Capability Ratios Predict Nothing

Robert J. Carroll
Florida State University

Brenton Kenkel
Vanderbilt University

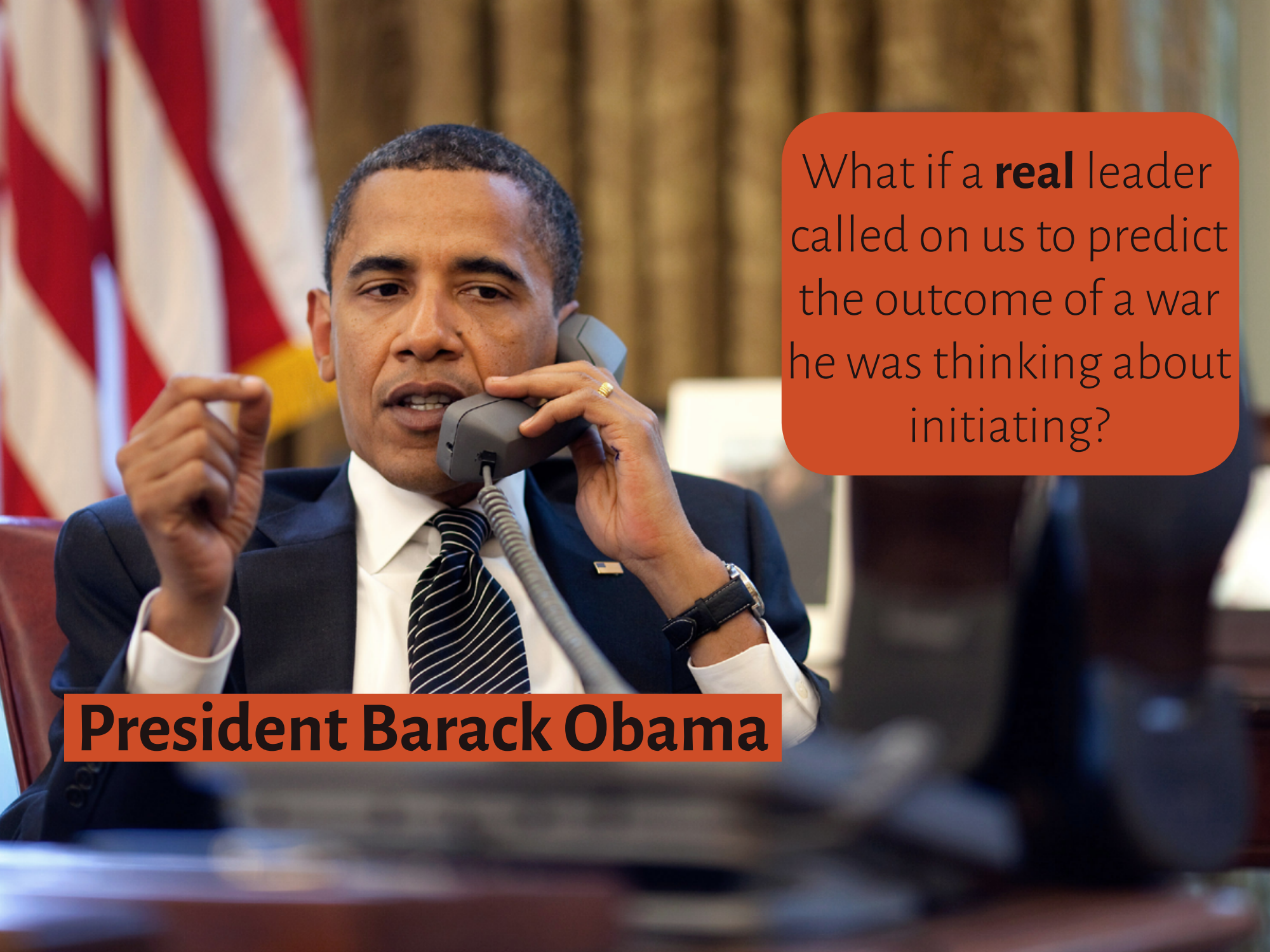
www.doe-scores.com/paper

a rather far-fetched thought experiment

a rather far-fetched thought experiment

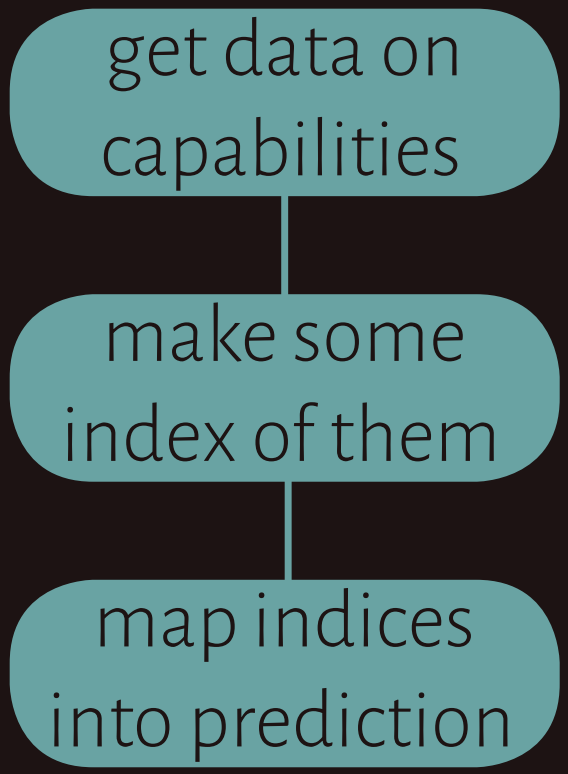
suppose policymakers had the good sense to listen to us


President Barack Obama


What if a **real** leader called on us to predict the outcome of a war he was thinking about initiating?
President Barack Obama

two plausible suggestions

two plausible suggestions



two plausible suggestions

President Obama probably isn't buying what we're selling


President Obama
~~probably~~ **isn't**
buying what we're
selling

reconsider our
two plausible suggestions

get data on capabilities

make some index of them

map indices into prediction

get data on outcomes

stick them into a model

fit and use to predict

so what's the problem with these?

get data on capabilities

make some index of them

map indices into prediction

get data on outcomes

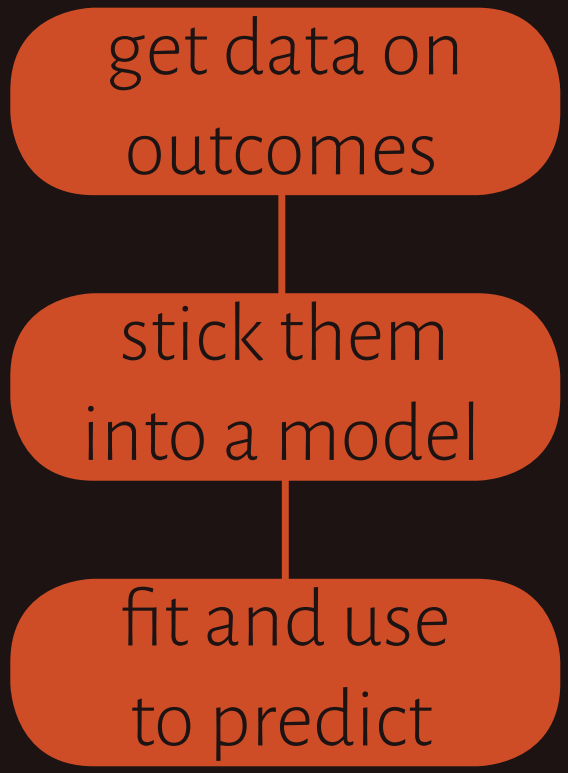
stick them into a model

fit and use to predict

so what's the problem with these?

not learning from history

not even fit *ad hoc*



so what's the problem with these?

not learning from history

not even fit
ad hoc

fighting the last battle

overfit
inflexible

so what's the problem with these?

measuring expectations is

really hard

and the theoretical stakes are quite high

Rationalist explanations for war

James D. Fearon

and the theoretical stakes are quite high

Rationalist explanations for war

James D. Fearon

Rationalist explanations for war

James D. Fearon

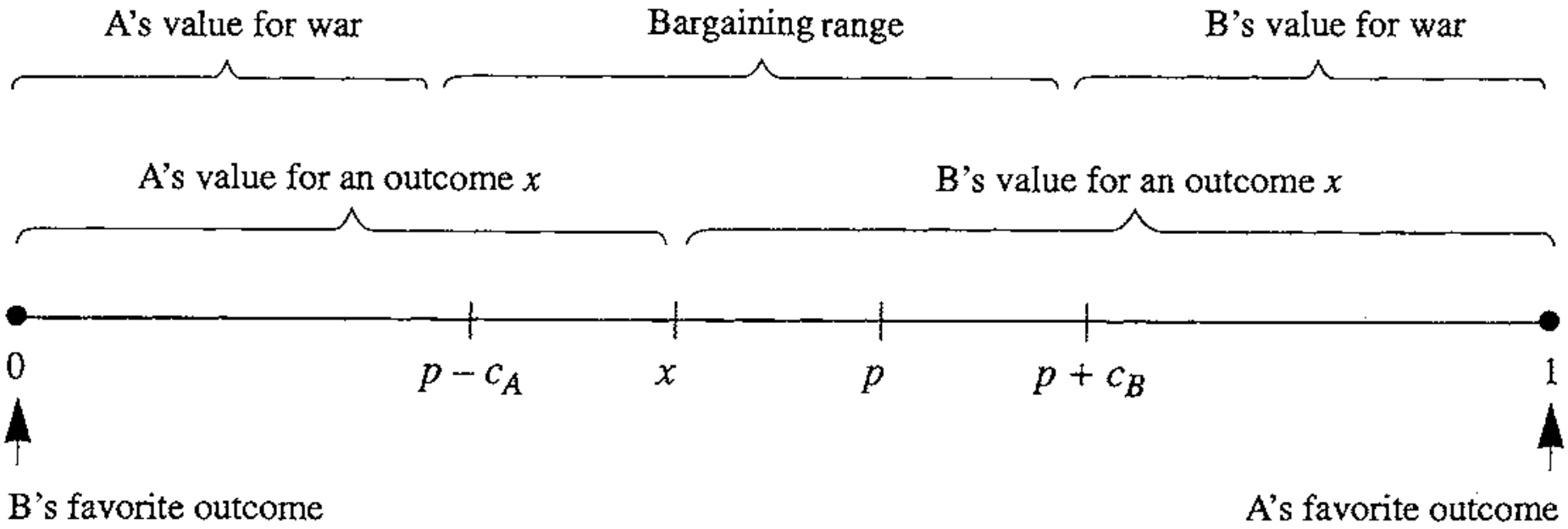


FIGURE 1. *The bargaining range*

Rationalist explanations for war

James D. Fearon

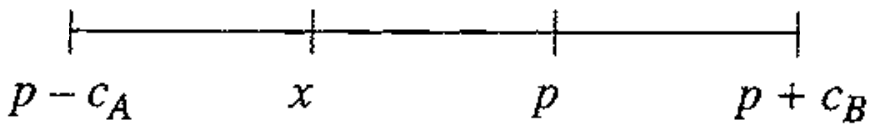


FIGURE 1. *The bargaining range*

Rationalist explanations for war

James D. Fearon

our workhorse model takes
dyadic relative power as **primitive**

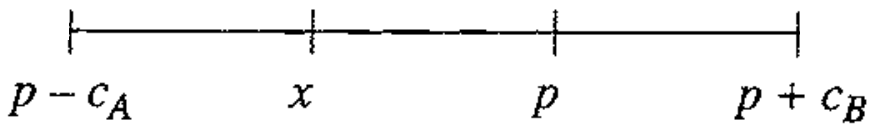


FIGURE 1. *The bargaining range*

Rationalist explanations for war

James D. Fearon

our workhorse model takes
dyadic relative power as **primitive**

this in turn affects
the way we go about
data analysis

this in turn affects
the way we go about
data analysis

War, Power, and Bargaining

William Reed Rice University
David H. Clark Binghamton University
Timothy Nordstrom University of Mississippi
Wonjae Hwang University of Tennessee

