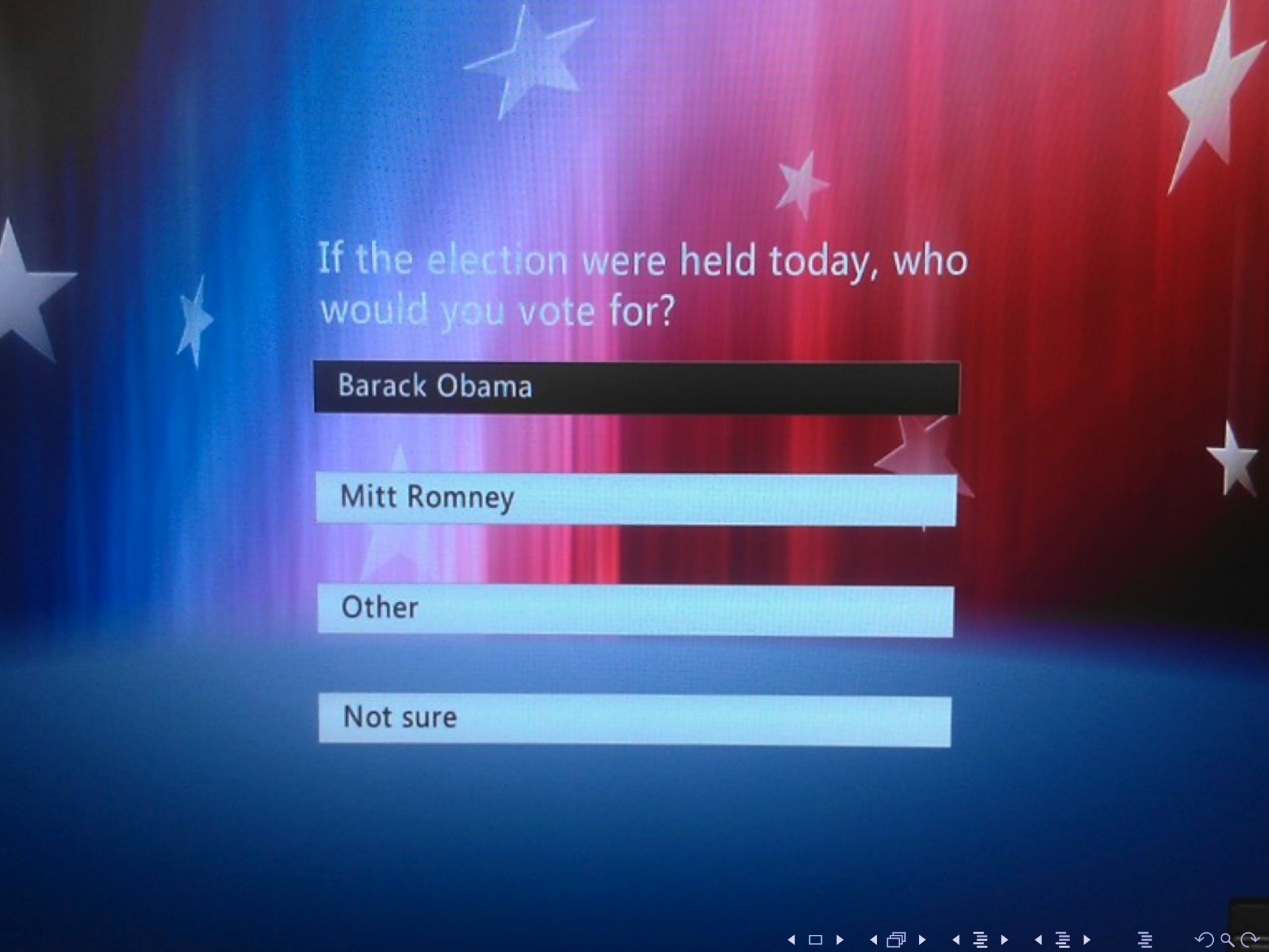
UFC 153

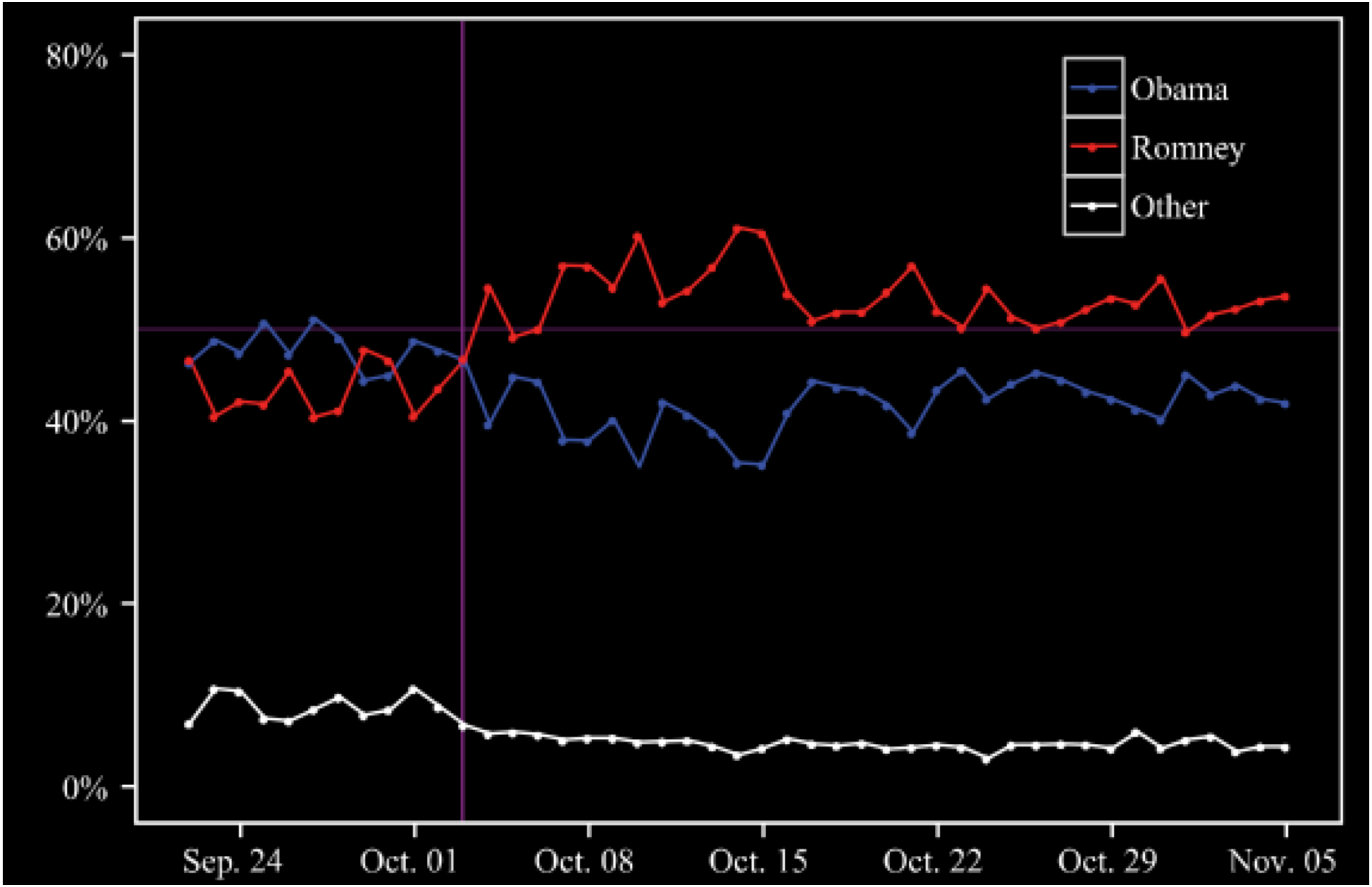
Press Conf

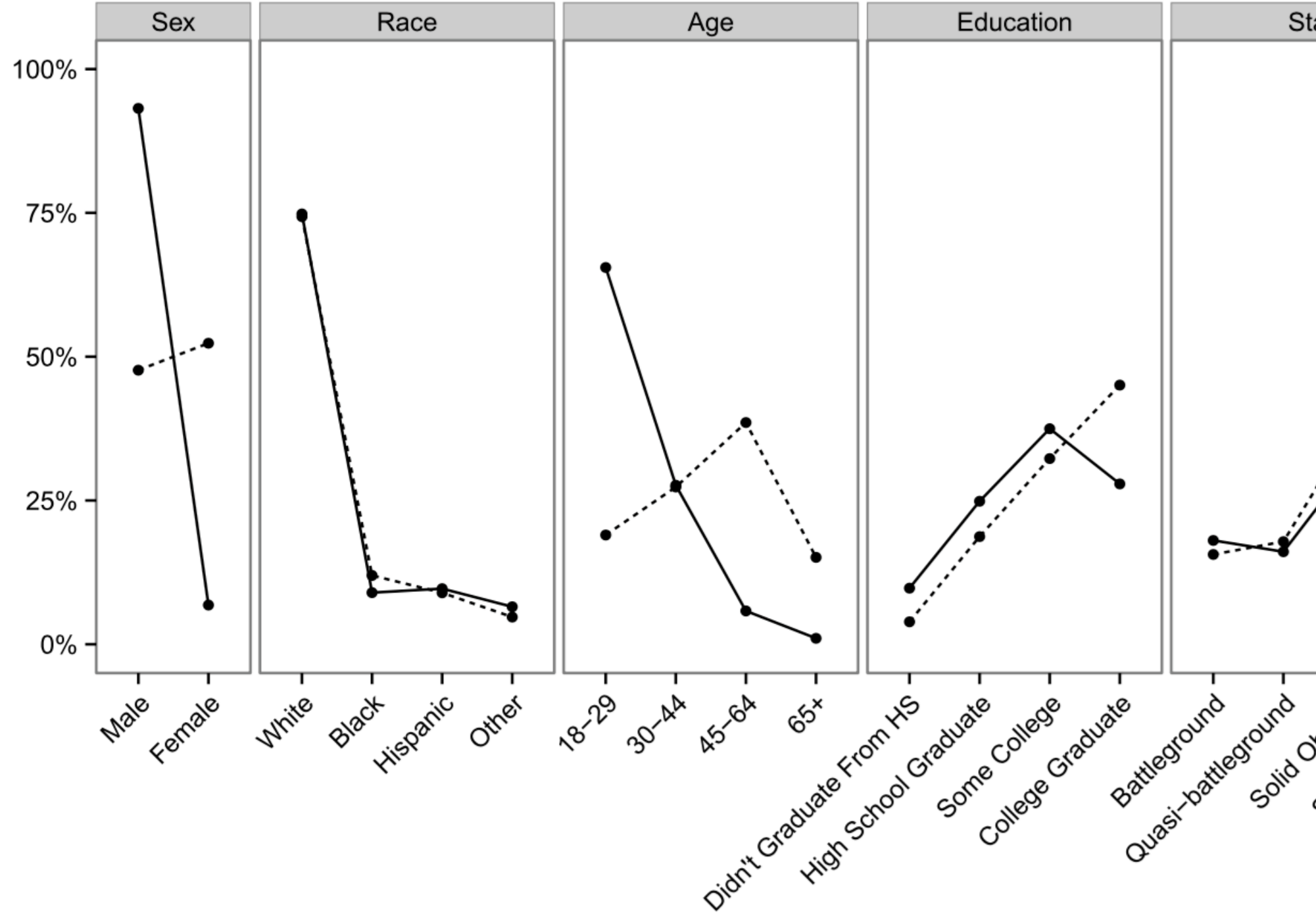
MLB.tv



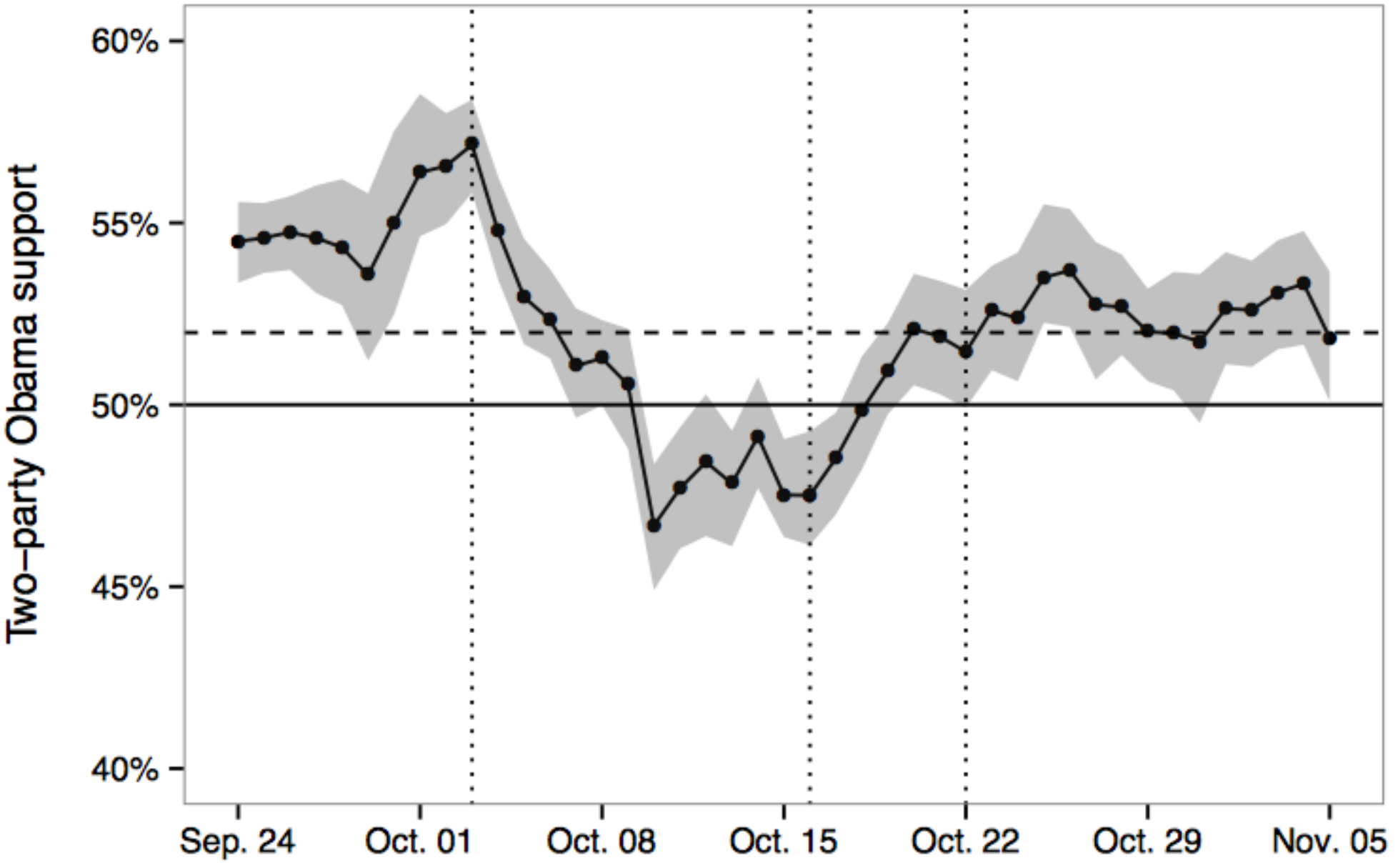
ADVERTISEMENT


If the election were held today, who would you vote for?
Barack Obama
Mitt Romney
Other
Not sure


80%
60%
40%
20%
0%
Obama
Romney
Other
Sep. 24
Oct. 01
Oct. 08
Oct. 15
Oct. 22
Oct. 29
Nov. 05


Sex
Race
Age
Education
Sta
100%
75%
50%
25%
0%
Male
Female
White
Black
Hispanic
Other
18–29
30–44
45–64
65+
Didn't Graduate From HS
High School Graduate
Some College
College Graduate
Battleground
Quasi−battleground
Solid O

Xbox estimates, adjusting for demographics:

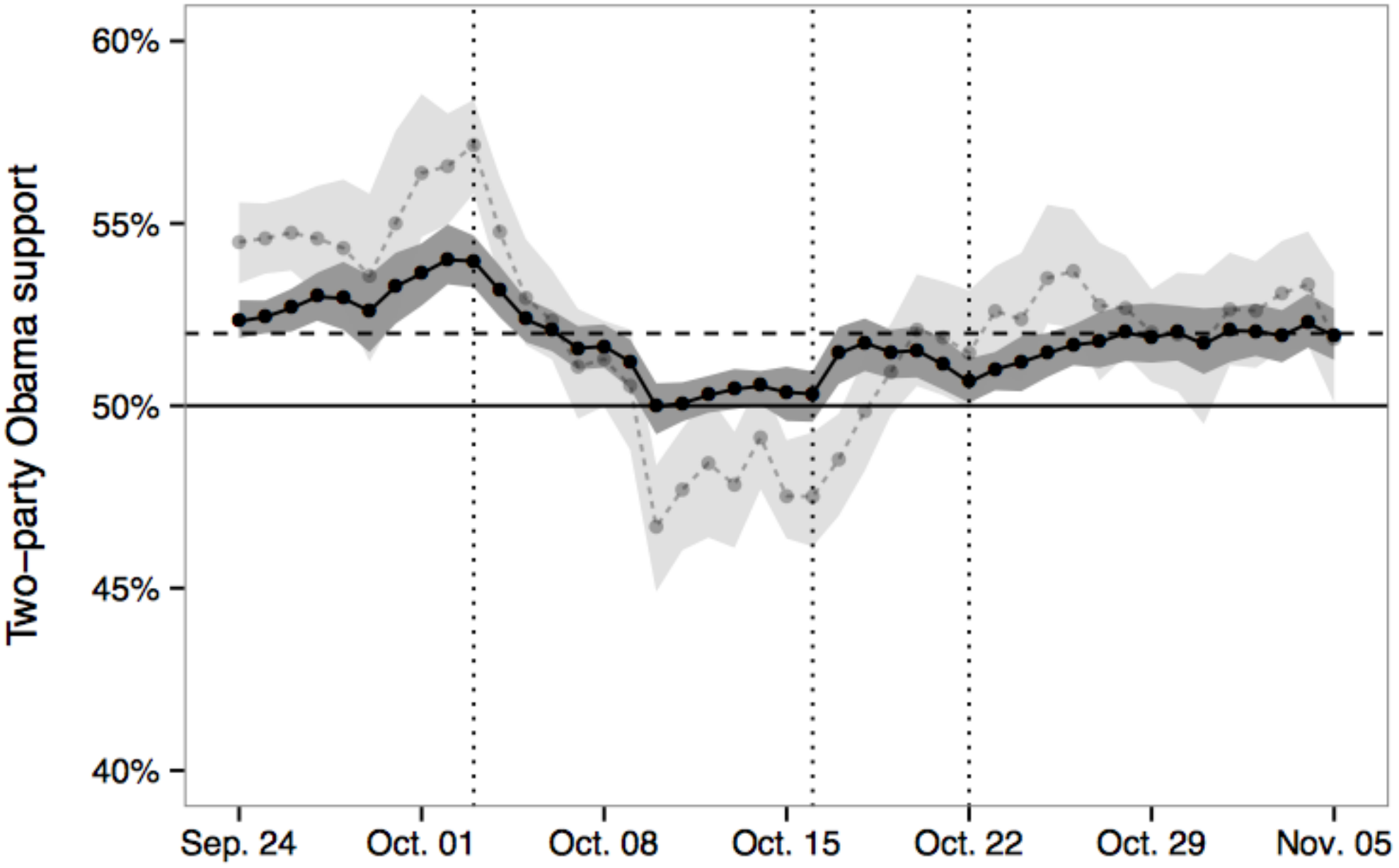


The Presidential
2012 DEBATES
