Microsoft Research

A combinatorial prediction market for the U.S. Elections

Miroslav Dudík

Thanks: S Lahaie, D Pennock, D Rothschild, D Osherson, A Wang, C Herget

Microsoft


ELECTION
2012

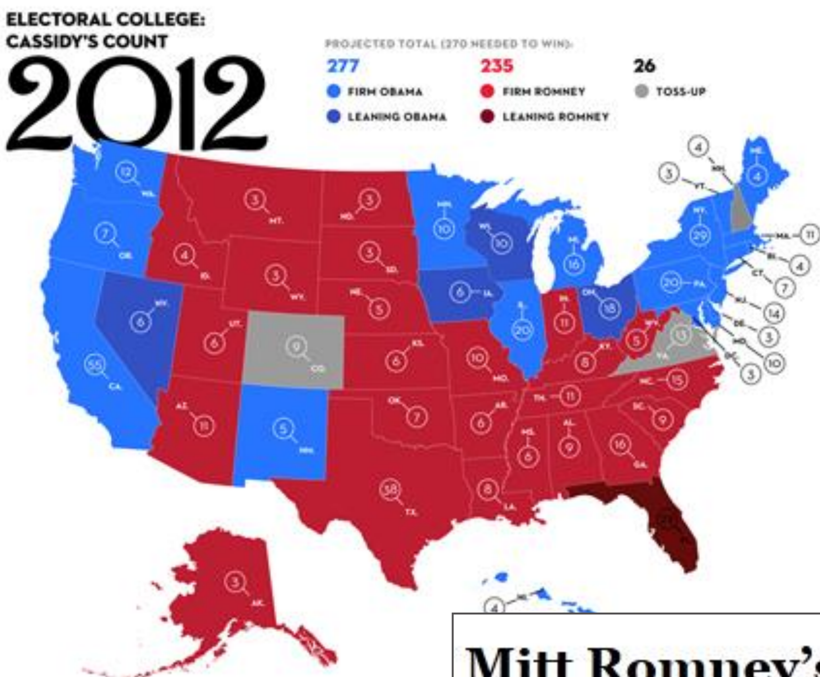

India
Elections
2014

OCTOBER 28, 2012

CASSIDY'S COUNT: CAN ROMNEY WIN WITHOUT OHIO?

POSTED BY JOHN CASSIDY

Share 244 Tweet 81 +1 PRINT + MORE 17 COMMENTS



Mitt Romney's road to presidency this fall looks narrow on electoral map

By Chris Cillizza, April 29, 2012

It's no secret that former Massachusetts governor Mitt Romney has a narrow path to win the presidency this fall. Nowhere is that reality more apparent than when examining the electoral map on which Romney and President Obama will battle in November.

A detailed analysis of Romney's various paths to the 270 electoral votes he would need to claim the presidency suggests he has a ceiling of

You're watching...

What are the Must-Win States for Romney?

Rasmussen Reports President Scott Rasmussen on the Presidential election.

OCTOBER 28, 2012

CASSIDY'S COUNT: CAN ROMNEY WIN WITHOUT OHIO?

POSTED BY JOHN CASSIDY

Share 244 Tweet 81 +1 PRI

Ohio: Romney 49%, Obama 49% (Romney Must Win VA, FL, And Either OH/WI)

RasmussenReports ^ | November 05, 2012

Posted on Tuesday, November 06, 2012 1:37:02 AM by Steelfish

FiveThirtyEight

Nate Silver's Political Calculus

September 12, 2012, 8:13 am | 181 Comments

Why Romney is losing must-win Ohio

By Peter Hamby, CNN Political Reporter

updated 5:15 PM EDT, Wed September 26, 2012

POLITICS

DEBATE BUMP PUTS ROMNEY WITHIN A POINT OF OBAMA IN MUST-WIN OHIO

Romney, Obama hit must-win states

One day before Election Day campaigns se

Jump to video

Virginia remains key to the roa

Below: Video Discuss Related

Mitt Romney's road to presidency this fall looks narrow on electoral map

By Chris Cillizza, April 29, 2012

It's no secret that former Massachusetts governor Mitt Romney has a narrow path to win the presidency this fall. Nowhere is that reality

the presidency suggests he has a ceiling of

Romney treating Ohio as a must-win state

By Chris Moody, Yahoo! News | The Ticket – Tue, Oct 16, 2012

You're watching...

What are the Must-Win States for Romney

Rasmussen Reports President Scott Rasmussen on the Presidential election.

OCTOBER 28, 2012

CASSIDY'S COUNT: CAN ROMNEY WIN WITHOUT OHIO?

POSTED BY JOHN CASSIDY

Ohio: Romney 49%, Obama 49% (Romney Must Win VA, FL, And Either OH/WI)

RasmussenReports ^ | November 05, 2012

Posted on Tuesday, November 06, 2012 1:37:02 AM by Steelfish

Total 2012 election spending: $7 billion

By *Jake Harper* | *Jan 31 2013* | *11:47 a.m.*

A new estimate from the Federal Election Commission puts total spending for the 2012 election at more than $7 billion -- $1 billion more than previously thought.

New FEC Chair Ellen Weintraub unveiled the latest estimate of the 2012 campaign's record-shattering cost at the agency's first open meeting of 2013, one that saw the departure of Cynthia Bauerly, one of the three Democratic commissioners. Though campaign spending was expected to break records after the Supreme Court's 2010 Citizens United decision that opened the door for unlimited contributions, the latest FEC estimate exceeds earlier expectations.

You're watching...

What are the Must-Win States for Romney

Rasmussen Reports President Scott Rasmussen on the Presidential election.

By Chris Moody, Yahoo! News | The Ticket – Tue, Oct 16, 2012

Polling

accurate, but costly

limited range of questions

limited timeliness

Polling

accurate, but costly
limited range of questions
limited timeliness

Prediction markets

accurate and cheap (after fixed cost)
broad range of questions
good timeliness

Outline

Prediction markets:
Setting and challenges

Addressing the challenges:
constraint generation

Empirical evaluation:
U.S. Elections 2008

Field experiment:
U.S. Elections 2012

Security

= proposition which becomes **true** or **false** at some point in future

"Romney will win Florida in Elections 2012"

Security

= proposition which becomes
true or **false** at some point in future

“Romney will win Florida
in Elections 2012”

Traders buy *shares* for some price: $0.45 per share

For each *share* of a security receive:

$1 if **true**

$0 if **false**

Market implementation: (automated) market maker

market maker
sets prices

if more shares bought,
price increases

the price equals the
consensus probability
of the event

Combinatorial securities:
more information

payoff is a function of common variables
e.g., **50 states** elect Obama or Romney

Combinatorial securities:
more information

Obama to lose FL, but win election

Obama to win >8 of 10 Northeastern states

Industry standard: ignore relationships

Treat them as independent markets:

Las Vegas	sports betting
Kentucky	horse racing
Wall Street	stock options
Betfair	political betting

Industry standard:
ignore relationships

Treat them as independent markets:

Las Vegas	sports betting
Kentucky	horse racing
Wall Street	stock options
Betfair	political betting

Problem:

arbitrage opportunities

Arbitrage

trading with guaranteed profits

Arbitrage

trading with guaranteed profits

receive $1 if true

Arbitrage

trading with guaranteed profits

price $0.40

price $0.50

Arbitrage

trading with guaranteed profits possible if prices *incoherent*

prices cannot be realized as probabilities

price \$0.40

price \$0.50

Arbitrage

trading with guaranteed profits possible if prices *incoherent*

prices cannot be realized as probabilities

price $0.40

price $0.50

Pricing without arbitrage: #P-hard

Industry standard = Ignore arbitrage

Arbitrage

trading with guaranteed profits
possible if prices *incoherent*

prices cannot be realized as probabilities

price $0.40

price $0.50

Pricing without arbitrage: #P-hard

Industry standard = Ignore arbitrage

- traders rewarded for computation instead of information
- poor information sharing

Our approach:
partial arbitrage removal

Separate *pricing* (must be fast)
and *information propagation*

- *fast pricing* in *independent markets*
  for tractably small groups of securities

- *in parallel:* constraint generation
  to *find* and *remove arbitrage*

Embedded in convex optimization
(with many nice properties).