this in turn affects
the way we go about
data analysis

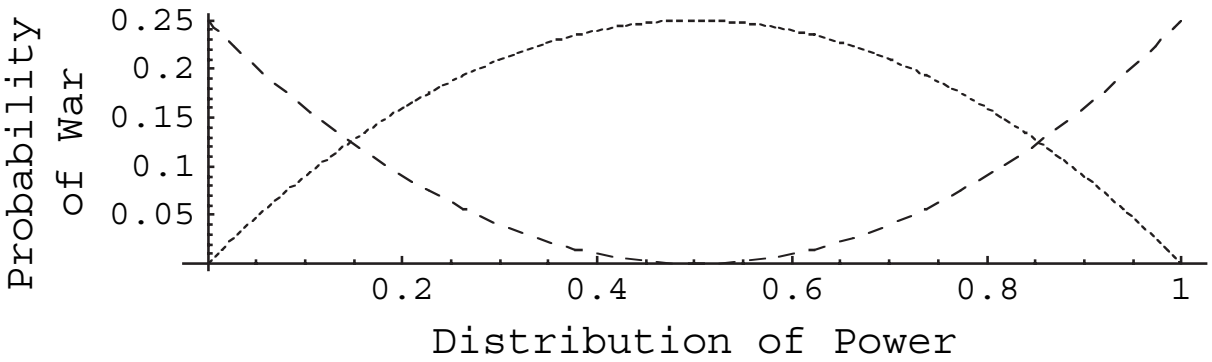
War, Power, and Bargaining

William Reed Rice University
David H. Clark Binghamton University
Timothy Nordstrom University of Mississippi
Wonjae Hwang University of Tennessee

War, Power, and Bargaining

William Reed Rice University
David H. Clark Binghamton University
Timothy Nordstrom University of Mississippi
Wonjae Hwang University of Tennessee

FIGURE 1 **Extant Theories of the Distribution of Power and the Chances of War: Theoretical expectations of balance of power (dashed parabola) and parity argument (dotted parabola) regarding how the distribution of power influences the probability of war.**



War, Power, and Bargaining

William Reed Rice University
David H. Clark Binghamton University
Timothy Nordstrom University of Mississippi
Wonjae Hwang University of Tennessee

War, Power, and Bargaining

William Reed Rice University
David H. Clark Binghamton University
Timothy Nordstrom University of Mississippi
Wonjae Hwang University of Tennessee

our measures of benefits (q) and power (p) (bounded between zero and one), and then computing:[13]

$$|q - p|$$

and

$$(q - p)^2$$

These terms exactly represent the terms and functional form in the equilibrium from Powell's (1999) model and provide the means for testing the relationships between the distributions of power and benefits and armed conflict. We can write the empirical model this way:

$$y^* = X\beta + Z\gamma + u \tag{1}$$

where y^* is a latent measure of conflict; its empirical realization, y equals one when $y^* > 0$ and zero otherwise. We use the Militarized Interstate Dispute data to measure this dichotomous indicator of conflict (Ghosn, Palmer, and Bremer 2004). Our dependent variable takes on a value of one in years when a dyad commences a militarized dispute, zero otherwise. Further,

War, Power, and Bargaining

William Reed Rice University
David H. Clark Binghamton University
Timothy Nordstrom University of Mississippi
Wonjae Hwang University of Tennessee

our measures of benefits (q) and power (p) (bounded between zero and one), and then computing:[13]

$$|q - p|$$

and

$$(q - p)^2$$

These terms exactly represent the terms and functional form in the equilibrium from Powell's (1999) model and provide the means for testing the relationships between the distributions of power and benefits and armed conflict. We can write the empirical model this way:

$$y^* = X\beta + Z\gamma + u \qquad (1)$$

where y^* is a latent measure of conflict; its empirical realization, y equals one when $y^* > 0$ and zero otherwise. We use the Militarized Interstate Dispute data to measure this dichotomous indicator of conflict (Ghosn, Palmer, and Bremer 2004). Our dependent variable takes on a value of one in years when a dyad commences a militarized dispute, zero otherwise. Further,

War, Power, and Bargaining

William Reed Rice University
David H. Clark Binghamton University
Timothy Nordstrom University of Mississippi
Wonjae Hwang University of Tennessee

our measures of benefits (q) and power (p) (bounded between zero and one), and then computing:[13]

$$|q - p|$$

and

$$(q - p)^2$$

These terms exactly represent the terms and functional form in the equilibrium from Powell's (1999) model and provide the means for testing the relationships between the distributions of power and benefits and armed conflict. We can write the empirical model this way:

$$y^* = X\beta + Z\gamma + u \qquad (1)$$

where y^* is a latent measure of conflict; its empirical realization, y equals one when $y^* > 0$ and zero otherwise. We use the Militarized Interstate Dispute data to measure this dichotomous indicator of conflict (Ghosn, Palmer, and Bremer 2004). Our dependent variable takes on a value of one in years when a dyad commences a militarized dispute, zero otherwise. Further,

Empirical Strategy

We adopt a dyad year design for all states and years that we have data on United Nations' roll-call data (1946–2000). We operationalize the distribution of power by measuring dyadic differences in military capabilities using the Correlates of War CINC score data (Singer, Bremer, and Stuckey 1972).[11] Pursuant to the discussion above, we operationalize the distribution of benefits by measuring the ideological distance between the two states in the dyad in terms of the distance between the states' revealed ideal point estimates derived from the UNGA voting data.[12]

The hypothesis we want to test is that divergences between the distributions of power and benefits will be associated with conflict whereas convergences between those distributions will not. Another way to describe the hypothesis is that the effect of the distribution of power depends on its relationship to the distribution of benefits (and we prefer to state this in terms of the distribution of power since the literature on power is

War, Power, and Bargaining

William Reed Rice University
David H. Clark Binghamton University
Timothy Nordstrom University of Mississippi
Wonjae Hwang University of Tennessee

our measures of benefits (q) and power (p) (bounded between zero and one), and then computing:[13]

$$|q - p|$$

and

$$(q - p)^2$$

These terms exactly represent the terms and functional form in the equilibrium from Powell's (1999) model and provide the means for testing the relationships between the distributions of power and benefits and armed conflict. We can write the empirical model this way:

$$y^* = X\beta + Z\gamma + u \qquad (1)$$

where y^* is a latent measure of conflict; its empirical realization, y equals one when $y^* > 0$ and zero otherwise. We use the Militarized Interstate Dispute data to measure this dichotomous indicator of conflict (Ghosn, Palmer, and Bremer 2004). Our dependent variable takes on a value of one in years when a dyad commences a militarized dispute, zero otherwise. Further,

Empirical Strategy

We adopt a dyad year design for all states and years that we have data on United Nations' roll-call data (1946–2000). We operationalize the distribution of power by measuring dyadic differences in military capabilities using the Correlates of War CINC score data (Singer, Bremer, and Stuckey 1972).[11] Pursuant to the discussion above, we operationalize the distribution of benefits by measuring the ideological distance between the two states in the dyad in terms of the distance between the states' revealed ideal point estimates derived from the UNGA voting data.[12]

The hypothesis we want to test is that divergences between the distributions of power and benefits will be associated with conflict whereas convergences between those distributions will not. Another way to describe the hypothesis is that the effect of the distribution of power depends on its relationship to the distribution of benefits (and we prefer to state this in terms of the distribution of power since the literature on power is

War, Power, and Bargaining

William Reed Rice University
David H. Clark Binghamton University
Timothy Nordstrom University of Mississippi
Wonjae Hwang University of Tennessee

scores of empirical IR papers proceed along similar lines

War, Power, and Bargaining

William Reed Rice University
David H. Clark Binghamton University
Timothy Nordstrom University of Mississippi
Wonjae Hwang University of Tennessee

scores of empirical IR papers proceed along similar lines

but don't just take my word for it

scores of empirical IR papers proceed along similar lines

but don’t just take my word for it

Colaresi and Thompson, 2005; Grieg, 2005; Senese, 2005; Sobek, 2005; Werner and Yuen, 2005; Allee and Huth, 2006; Bearce, Flanagan and Floros, 2006; Bennett, 2006; James, Park and Choi, 2006; Oneal and Tir, 2006; Pevehouse and Russett, 2006; Rasler and Thompson, 2006; Sobek, Abouharb and Ingram, 2006; Bearce and Bondanella, 2006; Chapman and Roeder, 2007; Clare, 2007; Crescenzi, 2007; Gartzke, 2007; Gent, 2007; Gibler, 2007; Lektzian and Sprecher, 2007; Long, Nordstrom and Baek, 2007; Maoz et al., 2007; Mitchell and Hensel, 2007; Morrow, 2007; Aydin, 2008; Beardsley, 2008; Braumoeller, 2008; Clark, Nordstrom and Reed, 2008; Fordham, 2008; Lo, Hashimoto and Reiter, 2008; Mattes, 2008; Salehyan, 2008*a*; Saleh 2009; Kroe ardt, 2010 ouva, 2010 lorey and 2011; Choi vahri and wnes and berg, Robi 2012; Uzon ibler and wiak and erski, 2013 and Leed ibler and Tir, 2014; Jung, 2014; Park and Colaresi, 2014; Potter and Baum, 2014; Wolford, 2014