#nbcpolitics

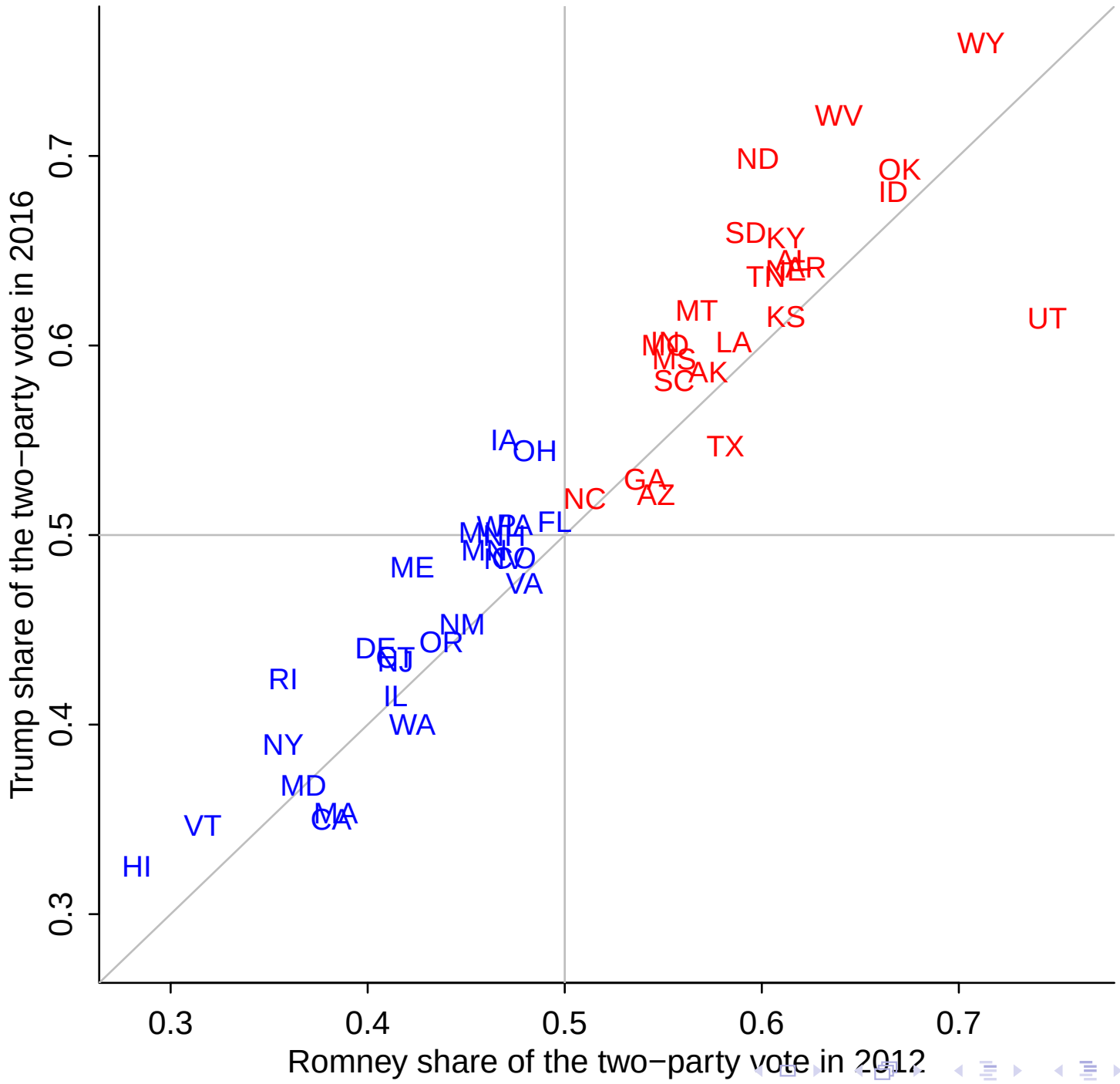
The experts speak!