Cost-based pricing

(Chen and Pennock 2007)

Setup:

- n securities
- $C : \mathbb{R}^n \to \mathbb{R}$ convex cost function
- $\boldsymbol{q} \in \mathbb{R}^n$ market state = #shares sold

Cost-based pricing

(Chen and Pennock 2007)

Setup:

- n securities
- $C: \mathbb{R}^n \to \mathbb{R}$ convex cost function
- $\boldsymbol{q} \in \mathbb{R}^n$ market state = #shares sold

$$\boldsymbol{q} = (\quad 100, \quad 400)$$

Cost-based pricing

(Chen and Pennock 2007)

Setup:

n securities

$C : \mathbb{R}^n \to \mathbb{R}$ convex cost function

$\boldsymbol{q} \in \mathbb{R}^n$ market state = #shares sold

$\boldsymbol{q} = (\quad 100, \quad 400)$

Trading:

$\boldsymbol{r} \in \mathbb{R}^n$ shares bought by a trader

cost: $C(\boldsymbol{q} + \boldsymbol{r}) - C(\boldsymbol{q})$

Cost-based pricing

(Chen and Pennock 2007)

Setup:

n securities

$C: \mathbb{R}^n \to \mathbb{R}$ convex cost function

$\boldsymbol{q} \in \mathbb{R}^n$ market state = #shares sold $\qquad \boldsymbol{q} = (100, 400)$

Trading:

$\boldsymbol{r} \in \mathbb{R}^n$ shares bought by a trader $\qquad \boldsymbol{r} = (0, 2)$

cost: $C(\boldsymbol{q} + \boldsymbol{r}) - C(\boldsymbol{q})$

Cost-based pricing

(Chen and Pennock 2007)

Setup:

- n securities
- $C: \mathbb{R}^n \to \mathbb{R}$ convex cost function
- $\boldsymbol{q} \in \mathbb{R}^n$ market state = #shares sold

$$\boldsymbol{q} = (100, 400)$$

Trading:

- $\boldsymbol{r} \in \mathbb{R}^n$ shares bought by a trader
- cost: $C(\boldsymbol{q} + \boldsymbol{r}) - C(\boldsymbol{q})$
- state updated: $\boldsymbol{q}' \leftarrow \boldsymbol{q} + \boldsymbol{r}$

$$\boldsymbol{r} = (0, 2)$$

$$\boldsymbol{q}' = (100, 402)$$

Cost-based pricing

(Chen and Pennock 2007)

Setup:

- n securities
- $C: \mathbb{R}^n \to \mathbb{R}$ convex cost function
- $\boldsymbol{q} \in \mathbb{R}^n$ market state = #shares sold

$\boldsymbol{q} = (100, \; 400)$

Trading:

- $\boldsymbol{r} \in \mathbb{R}^n$ shares bought by a trader
- cost: $C(\boldsymbol{q} + \boldsymbol{r}) - C(\boldsymbol{q})$
- state updated: $\boldsymbol{q}' \leftarrow \boldsymbol{q} + \boldsymbol{r}$
- instantaneous prices: $\nabla C(\boldsymbol{q})$

$\boldsymbol{r} = (0, \; 2)$

$\boldsymbol{q}' = (100, \; 402)$

$\nabla C(\boldsymbol{q}) = (\$0.70, \; \$0.75)$

Cost-based pricing

(Chen and Pennock 2007)

Setup:

- n securities
- $C\colon \mathbb{R}^n \to \mathbb{R}$ convex cost function
- $\boldsymbol{q} \in \mathbb{R}^n$ market state = #shares sold

$$\boldsymbol{q} = (100, \quad 400)$$

Trading:

- $\boldsymbol{r} \in \mathbb{R}^n$ shares bought by a trader
- cost: $C(\boldsymbol{q}+\boldsymbol{r}) - C(\boldsymbol{q})$
- state updated: $\boldsymbol{q}' \leftarrow \boldsymbol{q}+\boldsymbol{r}$
- instantaneous prices: $\nabla C(\boldsymbol{q})$

$$\boldsymbol{r} = (0, \quad 2)$$

$$\boldsymbol{q}' = (100, \quad 402)$$

$$\nabla C(\boldsymbol{q}) = (\$0.70, \quad \$0.75)$$

Can we just use existing approaches from graphical models?

MCMC—randomized, slow convergence

mean field—non-convex

belief propagation—lack of convergence

Can we just use existing approaches from graphical models?

MCMC—randomized, slow convergence

mean field—non-convex

belief propagation—lack of convergence

Problematic for pricing:

poor convergence → volatility

non-determinism → distorted incentives and user experience

Our approach

implement a coherent pricing scheme
on small groups of securities; e.g.,

number of shares
bought so far

priced $\dfrac{e^{q_1}}{e^{q_1} + e^{q_2}}$

priced $\dfrac{e^{q_2}}{e^{q_1} + e^{q_2}}$

Our approach

implement a coherent pricing scheme
on small groups of securities; e.g.,

priced $\dfrac{e^{q_1}}{e^{q_1} + e^{q_2}}$

priced $\dfrac{e^{q_2}}{e^{q_1} + e^{q_2}}$

Our approach

implement a coherent pricing scheme
on small groups of securities; e.g.,

priced $\frac{e^{q_1}}{e^{q_1}+e^{q_2}}$

priced $\frac{e^{q_2}}{e^{q_1}+e^{q_2}}$

detect incoherence *between groups*

act as an arbitrageur to restore coherence

Our approach

implement a coherent pricing scheme
on small groups of securities; e.g.,

priced $\frac{e^{q_1}}{e^{q_1}+e^{q_2}}$

priced $\frac{e^{q_2}}{e^{q_1}+e^{q_2}}$

detect incoherence *between groups*

act as an arbitrageur to restore coherence

caveat:

- difficult to detect incoherence in general
- we detect only a subset of violations

For U.S. Elections:
conjunction market

create 50 states (groups of size 2)

create all pairs of states (groups of size 4)

for conjunctions of 3 or more,
group with opposite disjunction:

$A \land B \land C$ with $\bar{A} \lor \bar{B} \lor \bar{C}$ (groups of size 2)

For U.S. Elections:
conjunction market

create 50 states (groups of size 2)

create all pairs of states (groups of size 4)

for conjunctions of 3 or more,
group with opposite disjunction:

$A \wedge B \wedge C$ with $\bar{A} \vee \bar{B} \vee \bar{C}$ (groups of size 2)

each group is independent market:
fast pricing

in parallel:
generate, **find**, and **fix** constraints
(via coordinate descent)

Local coherence

Pairs:

$$P[A \wedge B] + P[A \wedge \bar{B}] = P[A]$$

Larger conjunctions:

$$P[A_1 \wedge A_2 \wedge \cdots \wedge A_m] \leq P[A_i]$$

Clique constraints

For a disjunction $A_1 \vee \cdots \vee A_m$,
pick a subset $A_{i_1} \vee \cdots \vee A_{i_k}$

$$P[A_1 \vee \cdots \vee A_m] \geq P\left[A_{i_1} \vee \cdots \vee A_{i_k}\right]$$

Clique constraints

For a disjunction $A_1 \vee \cdots \vee A_m$,

pick a subset $A_{i_1} \vee \cdots \vee A_{i_k}$

$$P[A_1 \vee \cdots \vee A_m] \geq P\left[A_{i_1} \vee \cdots \vee A_{i_k}\right]$$

$$\geq \sum_{j=1}^{k} P\left[A_{i_j}\right] - \sum_{1 \leq j < l \leq k} P\left[A_{i_j} \wedge A_{i_l}\right]$$

Clique constraints

For a disjunction $A_1 \vee \cdots \vee A_m$,

pick a subset $A_{i_1} \vee \cdots \vee A_{i_k}$

$$P[A_1 \vee \cdots \vee A_m] \geq P\left[A_{i_1} \vee \cdots \vee A_{i_k}\right]$$
$$\geq \sum_{j=1}^{k} P\left[A_{i_j}\right] - \sum_{1 \leq j < l \leq k} P\left[A_{i_j} \wedge A_{i_l}\right]$$

#clique constraints exponential

→ find only the tightest one!

(approximate submodular maximization
via Feige et al. 2007)

Tree constraints

(Galambos and Simoneli 1996)

For a disjunction $A_1 \vee \cdots \vee A_m$,

$$P[A_1 \vee \cdots \vee A_m] \leq \sum_{i=1}^{m} P[A_i]$$

Tree constraints

(Galambos and Simoneli 1996)

For a disjunction $A_1 \vee \cdots \vee A_m$,

$$P[A_1 \vee \cdots \vee A_m] \leq \sum_{i=1}^{m} P[A_i] - \sum_{(i,j)\in T} P\left[A_i \wedge A_j\right]$$

where T is a spanning tree on nodes $1, \ldots, m$

Does it work?