scores of empirical IR papers proceed along similar lines

but don't just take my word for it

Colaresi and Thompson, 2005; Grieg, 2005; Senese, 2005; Sobek, 2005; Werner and Yuen, 2005; Allee and Huth, 2006; Bearce, Flanagan and Floros, 2006; Bennett, 2006; James, Park and Choi, 2006; Oneal and Tir, 2006; Pevehouse and Russett, 2006; Rasler and Thompson, 2006; Sobek, Abouharb and Ingram, 2006; Bearce and Bondanella, 2006; Chapman and Roeder, 2007; Clare, 2007; Crescenzi, 2007; Gartzke, 2007; Gent, 2007; Gibler, 2007; Lektzian and Sprecher, 2007; Long, Nordstrom and Baek, 2007; Maoz et al., 2007; Mitchell and Hensel, 2007; Morrow, 2007; Aydin, 2008; Beardsley, 2008; Braumoeller, 2008; Clark, Nordstrom and Reed, 2008; Fordham, 2008; Lo, Hashimoto and Reiter, 2008; Mattes, 2008; Salehyan, 2008*a*; Salehyan, 2008*b*; Weeks, 2008; Arena and Palmer, 2009; Bearce, Floros and McKibben, 2009; Kroenig, 2009; Langlois and Langlois, 2009; Maoz, 2009; Rider, 2009; Allee and Peinhardt, 2010; Dreyer 2010; Faussett and Volgy, 2010; Gibler and Tir, 2010; Lektzian, Prins and Souva, 2010; Mattes and Vonnahme, 2010; Reed and Chiba, 2010; Schultz, 2010; Shannon, Morey and Boehmke, 2010; Tir, 2010; Xiang, 2010; Benson, 2011; Brochmann and Hensel, 2011; Choi, 2011; Dafoe, 2011; Huth, Croco and Appel, 2011; Koga, 2011; Morey, 2011; Zawahri and Mitchell, 2011; Allen and Fordham, 2012; Colaresi, 2012; Crescenzi et al., 2012; Downes and Sechser, 2012; Findley, Piazza and Young, 2012; Huth, Croco and Appel, 2012; Kleinberg, Robinson and French, 2012; Mattes, 2012; Owsiak, 2012; Sobek, Foster and Robison 2012; Uzonyi, Souva and Golder, 2012; Weeks, 2012; Clare, 2013; Gartzke and Weisiger, 2013; Gibler and Hutchison, 2013; Huth, Croco and Appel, 2013; Kroenig, 2013; Mousseau, 2013; Oswiak and Rider, 2013; Rider, 2013; Sechser and Fuhrmann, 2013; Whang, McLean and Kuberski, 2013; Williams, 2013; Carnegie, 2014; Dafoe, Oneal and Russett, 2014; Fang, Johnson and Leeds 2014; Fuhrmann and Sechser, 2014; Gartzke and Weisiger, 2014; Gibler, 2014; Gibler and Tir, 2014; Jung, 2014; Park and Colaresi, 2014; Potter and Baum, 2014; Wolford, 2014

Colaresi and Thompson, 2005; Grieg, 2005; Senese, 2005; Sobek, 2005; Werner and Yuen, 2005; Allee and Huth, 2006; Bearce, Flanagan and Floros, 2006; Bennett, 2006; James, Park and Choi, 2006; Oneal and Tir, 2006; Pevehouse and Russett, 2006; Rasler and Thompson, 2006; Sobek, Abouharb and Ingram, 2006; Bearce and Bondanella, 2006; Chapman and Ro-
eder, 200 Lektzian
and Sprec d Hensel,
2007; Mo strom and
Reed, 200 an, 2008*a*;
Salehyan, ben, 2009;
Kroenig, Peinhardt,
2010; Dre nd Souva,
2010; Ma on, Morey
and Boeh sel, 2011;
Choi, 201 ; Zawahri
and Mitch ; Downes
and Sechs Kleinberg,
Robinson son 2012;
Uzonyi, S 13; Gibler
and Hutc 3; Oswiak
and Rider, 2013; Rider, 2013; Sechser and Fuhrmann, 2013; Whang, McLean and Kuberski, 2013; Williams, 2013; Carnegie, 2014; Dafoe, Oneal and Russett, 2014; Fang, Johnson and Leeds 2014; Fuhrmann and Sechser, 2014; Gartzke and Weisiger, 2014; Gibler, 2014; Gibler and Tir, 2014; Jung, 2014; Park and Colaresi, 2014; Potter and Baum, 2014; Wolford, 2014

it's probably worth thinking this through

the argument

the argument

we need a proxy that **predicts dispute outcomes** well

the argument

we need a proxy that **predicts dispute outcomes** well

the standard proxy (the **capabilty ratio**) predicts awfully

the argument

we need a proxy that **predicts dispute outcomes** well

the standard proxy (the **capabilty ratio**) predicts awfully

we use **predictive algorithms** to construct a better proxy

preview of the results

the capability ratio provides only a 1% improvement over a null model

conversely, our proxy—the **DOE score**—provides a 20% improvement over the null using the same data

in a series of 18 replications, the DOE score outperformed the capability ratio 14 times

roadmap

1. a refresher on the capability ratio
2. how we built a better proxy
3. what our proxy did on its summer vacation
4. wrap-up and conclusions

1

a refresher on the capability ratio

not that anybody here needs this, but...

not that anybody here needs this, but...

let's make a CINC score, anyway

not that anybody here needs this, but...

begin with data on state *i* in year *t*

not that anybody here needs this, but...

begin with data on state *i* in year *t*

$$
\begin{aligned}
& \text{troops}_{it} \\
+ \quad & \text{spending}_{it} \\
+ \quad & \text{population}_{it} \\
+ \quad & \text{urban}_{it} \\
+ \quad & \text{energy}_{it} \\
+ \quad & \text{ironsteel}_{it}
\end{aligned}
$$

not that anybody here needs this, but...

begin with data on state *i* in year *t*
we refer to these as the **components**

$$
\begin{aligned}
& \text{troops}_{it} \\
+ \quad & \text{spending}_{it} \\
+ \quad & \text{population}_{it} \\
+ \quad & \text{urban}_{it} \\
+ \quad & \text{energy}_{it} \\
+ \quad & \text{ironsteel}_{it}
\end{aligned}
$$

not that anybody here needs this, but...

divide each by its world total

$$
\begin{aligned}
& \text{troops}_{it} \\
+ \quad & \text{spending}_{it} \\
+ \quad & \text{population}_{it} \\
+ \quad & \text{urban}_{it} \\
+ \quad & \text{energy}_{it} \\
+ \quad & \text{ironsteel}_{it}
\end{aligned}
$$

not that anybody here needs this, but...

divide each by its world total

$$
\begin{aligned}
& \text{troops}_{it}/\Sigma_j\text{troops}_{jt} \\
+ \quad & \text{spending}_{it}/\Sigma_j\text{spending}_{jt} \\
+ \quad & \text{population}_{it}/\Sigma_j\text{population}_{jt} \\
+ \quad & \text{urban}_{it}/\Sigma_j\text{urban}_{jt} \\
+ \quad & \text{energy}_{it}/\Sigma_j\text{energy}_{jt} \\
+ \quad & \text{ironsteel}_{it}/\Sigma_j\text{ironsteel}_{jt}
\end{aligned}
$$

not that anybody here needs this, but...

divide each by its world total
we refer to these as the **proportions**

$$
\begin{aligned}
& \text{troops}_{it}/\Sigma_j\text{troops}_{jt} \\
+ \quad & \text{spending}_{it}/\Sigma_j\text{spending}_{jt} \\
+ \quad & \text{population}_{it}/\Sigma_j\text{population}_{jt} \\
+ \quad & \text{urban}_{it}/\Sigma_j\text{urban}_{jt} \\
+ \quad & \text{energy}_{it}/\Sigma_j\text{energy}_{jt} \\
+ \quad & \text{ironsteel}_{it}/\Sigma_j\text{ironsteel}_{jt}
\end{aligned}
$$

not that anybody here needs this, but...

divide each by its world total
we refer to these as the **proportions**

$$
\begin{aligned}
& \text{troops}_{it}/\Sigma_j\text{troops}_{jt} \\
+ \quad & \text{spending}_{it}/\Sigma_j\text{spending}_{jt} \\
+ \quad & \text{population}_{it}/\Sigma_j\text{population}_{jt} \\
+ \quad & \text{urban}_{it}/\Sigma_j\text{urban}_{jt} \\
+ \quad & \text{energy}_{it}/\Sigma_j\text{energy}_{jt} \\
+ \quad & \text{ironsteel}_{it}/\Sigma_j\text{ironsteel}_{jt}
\end{aligned}
$$

we just made CINC sensitive to
system membership

not that anybody here needs this, but...

divide each by its world total
we refer to these as the **proportions**

$$
\begin{aligned}
& \text{troops}_{it}/\Sigma_j\text{troops}_{jt} \\
+ \quad & \text{spending}_{it}/\Sigma_j\text{spending}_{jt} \\
+ \quad & \text{population}_{it}/\Sigma_j\text{population}_{jt} \\
+ \quad & \text{urban}_{it}/\Sigma_j\text{urban}_{jt} \\
+ \quad & \text{energy}_{it}/\Sigma_j\text{energy}_{jt} \\
+ \quad & \text{ironsteel}_{it}/\Sigma_j\text{ironsteel}_{jt}
\end{aligned}
$$

not that anybody here needs this, but...

weight each proportion equally

$$
\begin{aligned}
& \text{troops}_{it}/\Sigma_j \text{troops}_{jt} \\
+ \quad & \text{spending}_{it}/\Sigma_j \text{spending}_{jt} \\
+ \quad & \text{population}_{it}/\Sigma_j \text{population}_{jt} \\
+ \quad & \text{urban}_{it}/\Sigma_j \text{urban}_{jt} \\
+ \quad & \text{energy}_{it}/\Sigma_j \text{energy}_{jt} \\
+ \quad & \text{ironsteel}_{it}/\Sigma_j \text{ironsteel}_{jt}
\end{aligned}
$$

not that anybody here needs this, but...

weight each proportion equally

$$
\begin{aligned}
\mathrm{CINC}_{it} = \quad & \tfrac{1}{6} \times \mathrm{troops}_{it}/\Sigma_j \mathrm{troops}_{jt} \\
& + \tfrac{1}{6} \times \mathrm{spending}_{it}/\Sigma_j \mathrm{spending}_{jt} \\
& + \tfrac{1}{6} \times \mathrm{population}_{it}/\Sigma_j \mathrm{population}_{jt} \\
& + \tfrac{1}{6} \times \mathrm{urban}_{it}/\Sigma_j \mathrm{urban}_{jt} \\
& + \tfrac{1}{6} \times \mathrm{energy}_{it}/\Sigma_j \mathrm{energy}_{jt} \\
& + \tfrac{1}{6} \times \mathrm{ironsteel}_{it}/\Sigma_j \mathrm{ironsteel}_{jt}
\end{aligned}
$$

not that anybody here needs this, but...

weight each proportion equally

$$
\begin{aligned}
\text{CINC}_{it} = \quad & \tfrac{1}{6} \times \text{troops}_{it}/\Sigma_j \text{troops}_{jt} \\
& + \tfrac{1}{6} \times \text{spending}_{it}/\Sigma_j \text{spending}_{jt} \\
& + \tfrac{1}{6} \times \text{population}_{it}/\Sigma_j \text{population}_{jt} \\
& + \tfrac{1}{6} \times \text{urban}_{it}/\Sigma_j \text{urban}_{jt} \\
& + \tfrac{1}{6} \times \text{energy}_{it}/\Sigma_j \text{energy}_{jt} \\
& + \tfrac{1}{6} \times \text{ironsteel}_{it}/\Sigma_j \text{ironsteel}_{jt}
\end{aligned}
$$

equal weighting is entirely *ad hoc*

not that anybody here needs this, but...