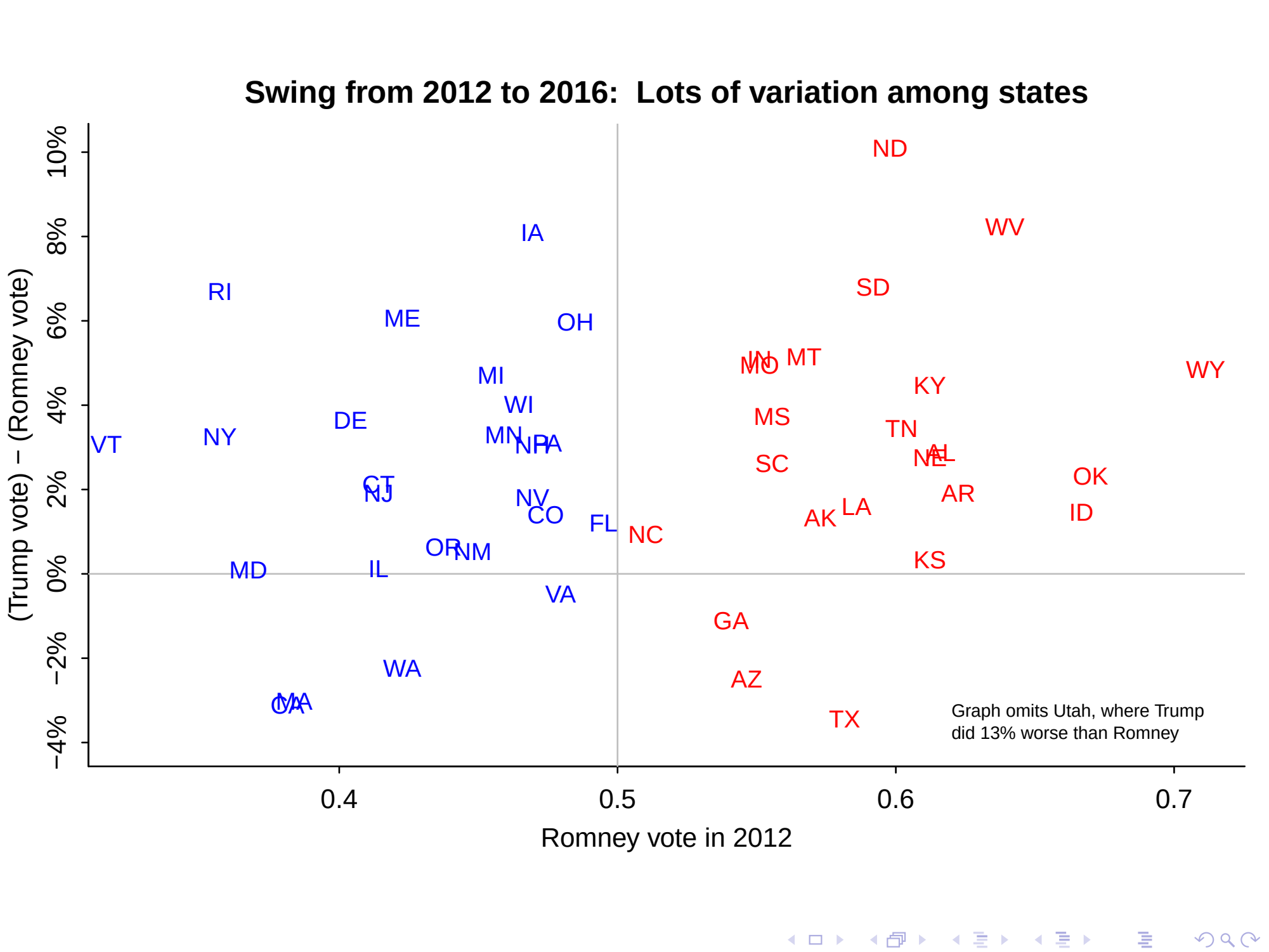
- Karl Rove, *Wall Street Journal*, 7 Oct: "Mr. Romney's bounce is significant."
- Nate Silver, *New York Times*, 6 Oct: "Mr. Romney has not only improved his own standing but also taken voters away from Mr. Obama's column."