Tested using a survey of Election 2008:

singletons, pairs, triples

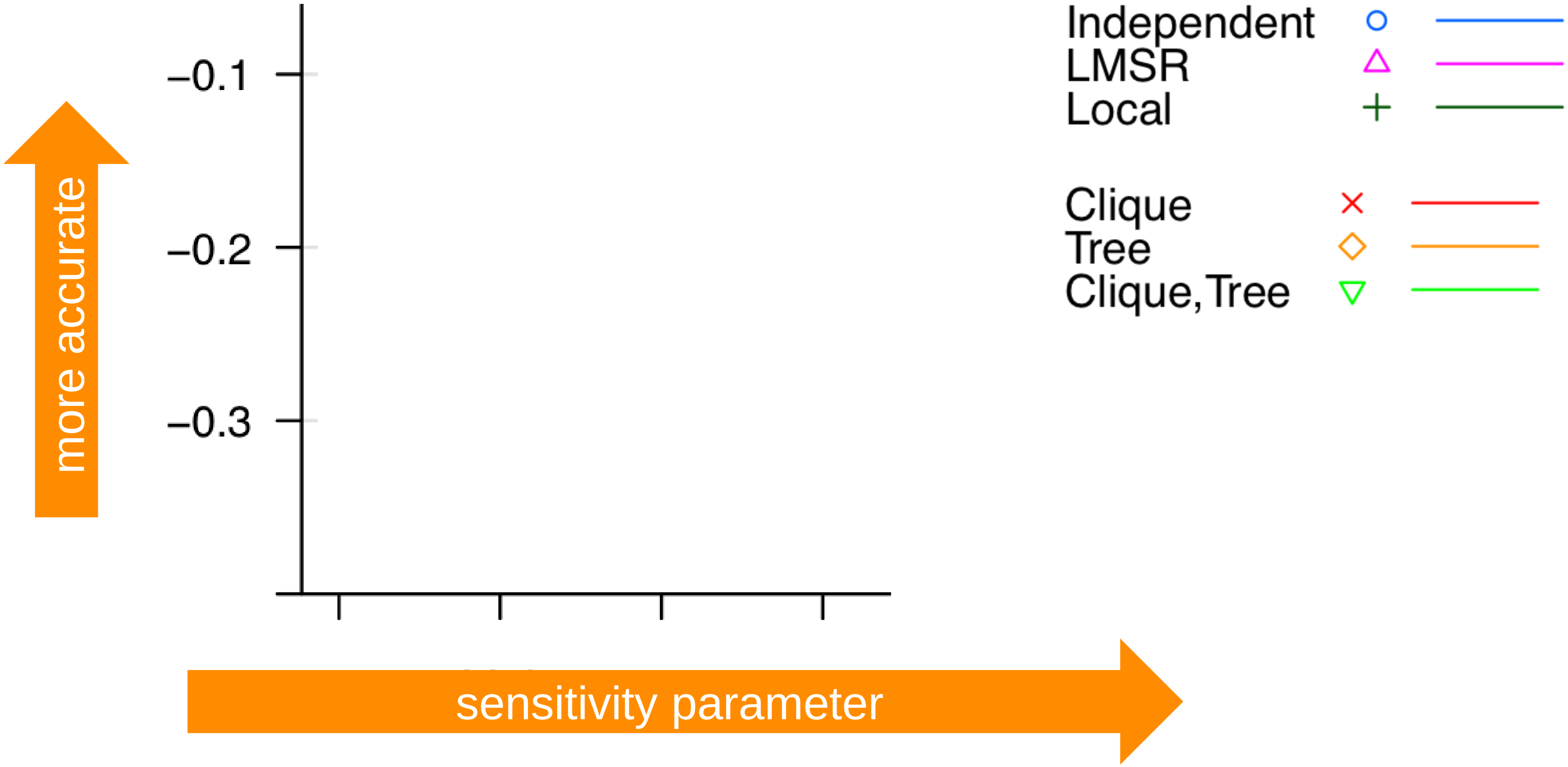
Small data set—compare with exact:

10 states, 30k trades

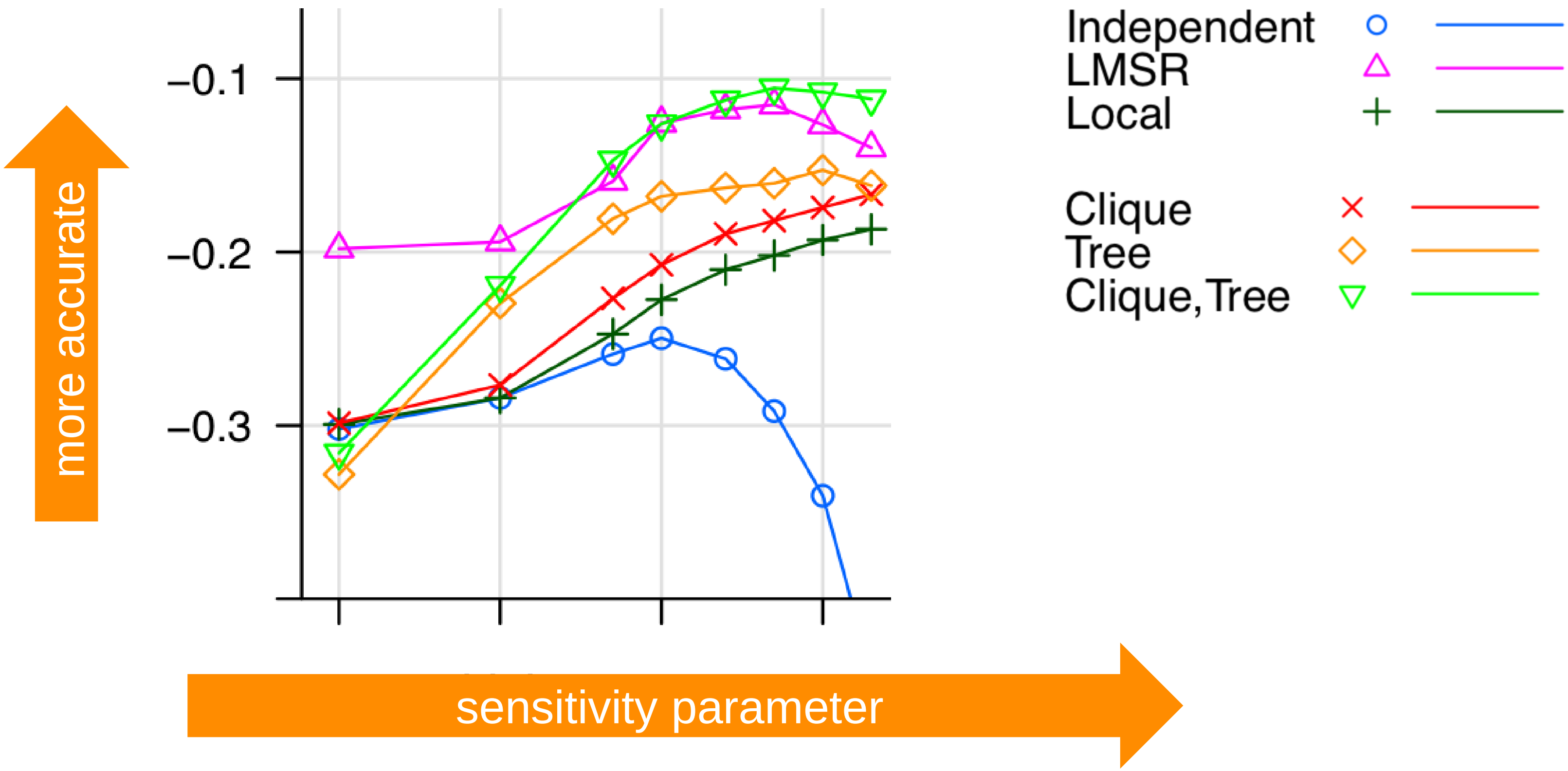
Large data set—compare with independent:

50 states, 300k trades

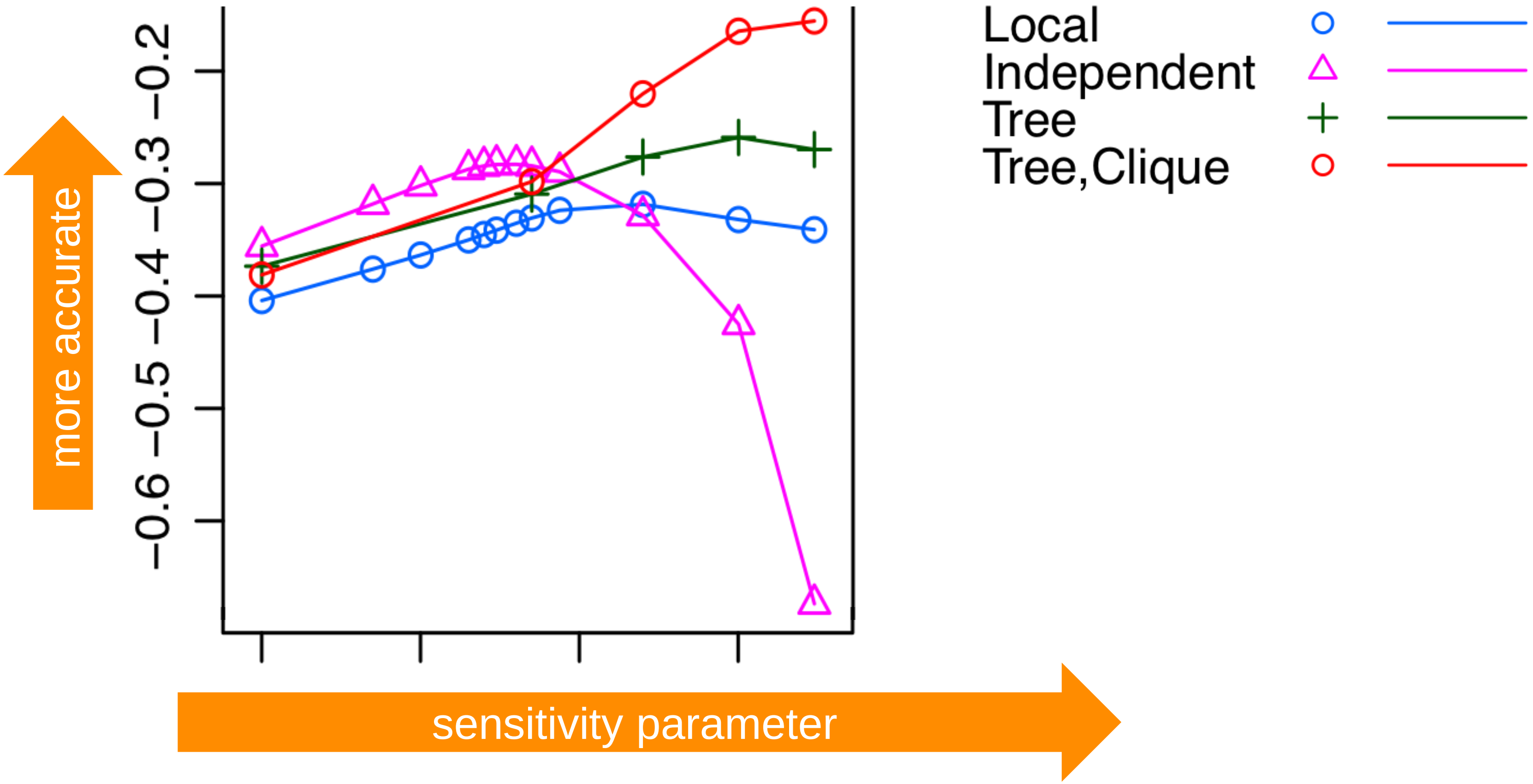
Small data set:
10 states



Small data set:
10 states



Large data set:
50 states, 300k trades



No really,
does it work?

WiseQ Game

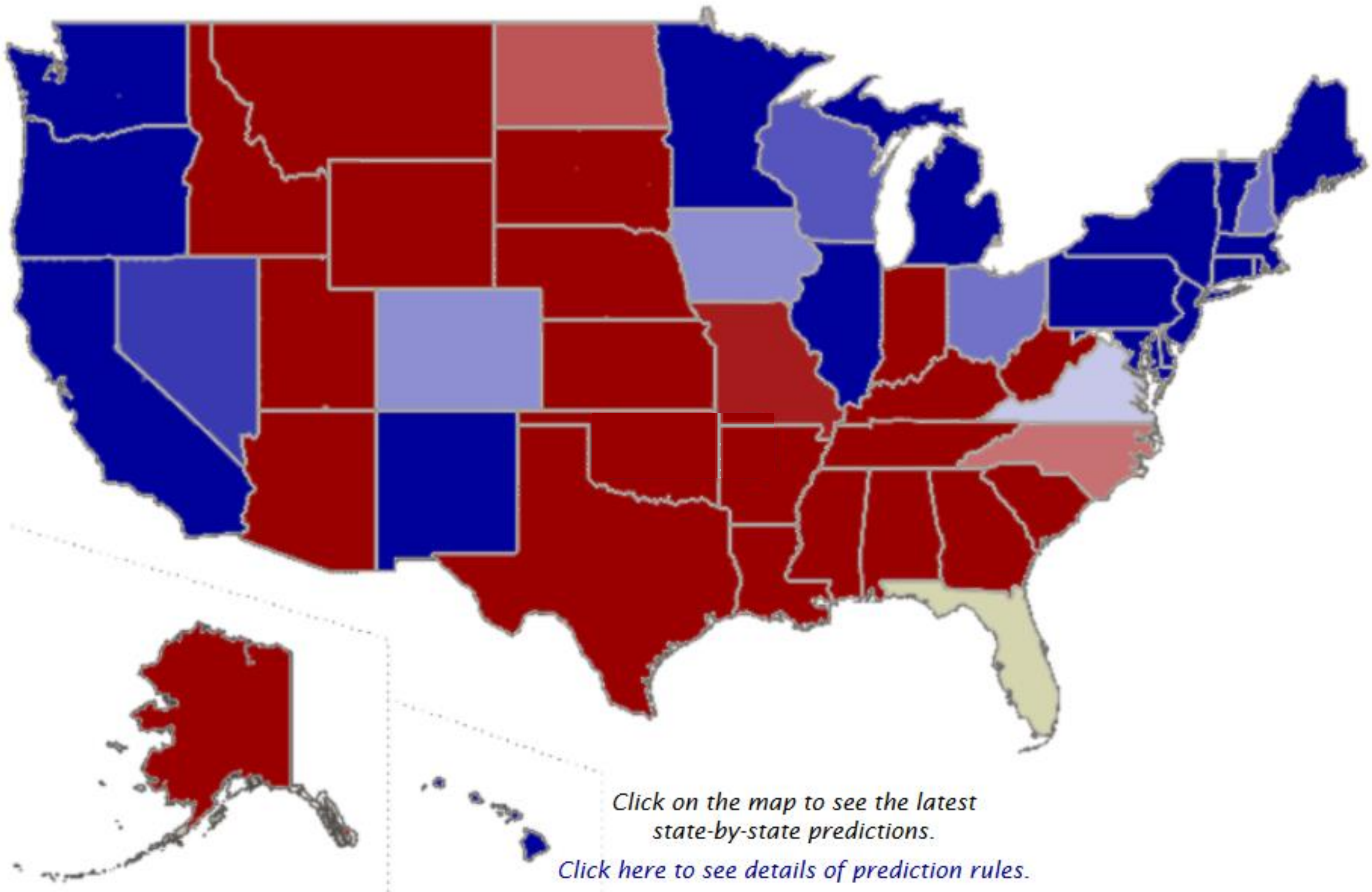
(launched September 16, 2012)

WiseQ Game - Elections 2012 *(Beta)*

MAKE A PREDICTION

President | Senate | Governors | Leaderboards | My Portfolio | My Leagues | Forum | FAQ

Welcome,
miro!
Logout



Click on the map to see the latest state-by-state predictions.

Click here to see details of prediction rules.

CURRENT ODDS

Click a state for current odds.