weight each proportion equally

$$
\begin{aligned}
\mathrm{CINC}_{it} = \quad & \tfrac{1}{6} \times \mathrm{troops}_{it}/\Sigma_j \mathrm{troops}_{jt} \\
& + \tfrac{1}{6} \times \mathrm{spending}_{it}/\Sigma_j \mathrm{spending}_{jt} \\
& + \tfrac{1}{6} \times \mathrm{population}_{it}/\Sigma_j \mathrm{population}_{jt} \\
& + \tfrac{1}{6} \times \mathrm{urban}_{it}/\Sigma_j \mathrm{urban}_{jt} \\
& + \tfrac{1}{6} \times \mathrm{energy}_{it}/\Sigma_j \mathrm{energy}_{jt} \\
& + \tfrac{1}{6} \times \mathrm{ironsteel}_{it}/\Sigma_j \mathrm{ironsteel}_{jt}
\end{aligned}
$$

not that anybody here needs this, but...

so now what?

well, now you make a capability ratio

well, now you make a capability ratio

$$\text{Capability Ratio}_{AB,t} = \frac{\text{CINC}_{At}}{\text{CINC}_{At} + \text{CINC}_{Bt}}$$

well, now you make a capability ratio

constant returns to scale is *ad hoc*

$$\text{Capability Ratio}_{AB,t} = \frac{CINC_{At}}{CINC_{At} + CINC_{Bt}}$$

well, now you make a capability ratio

$$\text{Capability Ratio}_{AB,t} = \frac{\text{CINC}_{At}}{\text{CINC}_{At} + \text{CINC}_{Bt}}$$

well, now you make a capability ratio

$$\text{Capability Ratio}_{AB,t} =$$

and who said a ratio was necessarily best?

why the capability ratio may not be best

why the capability ratio may not be best

constant returns to scale is *ad hoc*

and who said a ratio was necessarily best?

equal weighting is entirely *ad hoc*

we just made CINC sensitive to system membership

why the capability ratio may not be best

these are the kinds of issues that make one wonder whether the capability ratio can be a useful predictor of dispute outcomes

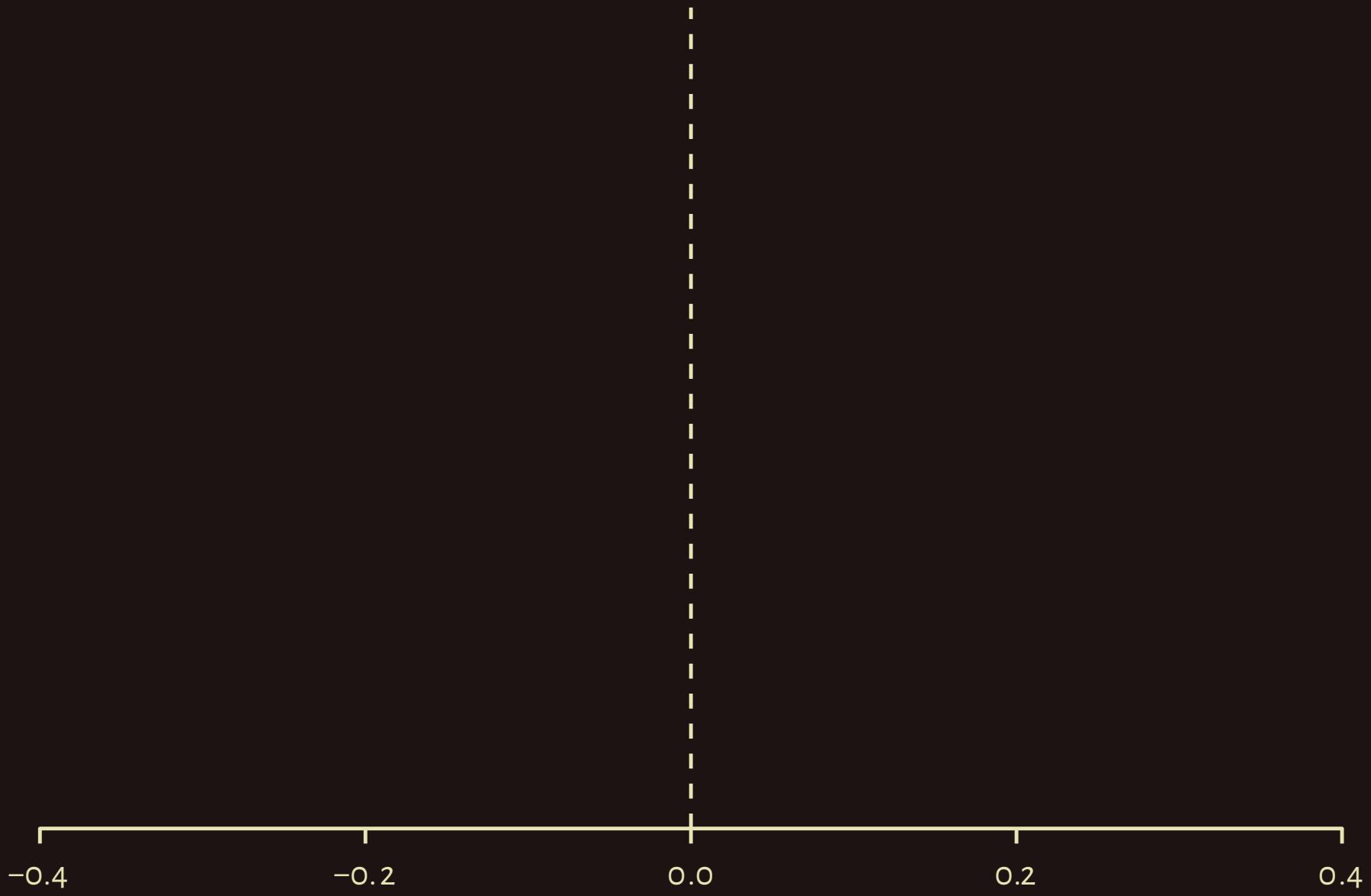
why the capability ratio may not be best

these are the kinds of issues that make one wonder whether the capability ratio can be a useful predictor of dispute outcomes

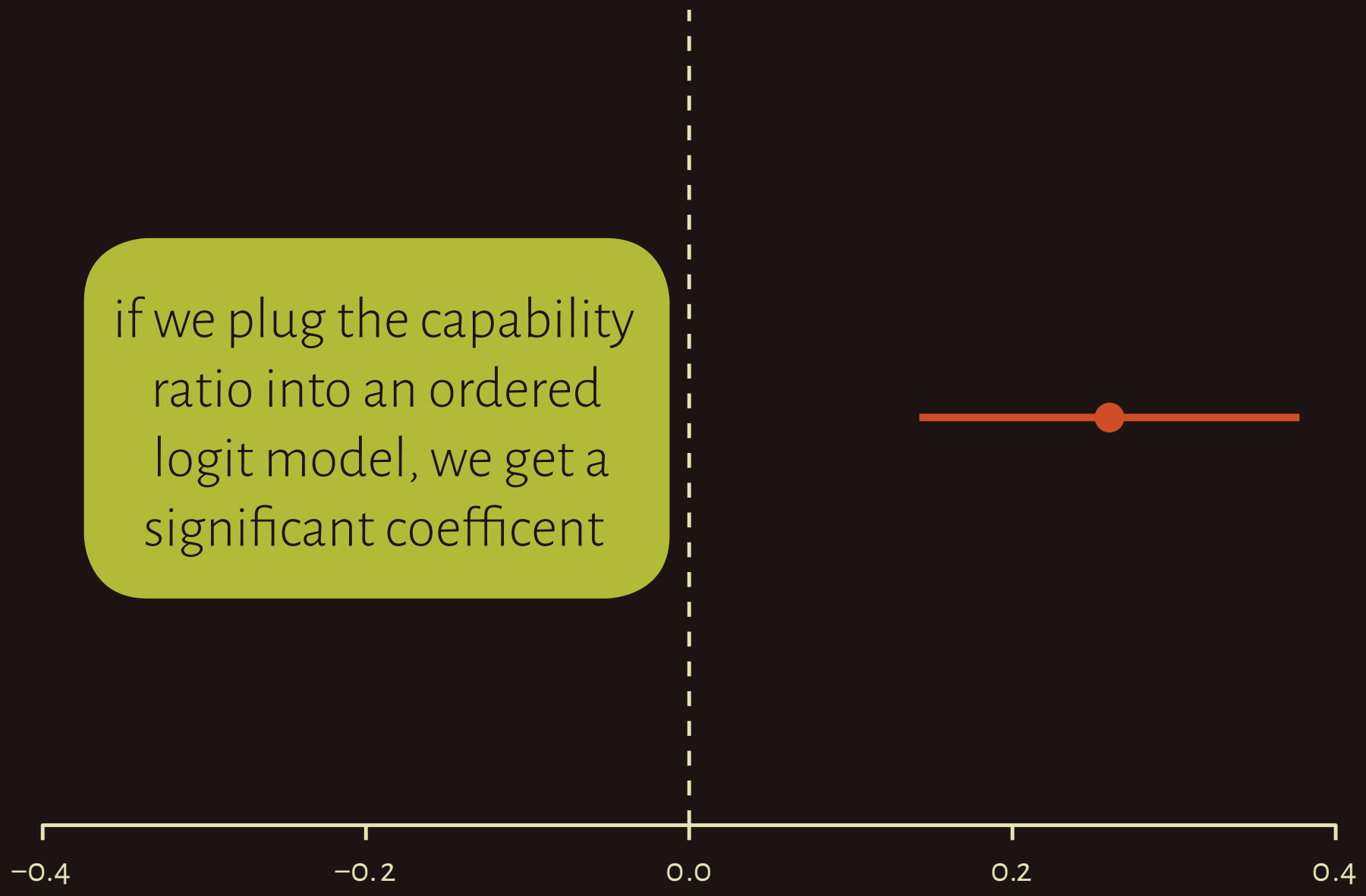
so, we tried to use the capability ratio as a predictor of dispute outcomes using MID data (**88%** of the 1,740 MIDs end in stalemate)