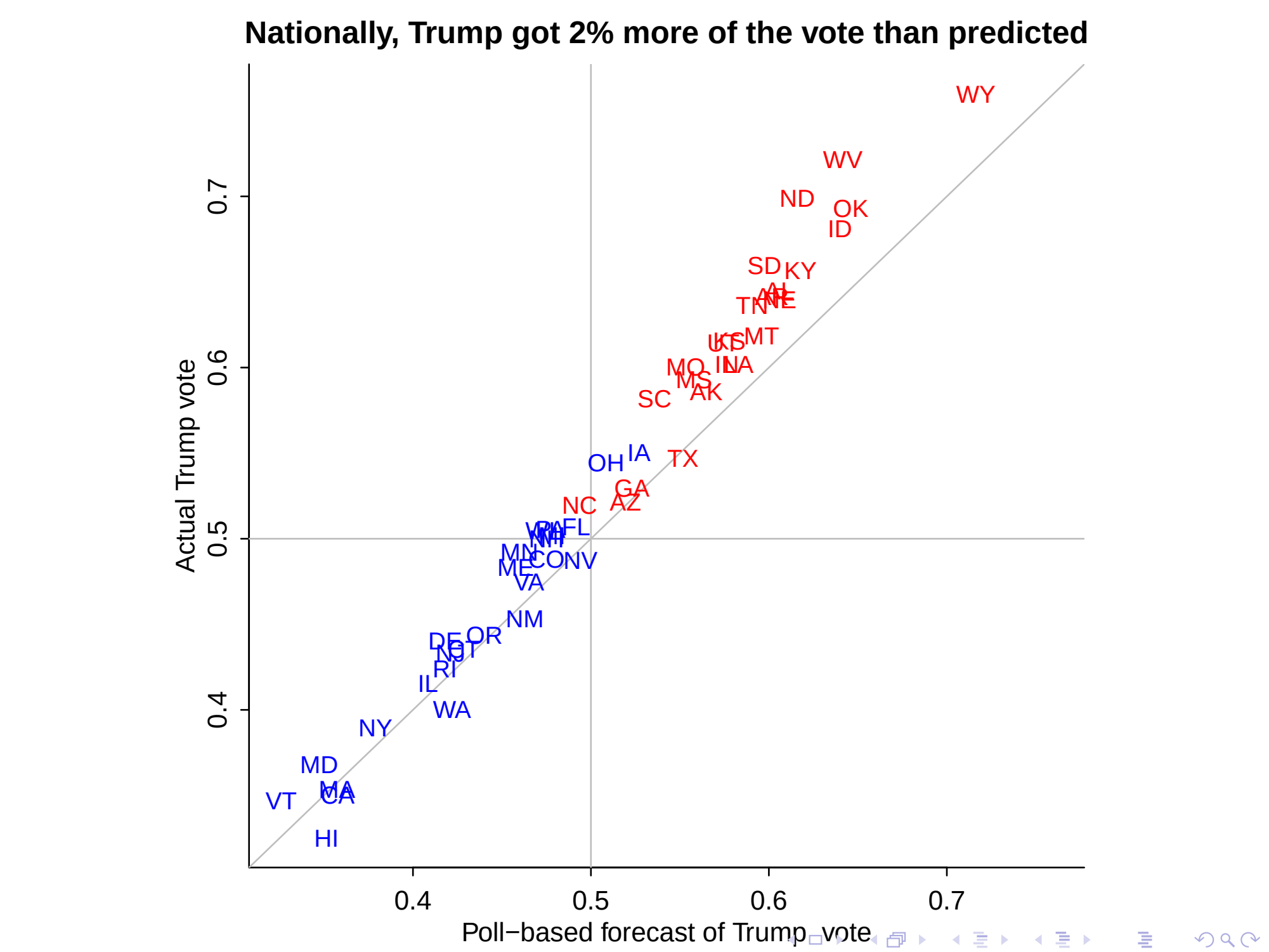
Xbox estimates, adjusting for demographics and partisanship:

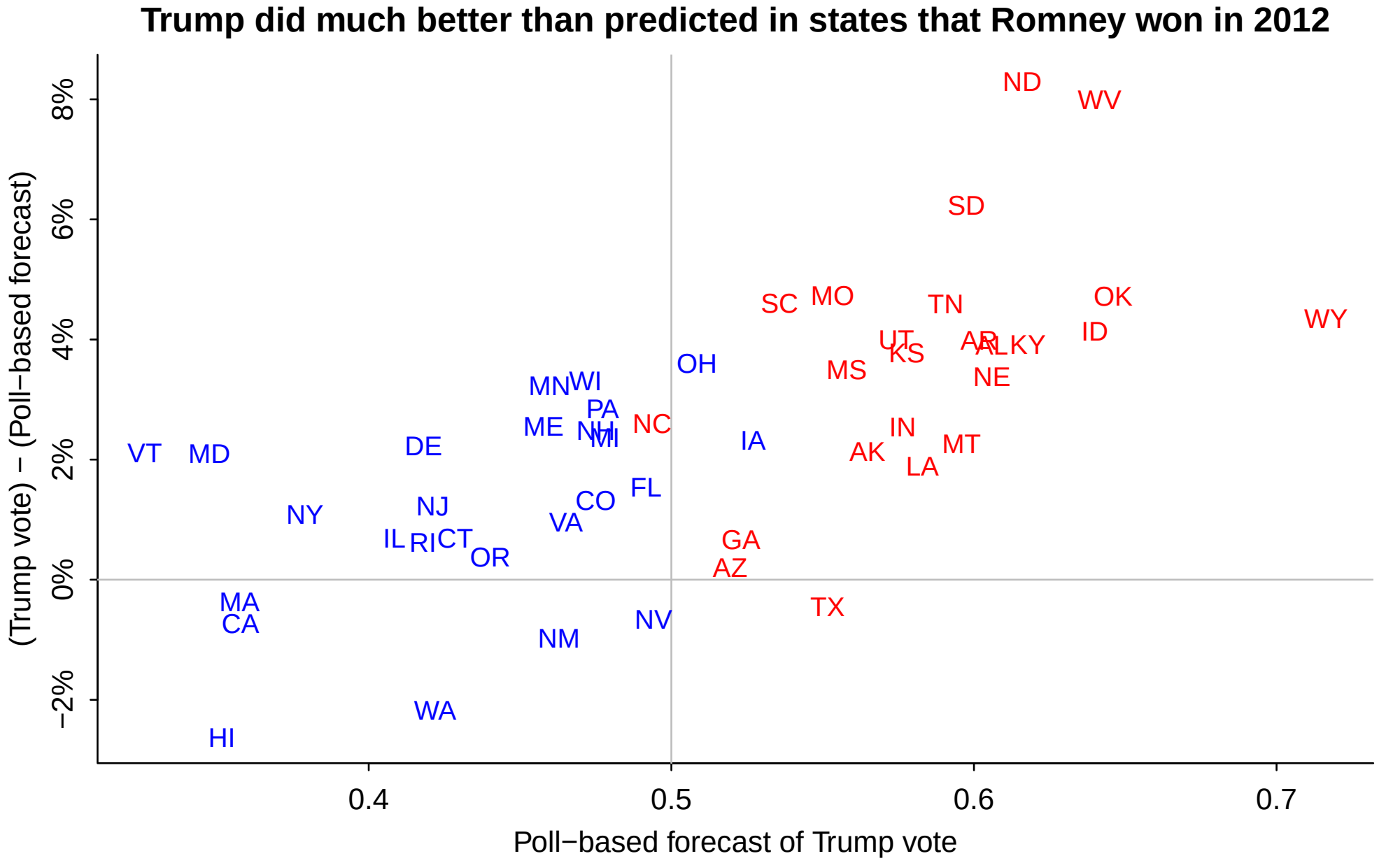


Why multilevel regression?


Nationally, Trump got 2% more of the vote than Romney
Trump share of the two−party vote in 2016
Romney share of the two−party vote in 2012
0.3
0.4
0.5
0.6
0.7
WY
WV
ND
OK
ID
SD
KY
TN
MT
KS
UT
LA
SC
AK
TX
IA
OH
GA
AZ
NC
FL
ME
VA
NM
OR
DE
NJ
RI
IL
WA
NY
MD
VT
HI


Swing from 2012 to 2016: Lots of variation among states
(Trump vote) − (Romney vote)
10%
8%
6%
4%
2%
0%
−2%
−4%
0.4
0.5
0.6
0.7
Romney vote in 2012
ND
WV
SD
WY
MT
KY
MS
TN
SC
OK
AR
ID
LA
AK
KS
NC
FL
NV
CO
IA
RI
ME
OH
MI
WI
DE
VT
NY
MD
IL
OR
NM
VA
GA
WA
AZ
TX
Graph omits Utah, where Trump
did 13% worse than Romney


Nationally, Trump got 2% more of the vote than predicted
Actual Trump vote
Poll−based forecast of Trump vote
0.4
0.5
0.6
0.7
WY
WV
ND
OK
ID
SD
KY
TN
MT
MO
IN
MS
AK
SC
OH
IA
TX
GA
NC
AZ
FL
CO
NV
VA
NM
OR
RI
IL
WA
NY
MD
VT
CA
HI


Trump did much better than predicted in states that Romney won in 2012
(Trump vote) − (Poll−based forecast)
8%
6%
4%
2%
0%
−2%
0.4
0.5
0.6
0.7
Poll−based forecast of Trump vote
ND
WV
SD
SC
MO
TN
OK
WY
ID
UT
AR
KY
KS
OH
MS
NE
MN
WI
PA
ME
NH
NC
IN
VT
MD
DE
IA
AK
MT
LA
FL
NY
NJ
CO
VA
IL
RI
CT
OR
GA
AZ
MA
CA
TX
NV
NM
WA
HI

Open problems in MRP

- Deep interactions
- Non-census variables
- Survey weights
- Cluster sampling
- Estimating regression coefficients
- Building trust in results