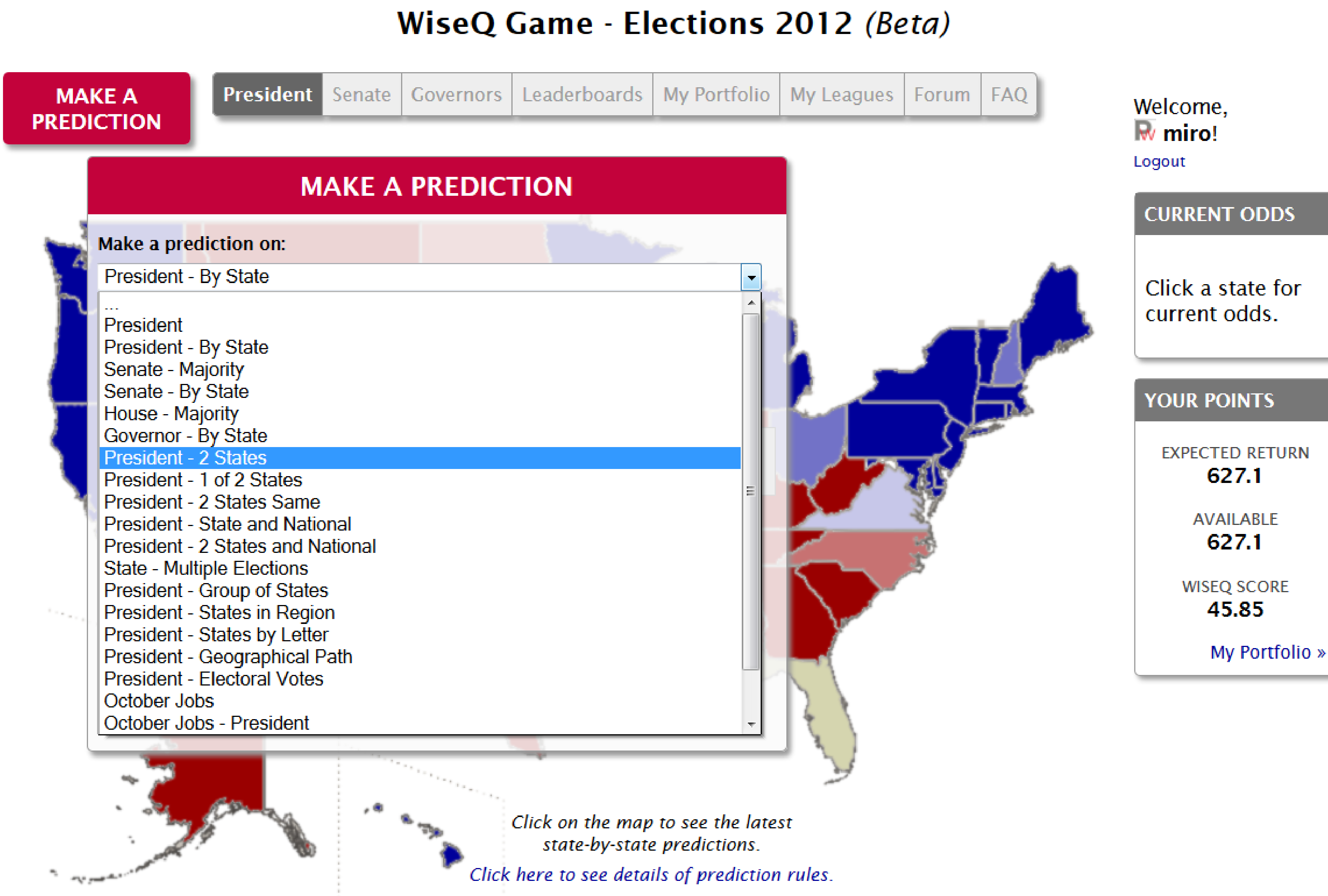
YOUR POINTS

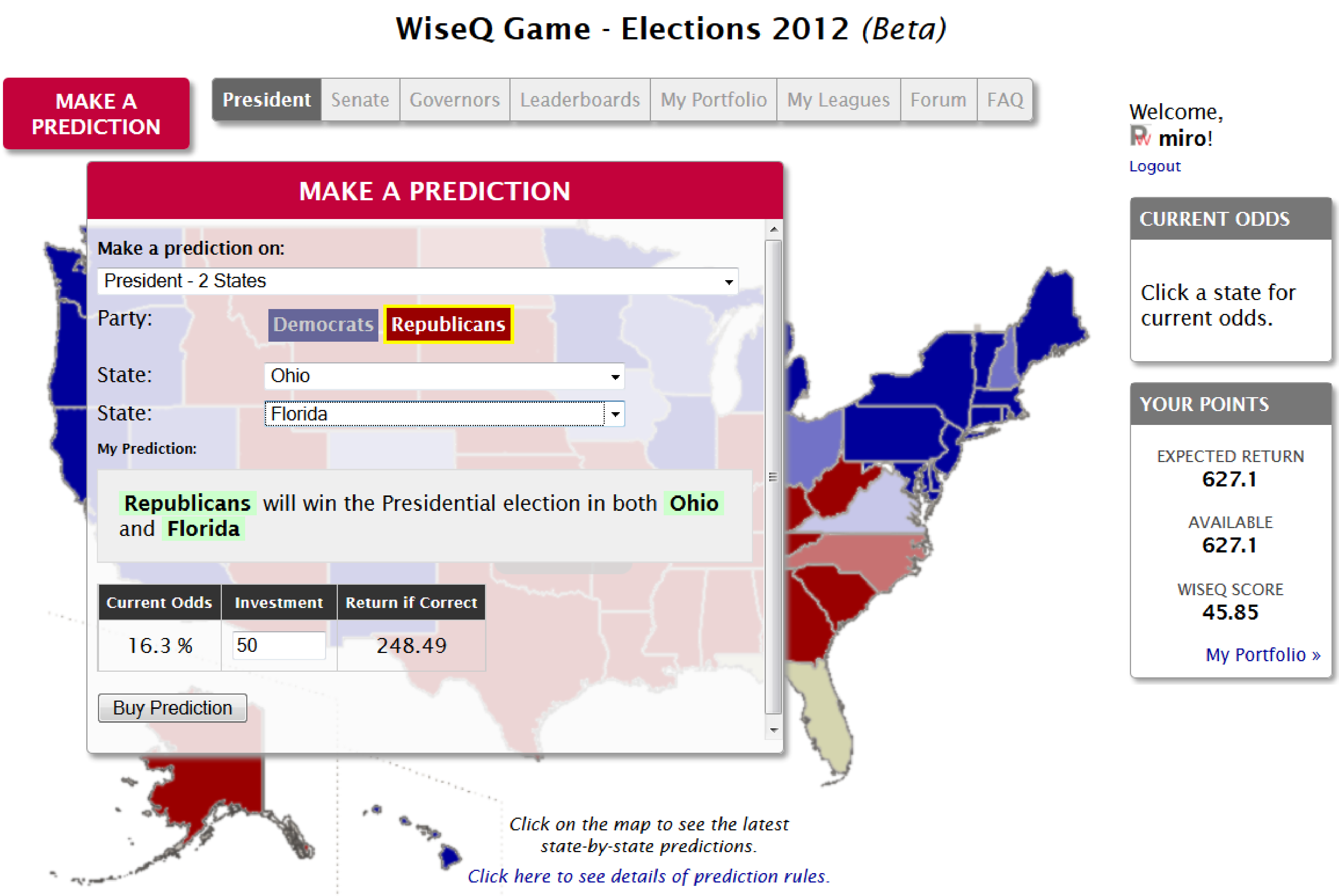
EXPECTED RETURN
627.1

AVAILABLE
627.1

WISEQ SCORE
45.85

My Portfolio »


WiseQ Game - Elections 2012 (Beta)
MAKE A PREDICTION
President
Senate
Governors
Leaderboards
My Portfolio
My Leagues
Forum
FAQ
Welcome,
miro!
Logout
MAKE A PREDICTION
Make a prediction on:
President - By State
...
President
President - By State
Senate - Majority
Senate - By State
House - Majority
Governor - By State
President - 2 States
President - 1 of 2 States
President - 2 States Same
President - State and National
President - 2 States and National
State - Multiple Elections
President - Group of States
President - States in Region
President - States by Letter
President - Geographical Path
President - Electoral Votes
October Jobs
October Jobs - President
CURRENT ODDS
Click a state for current odds.
YOUR POINTS
EXPECTED RETURN
627.1
AVAILABLE
627.1
WISEQ SCORE
45.85
My Portfolio »
Click on the map to see the latest state-by-state predictions.
Click here to see details of prediction rules.


WiseQ Game - Elections 2012 *(Beta)*
MAKE A PREDICTION
President
Senate
Governors
Leaderboards
My Portfolio
My Leagues
Forum
FAQ
Welcome,
miro!
Logout
MAKE A PREDICTION
Make a prediction on:
President - 2 States
Party:
Democrats
Republicans
State:
Ohio
State:
Florida
My Prediction:
Republicans will win the Presidential election in both Ohio and Florida
Current Odds
Investment
Return if Correct
16.3 %
50
248.49
Buy Prediction
CURRENT ODDS
Click a state for current odds.
YOUR POINTS
EXPECTED RETURN
627.1
AVAILABLE
627.1
WISEQ SCORE
45.85
My Portfolio »
Click on the map to see the latest state-by-state predictions.
Click here to see details of prediction rules.