why the capability ratio may not be best

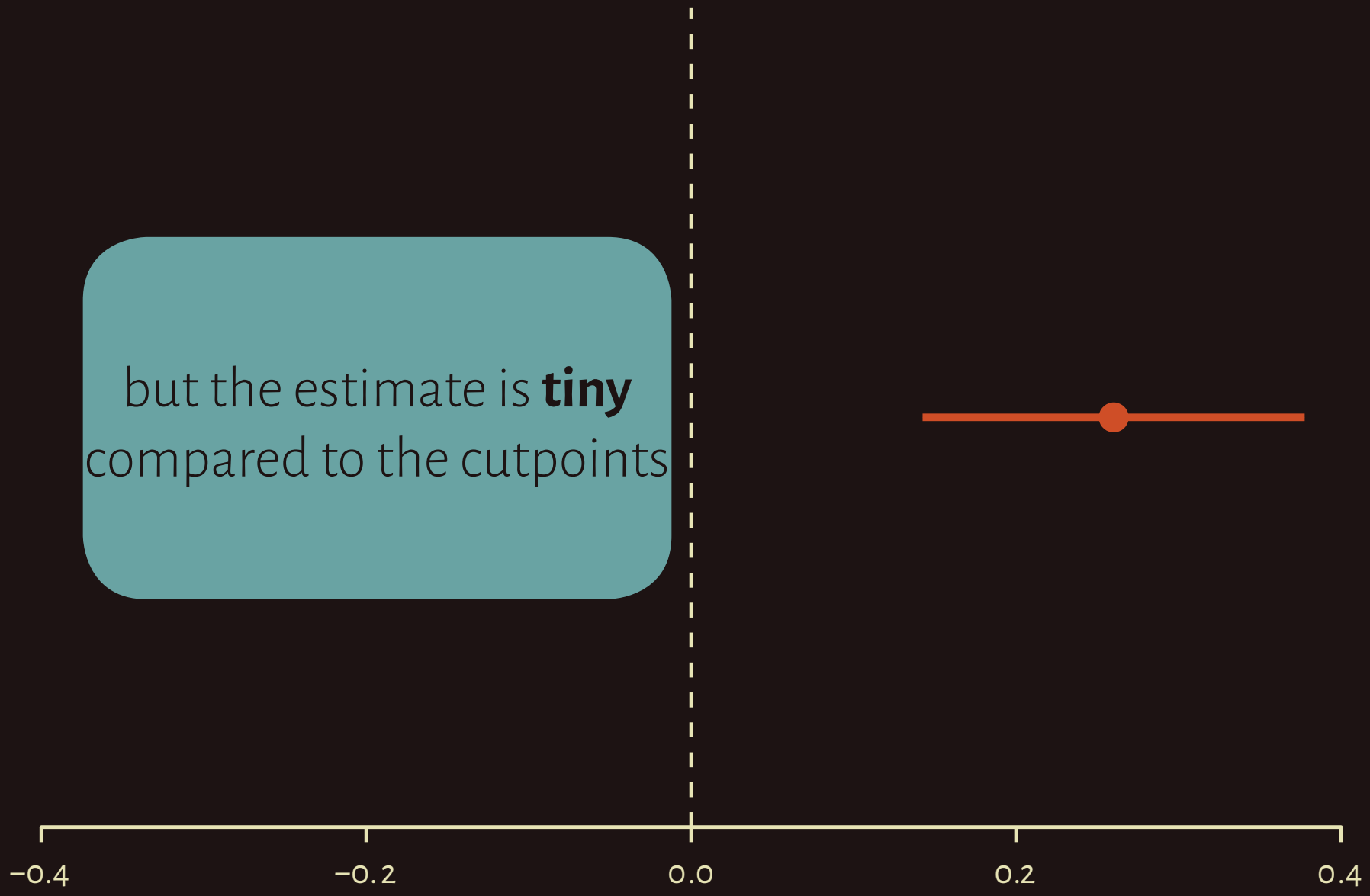
it did not go well


−0.4
−0.2
0.0
0.2
0.4

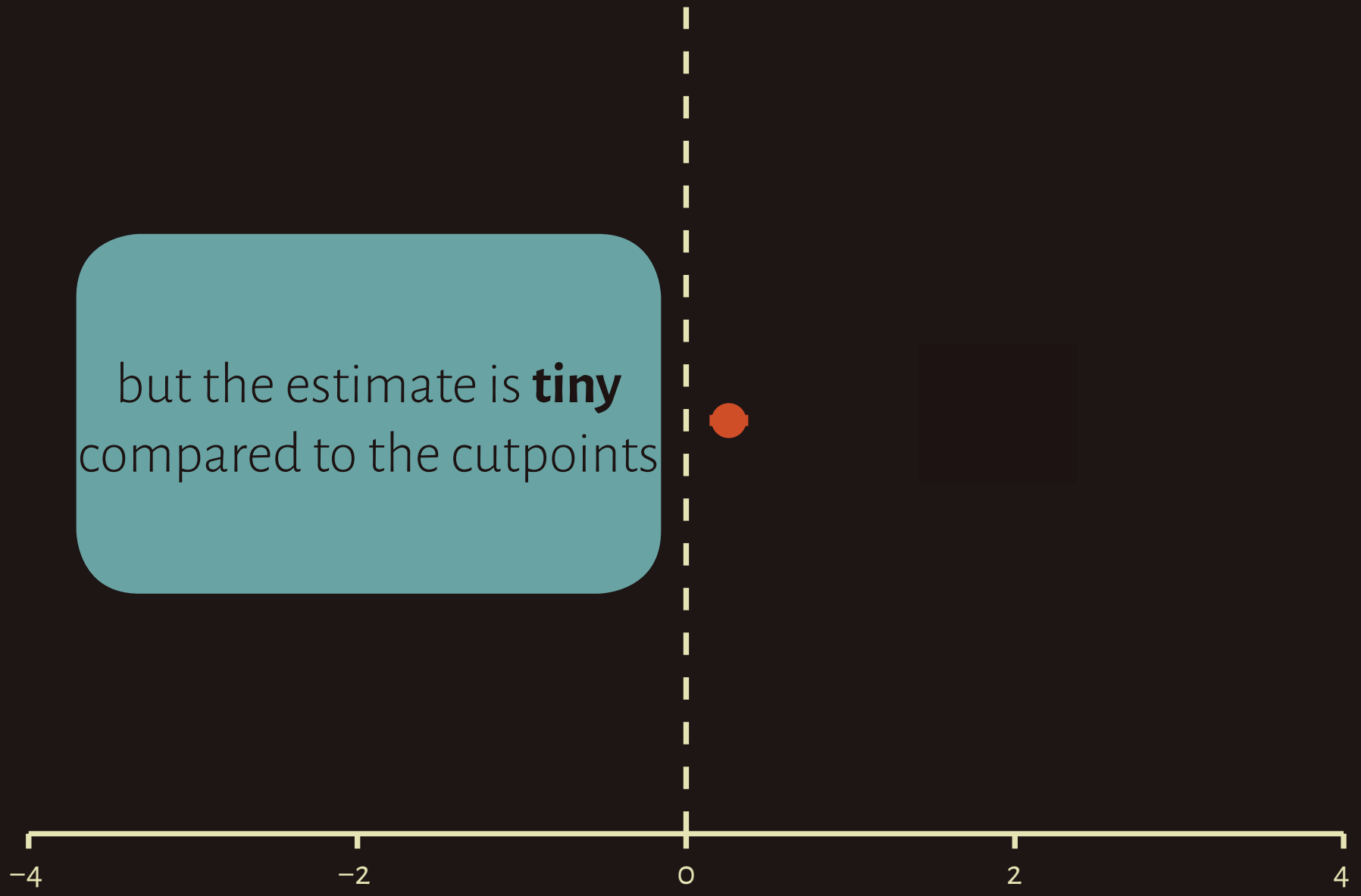
if we plug the capability ratio into an ordered logit model, we get a significant coefficent