WiseQ by numbers

437 active users

3,137 trades

514 distinct bundles traded

10^{33} possible outcomes

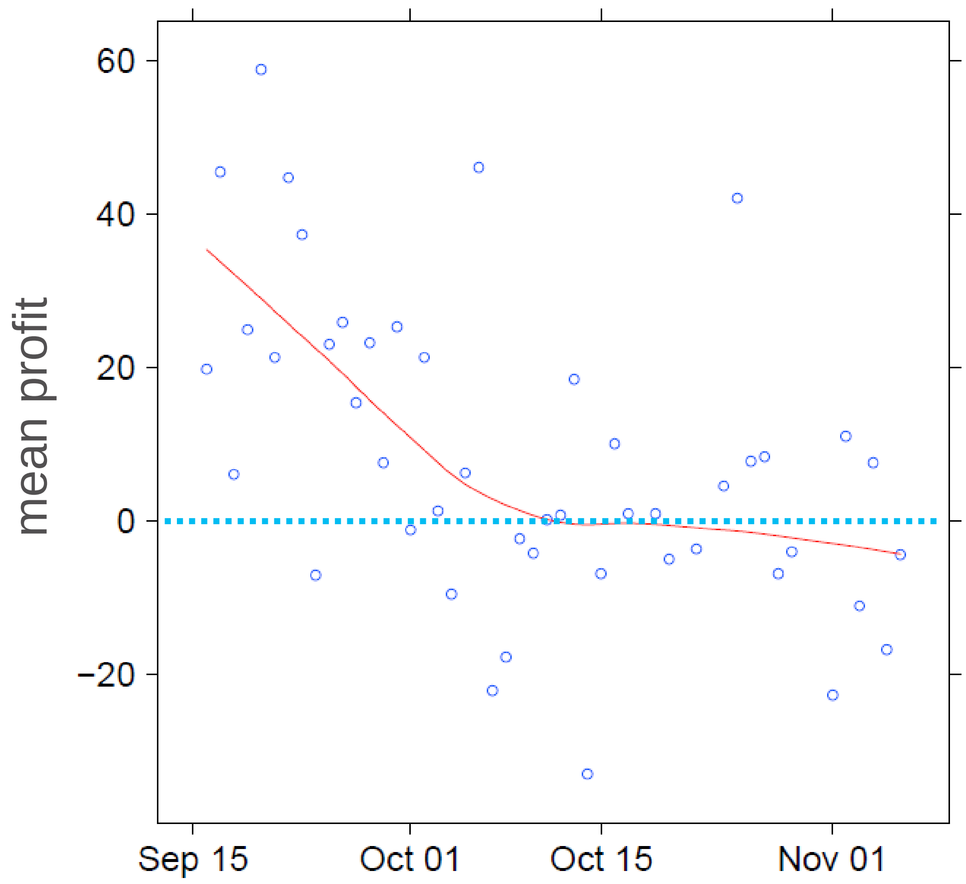
44.5 million possible bundles allowed by our menu

17,222 securities in **2,840** small markets

20,983 coherence constraints

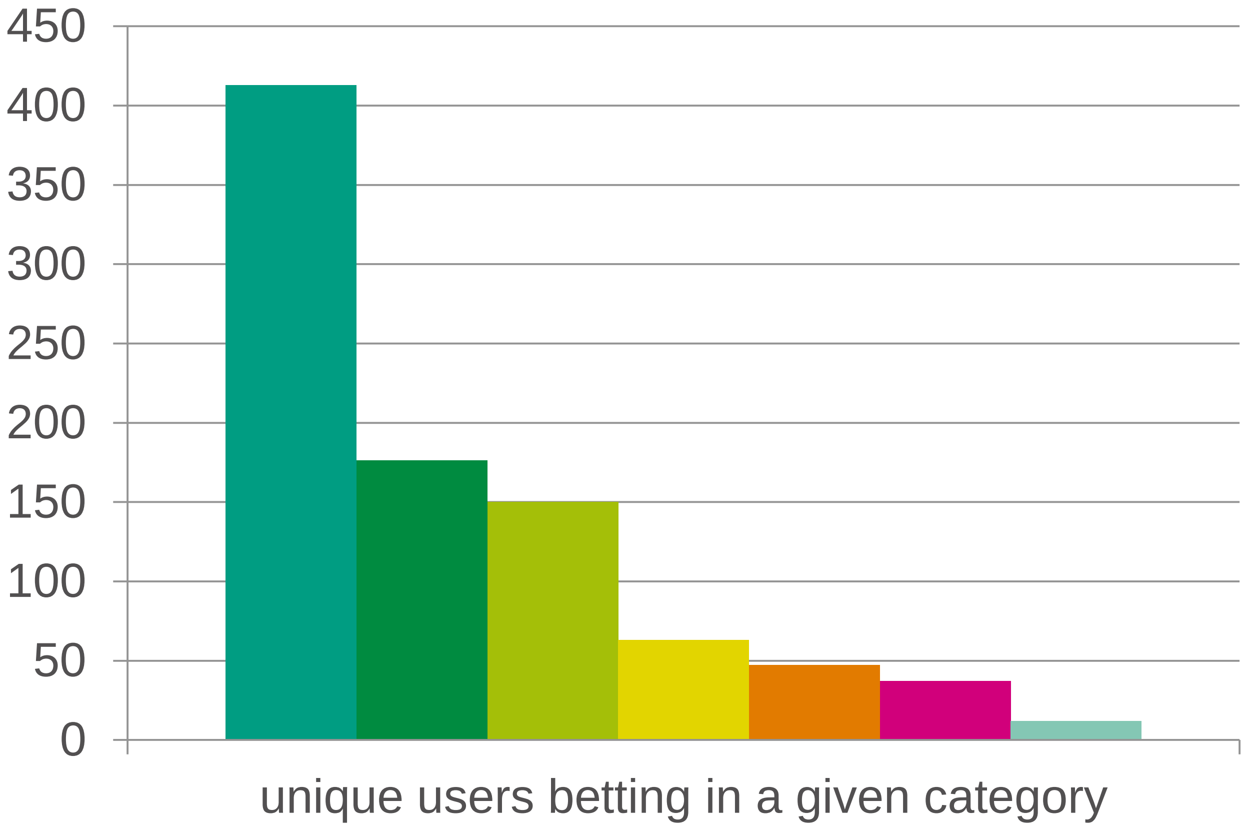
Did market absorb information from users?

Did market absorb information from users?

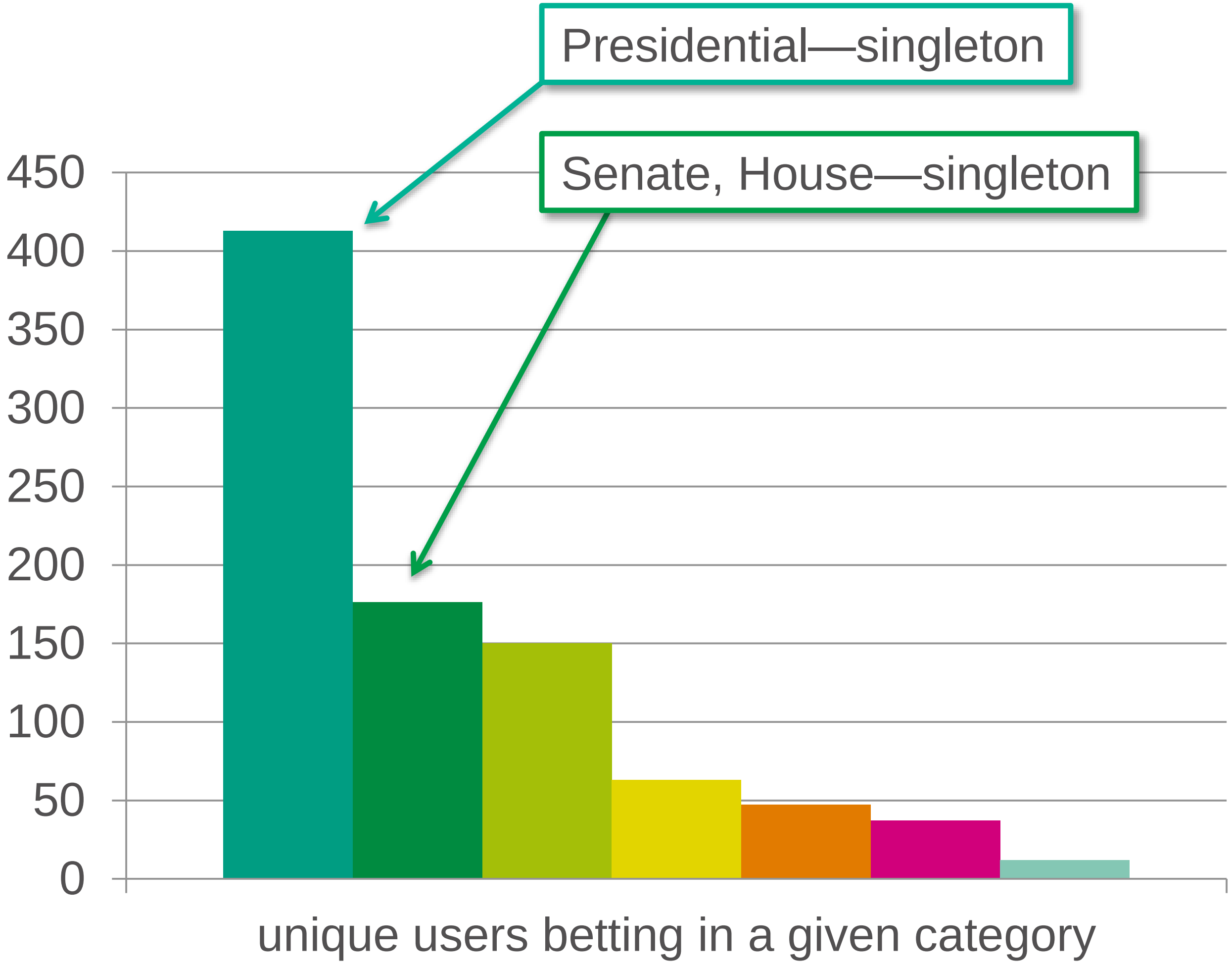


Did users place combinatorial bets?

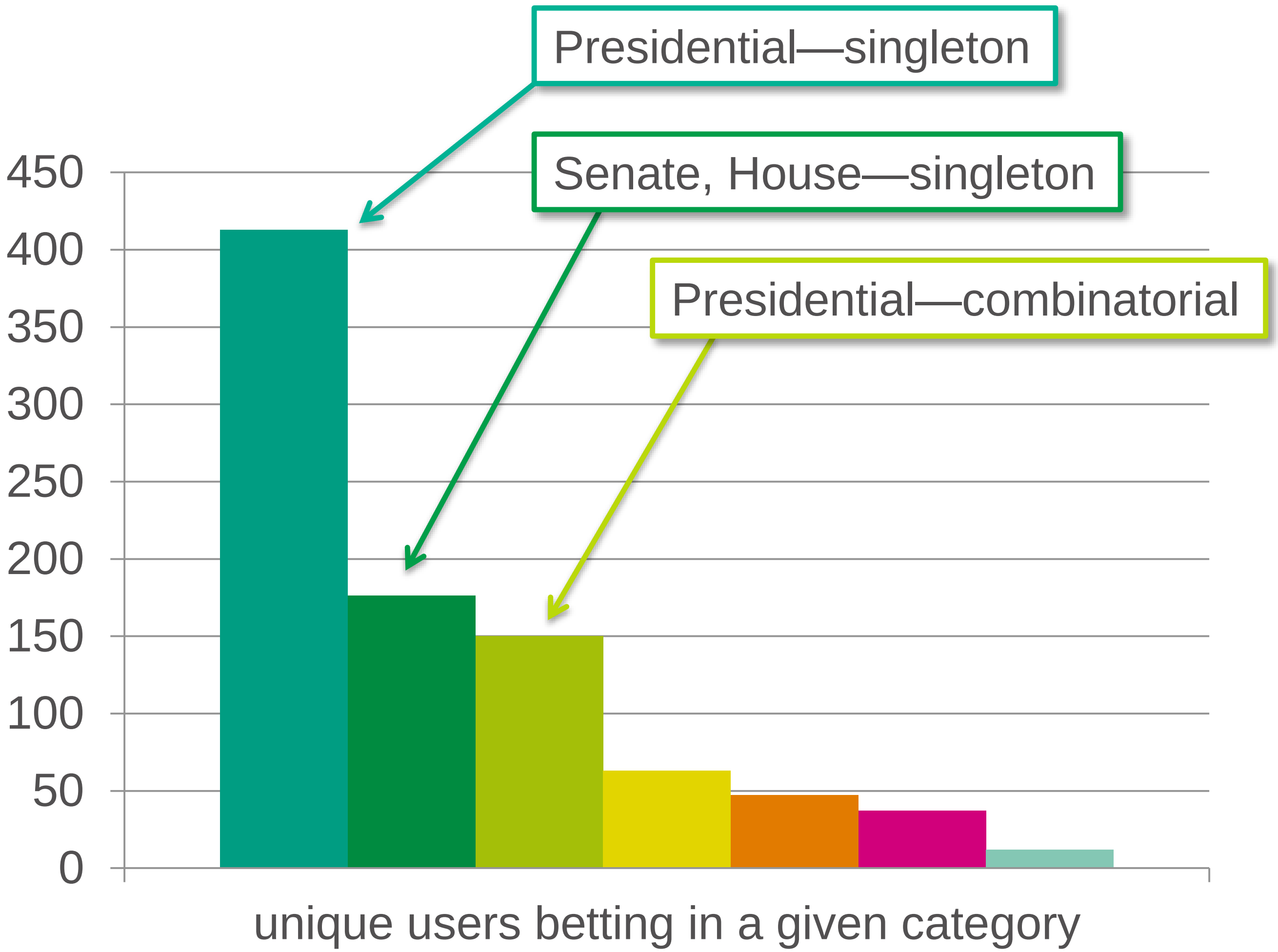
Did users place combinatorial bets?



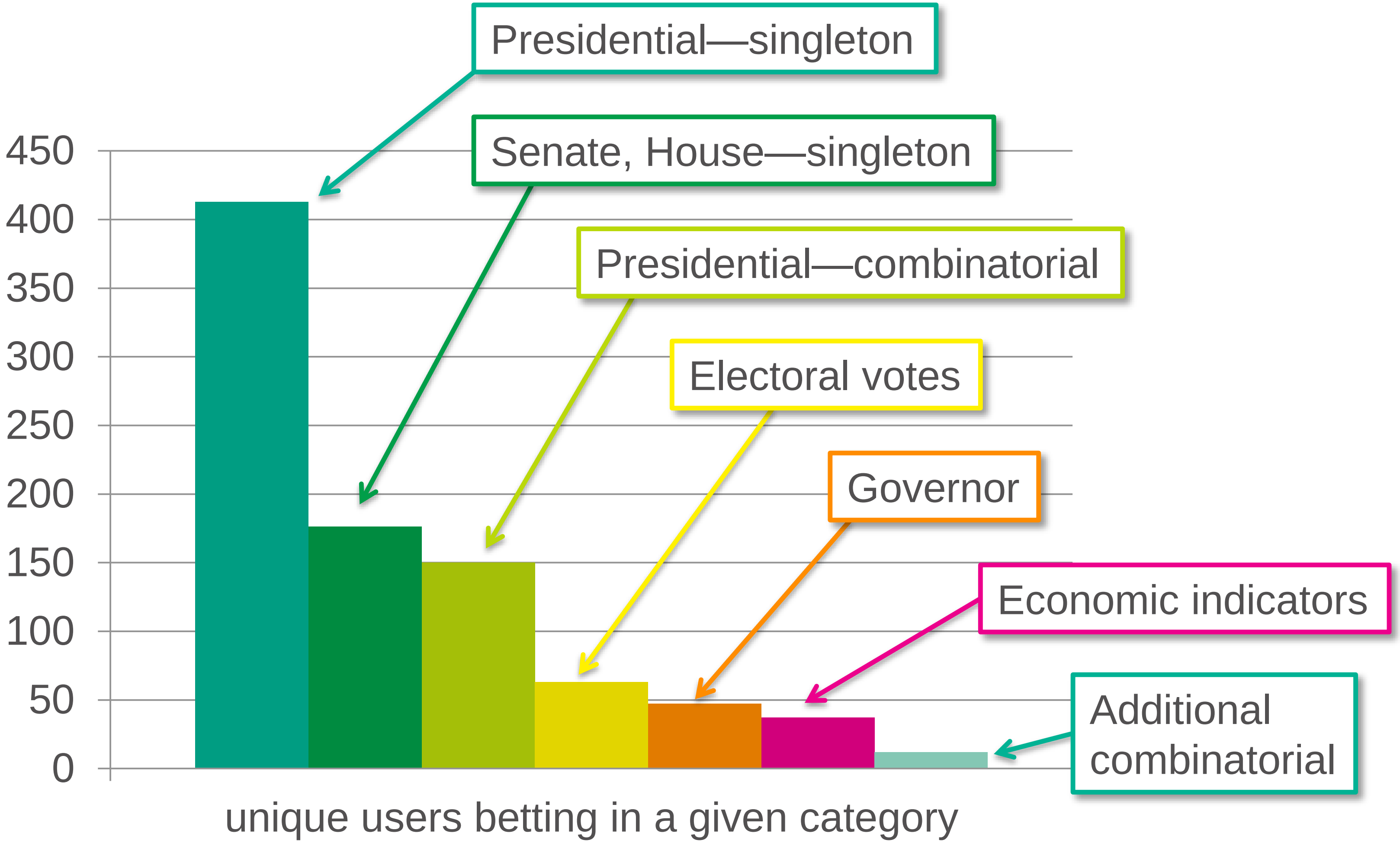
Did users place combinatorial bets?