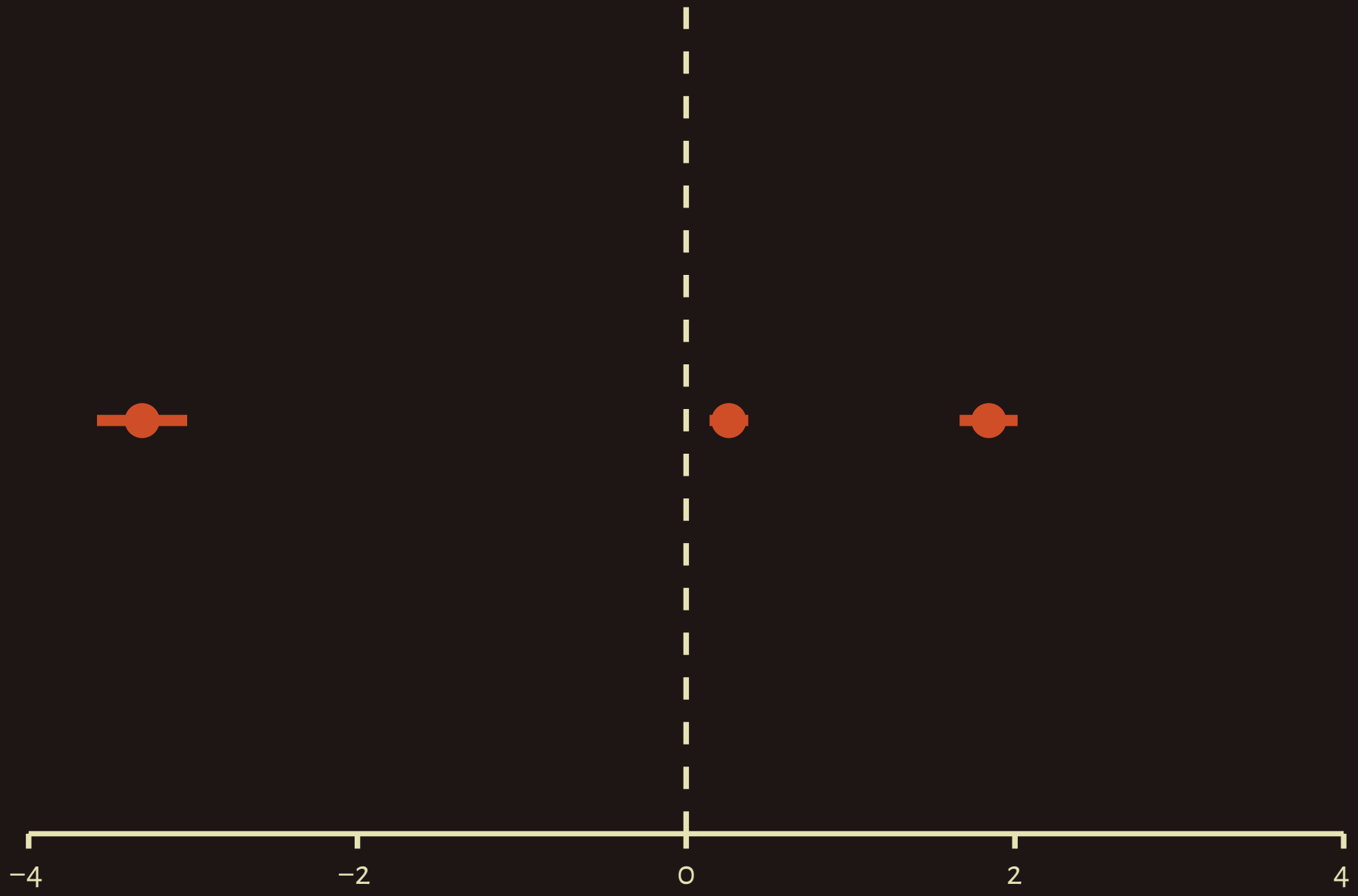


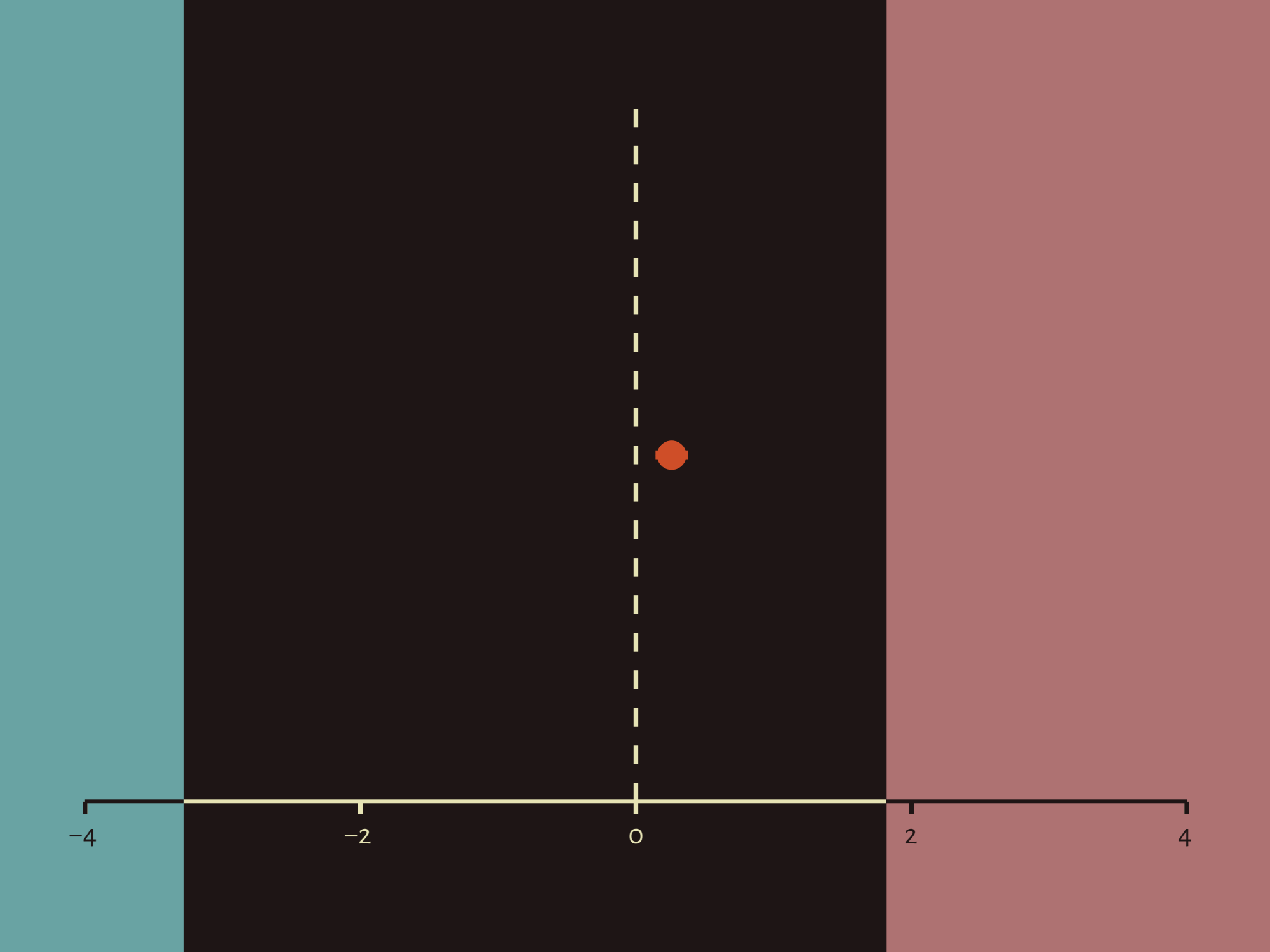
but the estimate is **tiny**
compared to the cutpoints

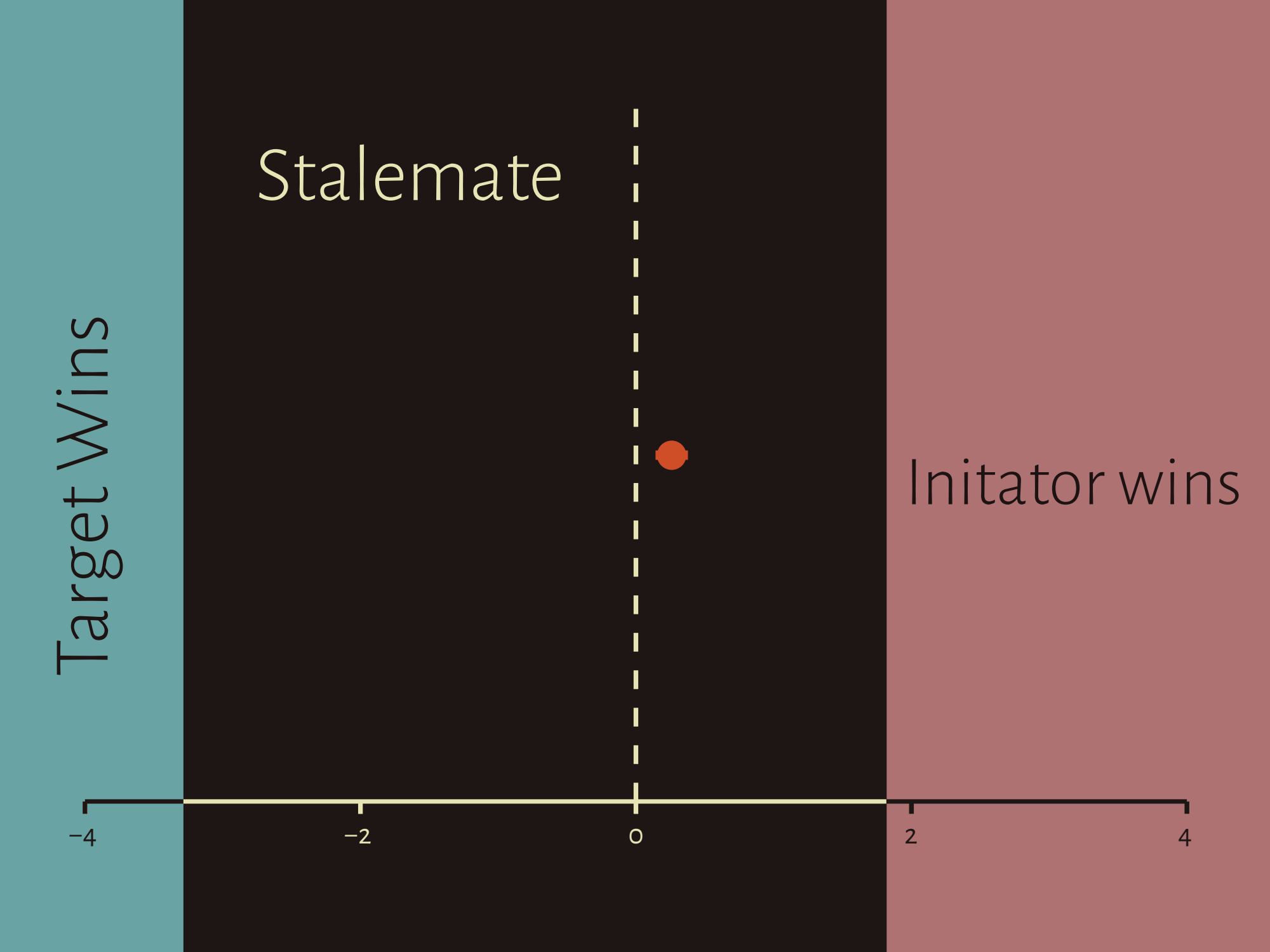


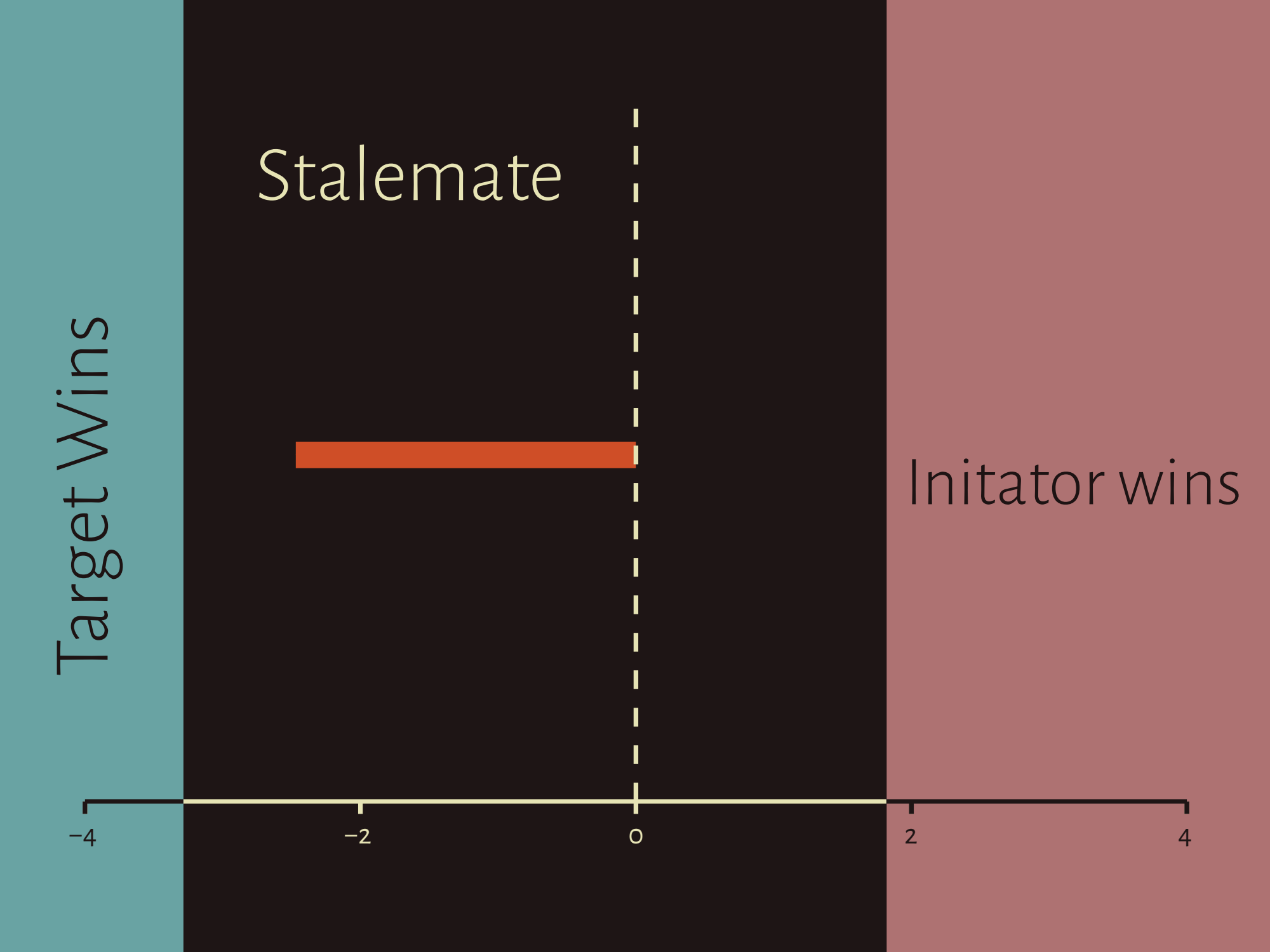
but the estimate is **tiny**
compared to the cutpoints