Did users place combinatorial bets?

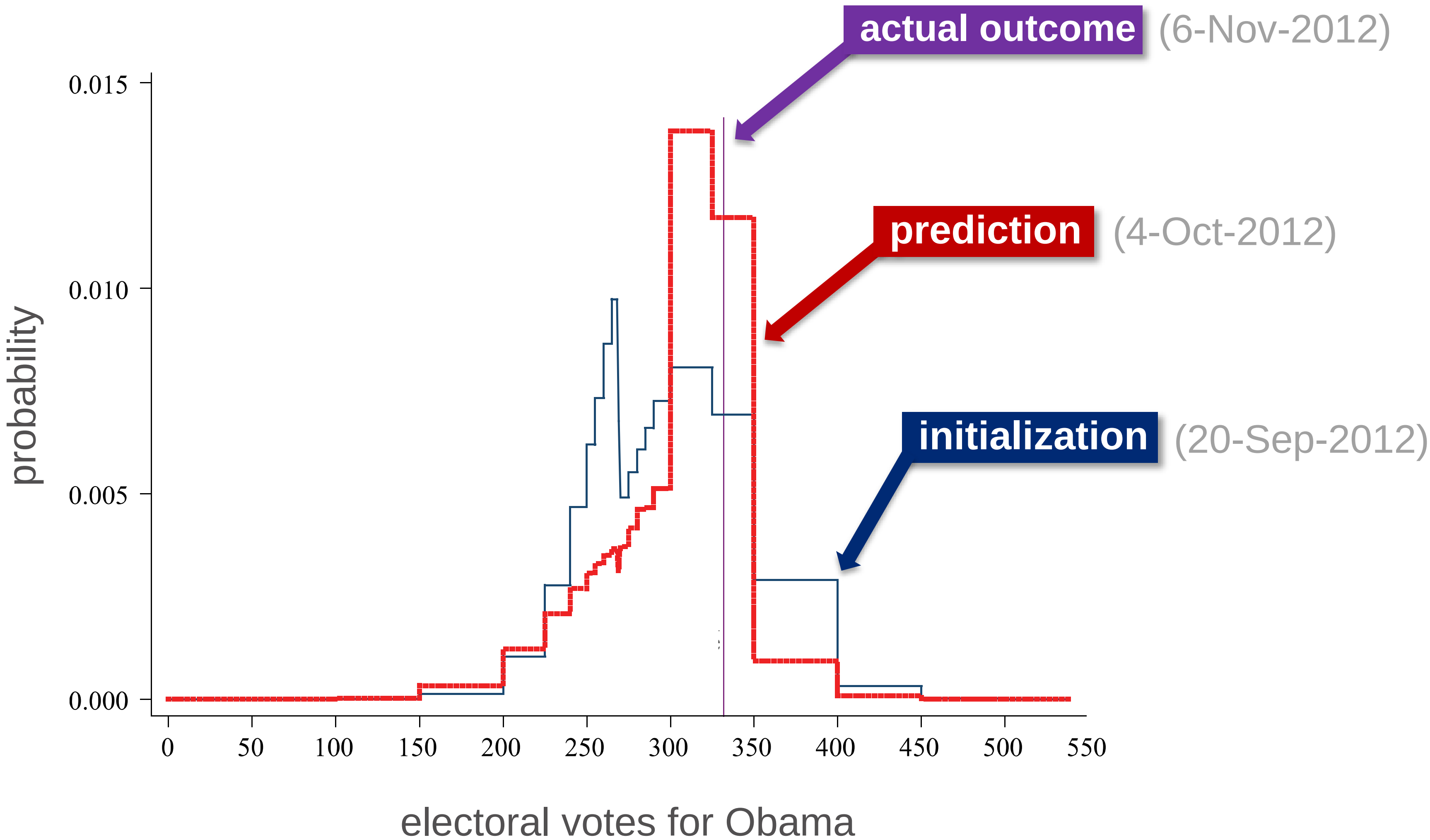


Did users place combinatorial bets?

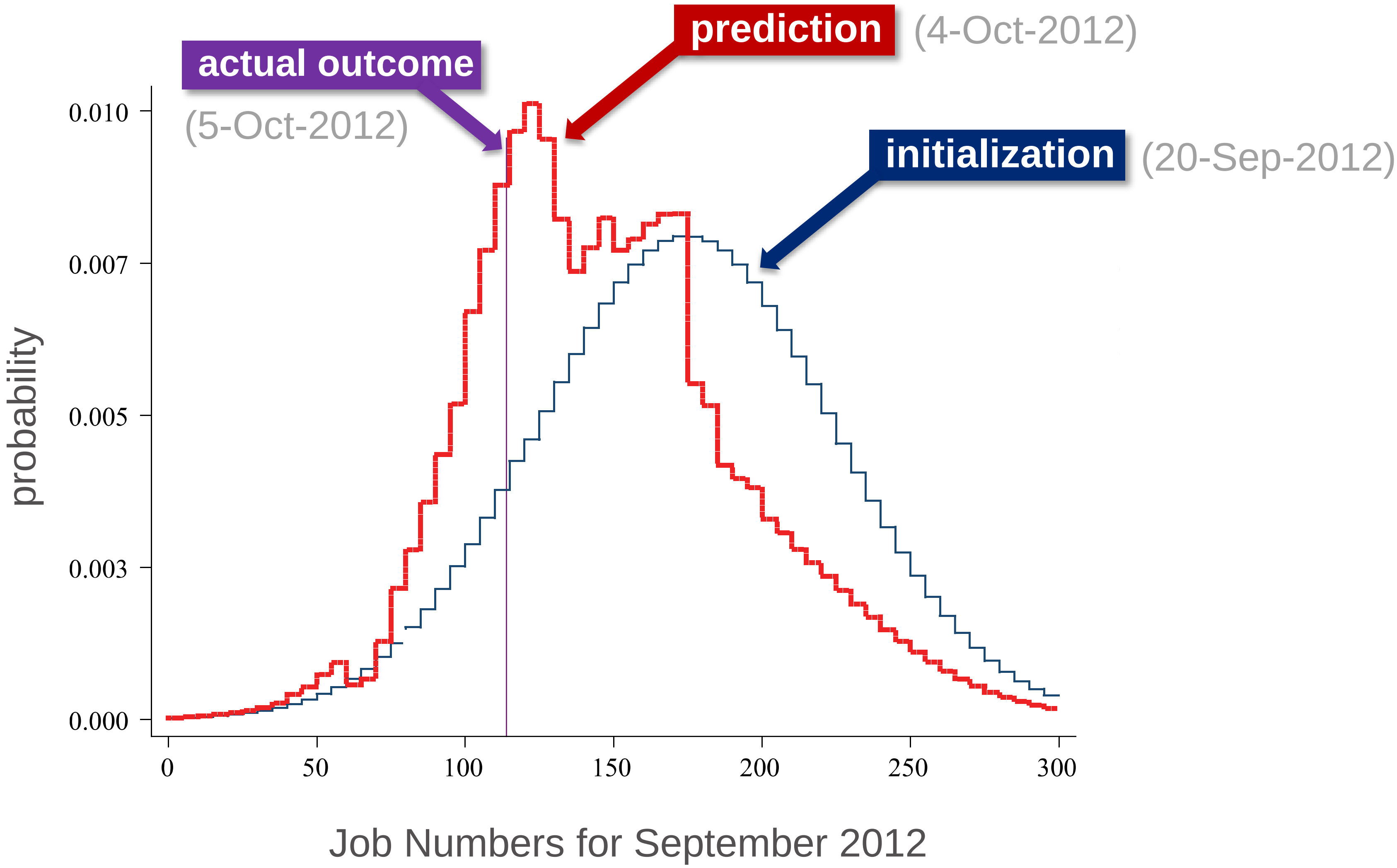


Numerical predictions:
electoral votes

Numerical predictions:
electoral votes



Numerical predictions:
job numbers



Summary

independent markets + constraints:
tractable and *accurate*

combinatorial markets can succeed with
moderate numbers of users
even on *huge outcome spaces*

meaningful forecasts for
challenging, but *relevant* outcomes:
combinatorial and *numerical*