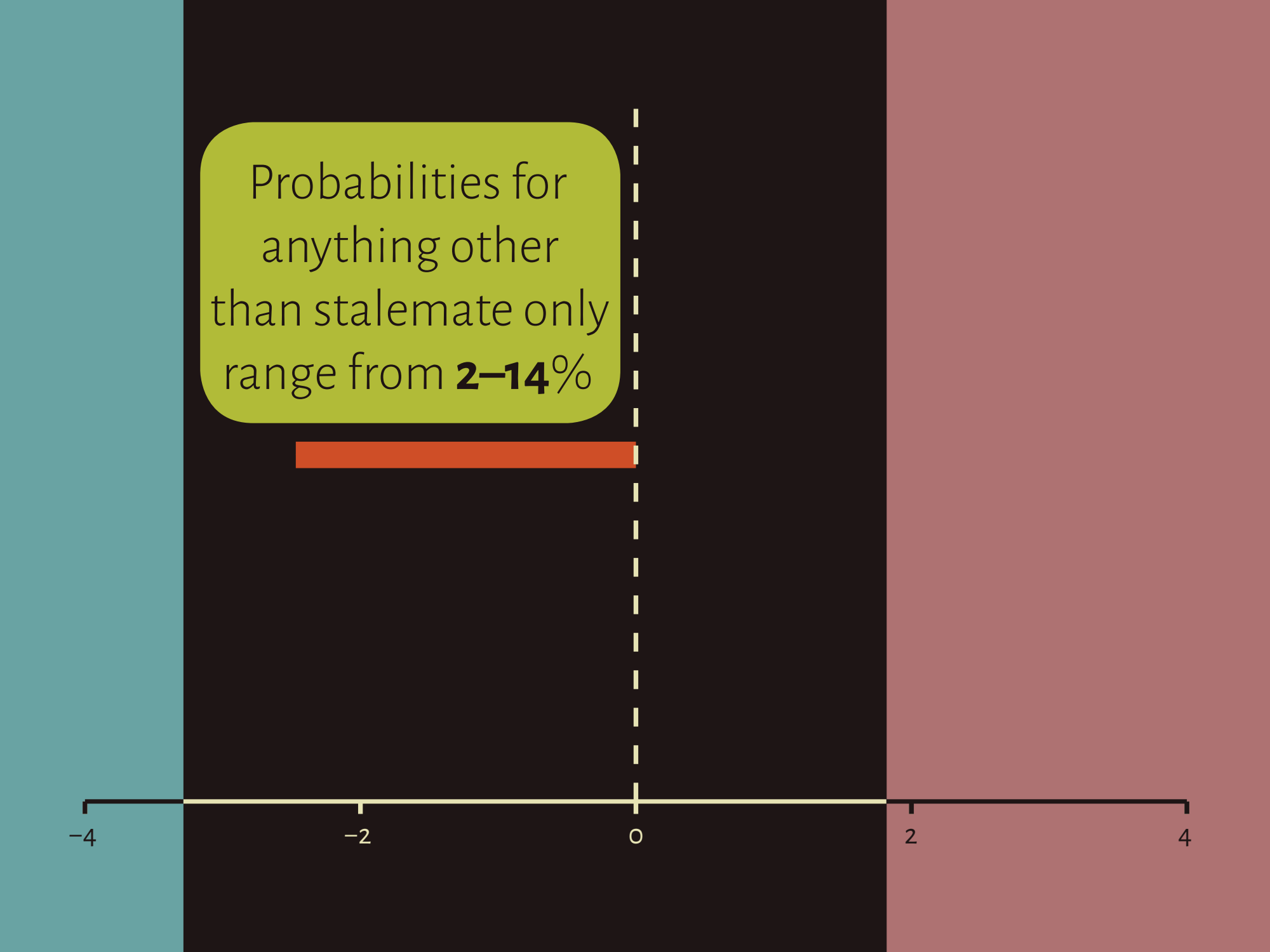

but the estimate is **tiny**
compared to the cutpoints
−4
−2
0
2
4

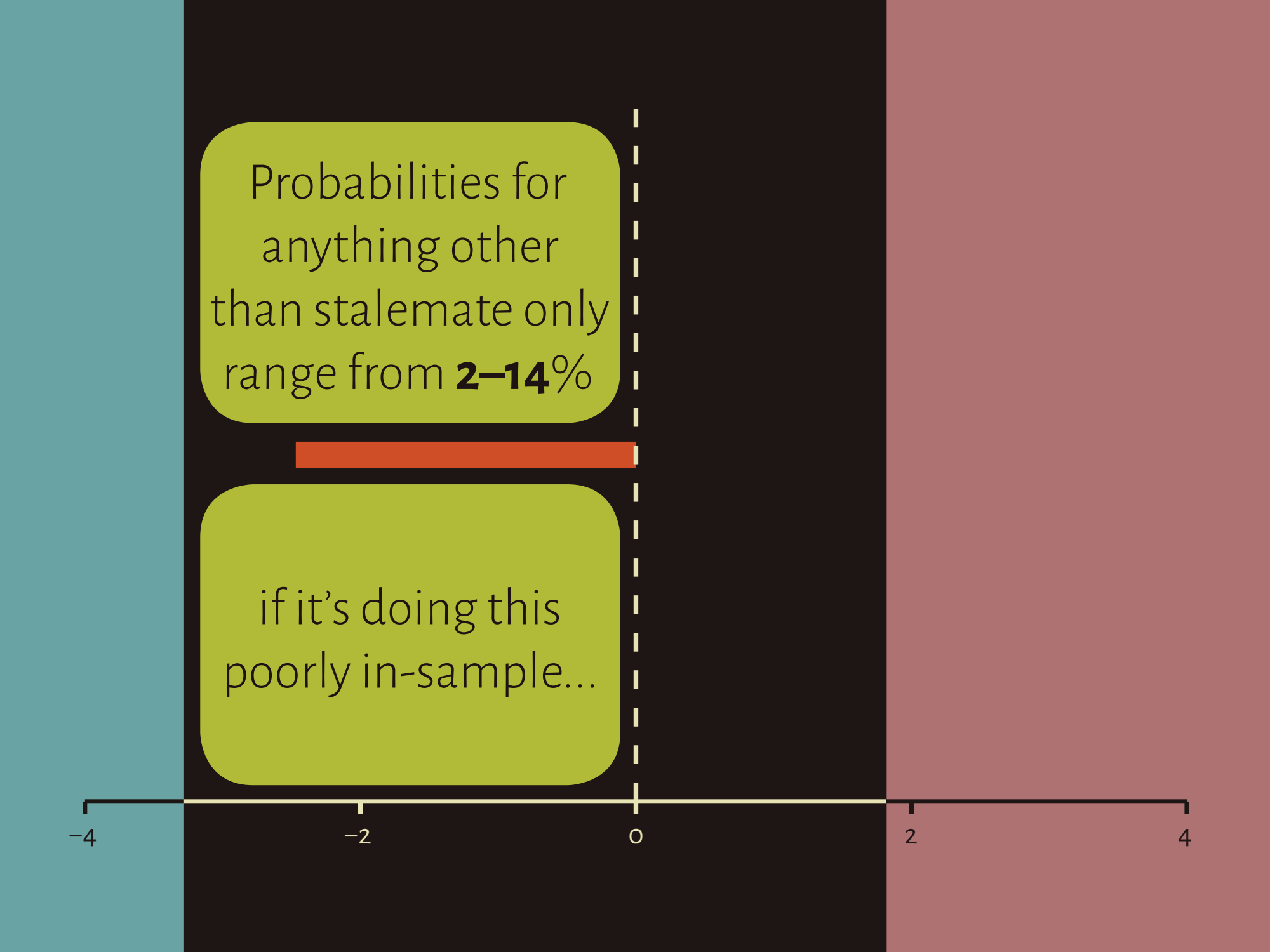

−4
−2
0
2
4


−4
−2
0
2
4


Stalemate
Target Wins
Initator wins
−4
−2
0
2
4


Stalemate
Target Wins
Initator wins
−4
−2
0
2
4


Probabilities for anything other than stalemate only range from **2–14**%
−4
−2
0
2
4


Probabilities for anything other than stalemate only range from **2–14**%
if it's doing this poorly in-sample...
−4
−2
0
2
4

Punchline One

capability ratios predict nothing

how we built a better proxy

2

how we built
a better proxy
without collecting
any new data

2

step one

step one

LISTEN TO THE DATA

step one

LISTEN TO THE DATA

...but not **greedily**

tenets

give the data as many looks as possible

- components
- proportions
- allow importance to vary over time

take out-of-sample performance seriously

- cross-validation to measure prediction

blind with science

- use a lot of flexible algorithms
- find the best weighted average of them

a little more meat on that bone

a little more meat on that bone

we use 7 estimators...

ordered logit

random forests

CART

C5.0

support vector machine

k-nearest neighbor

averaged neural networks

a little more meat on that bone

we use 7 estimators...

ordered logit
random forests
CART
C5.0
support vector machine
k-nearest neighbor
averaged neural networks

...on 4 specifications...

raw components
raw components + year
annual proportions
annual proportions + year

a little more meat on that bone

we use 7 estimators...

ordered logit
random forests
CART
C5.0
support vector machine
k-nearest neighbor
averaged neural networks

...on 4 specifications...

raw components
raw components + year
annual proportions
annual proportions + year

...plus miscellany.

null model
ordered logit on C.R.

a little more meat on that bone

we use 7 estimators...

ordered logit
random forests
CART
C5.0
support vector machine
k-nearest neighbor
averaged neural networks

...on 4 specifications...

raw components
raw components + year
annual proportions
annual proportions + year

...plus miscellany.

null model
ordered logit on C.R.

that's 30 total models

measuring predictive power

cross validation:

1. split
2. fit
3. repeat

measuring predictive power

cross validation:

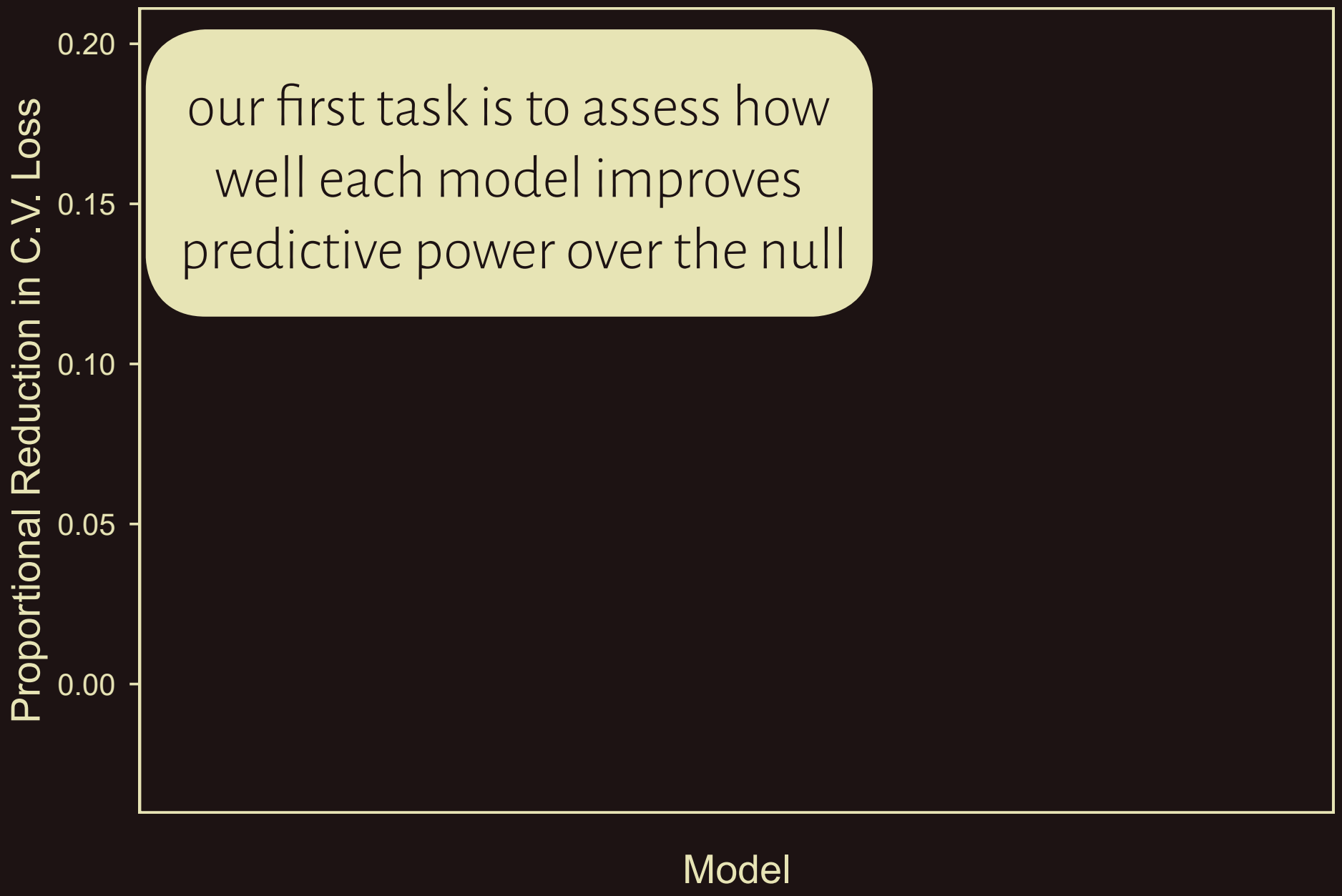
1. split
2. fit
3. repeat

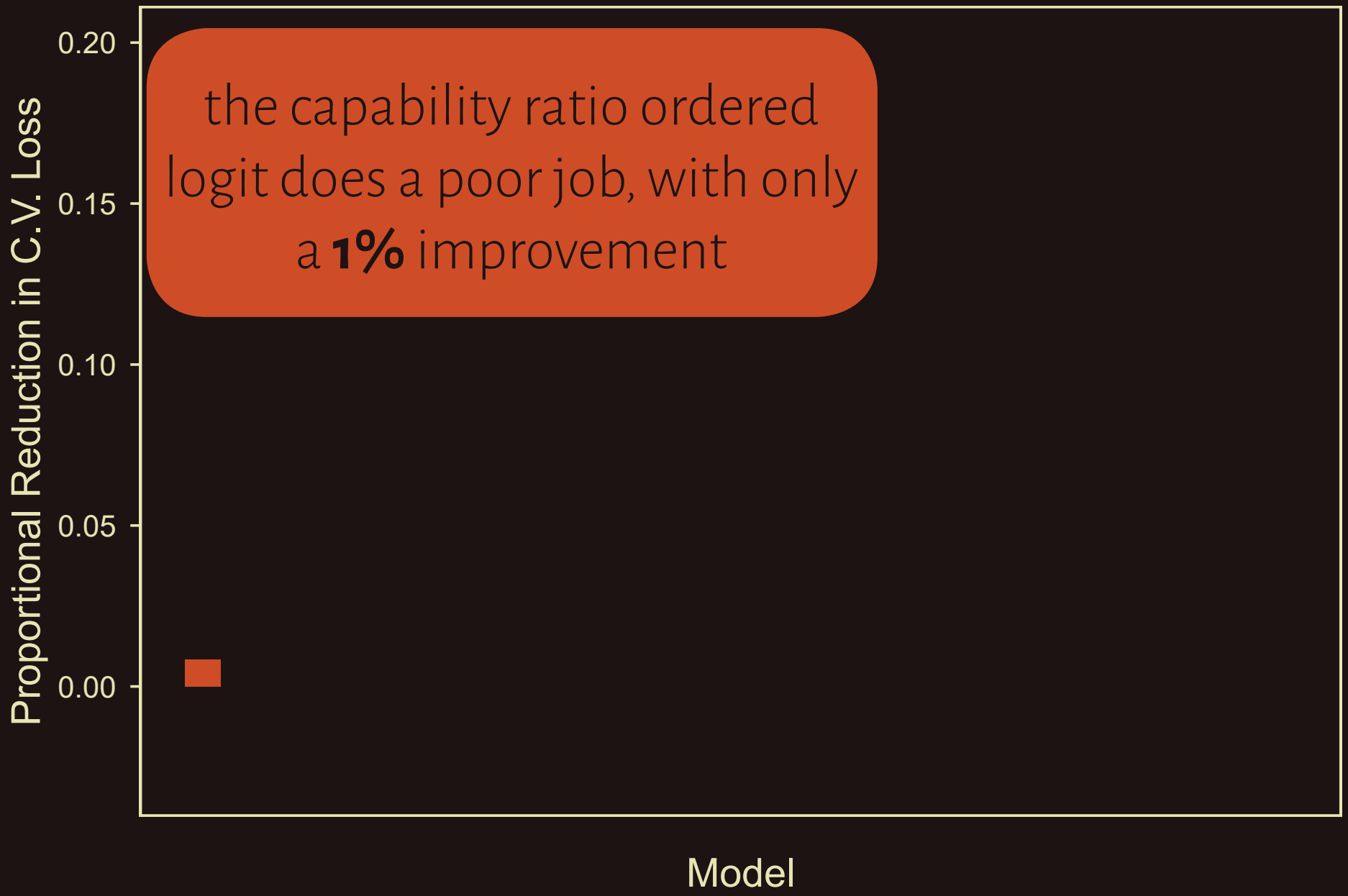
used for:

1. evaluating individual models
2. constructing best weighted average of them

our first task is to assess how well each model improves predictive power over the null

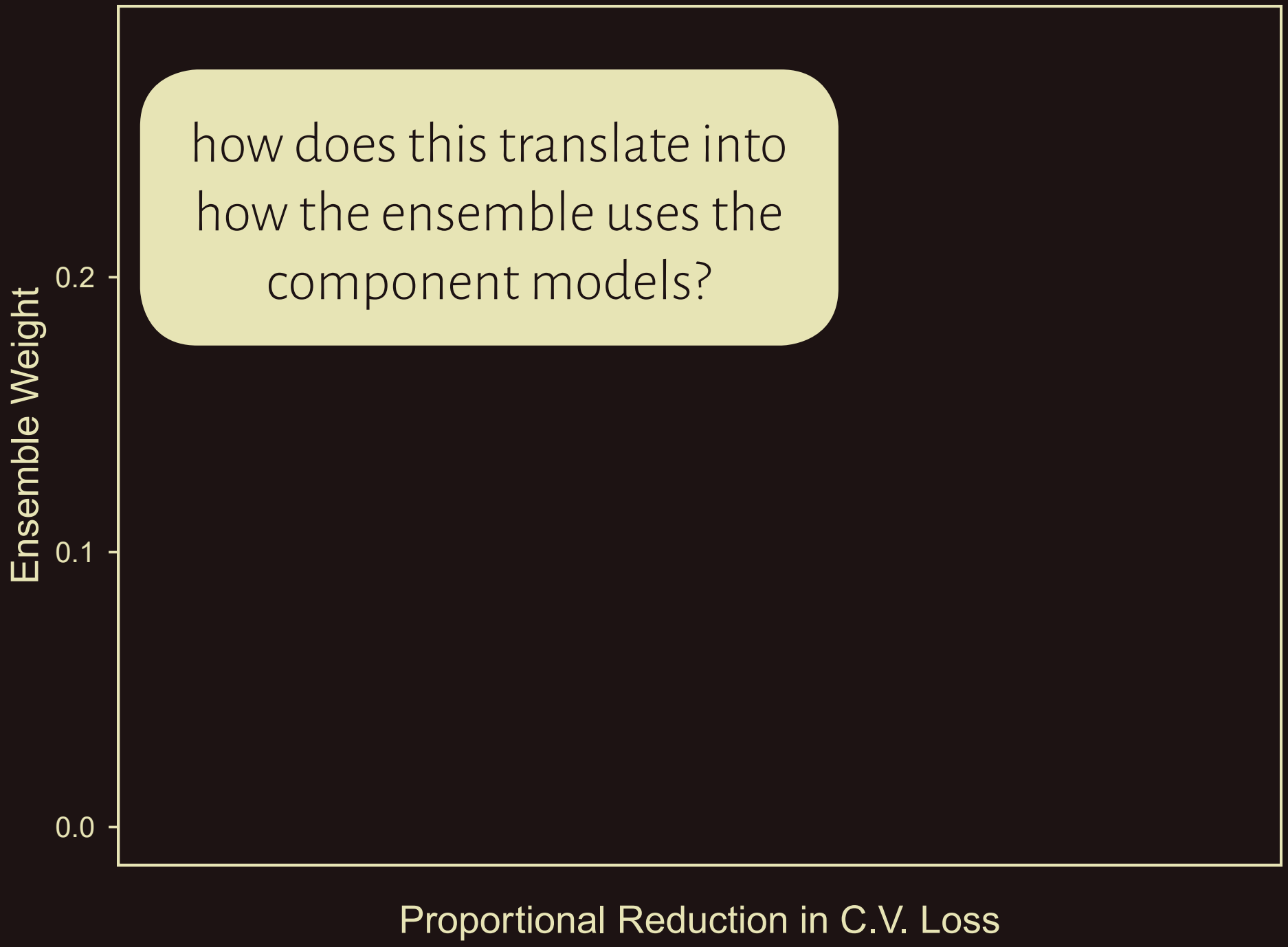
our first task is to assess how well each model improves predictive power over the null

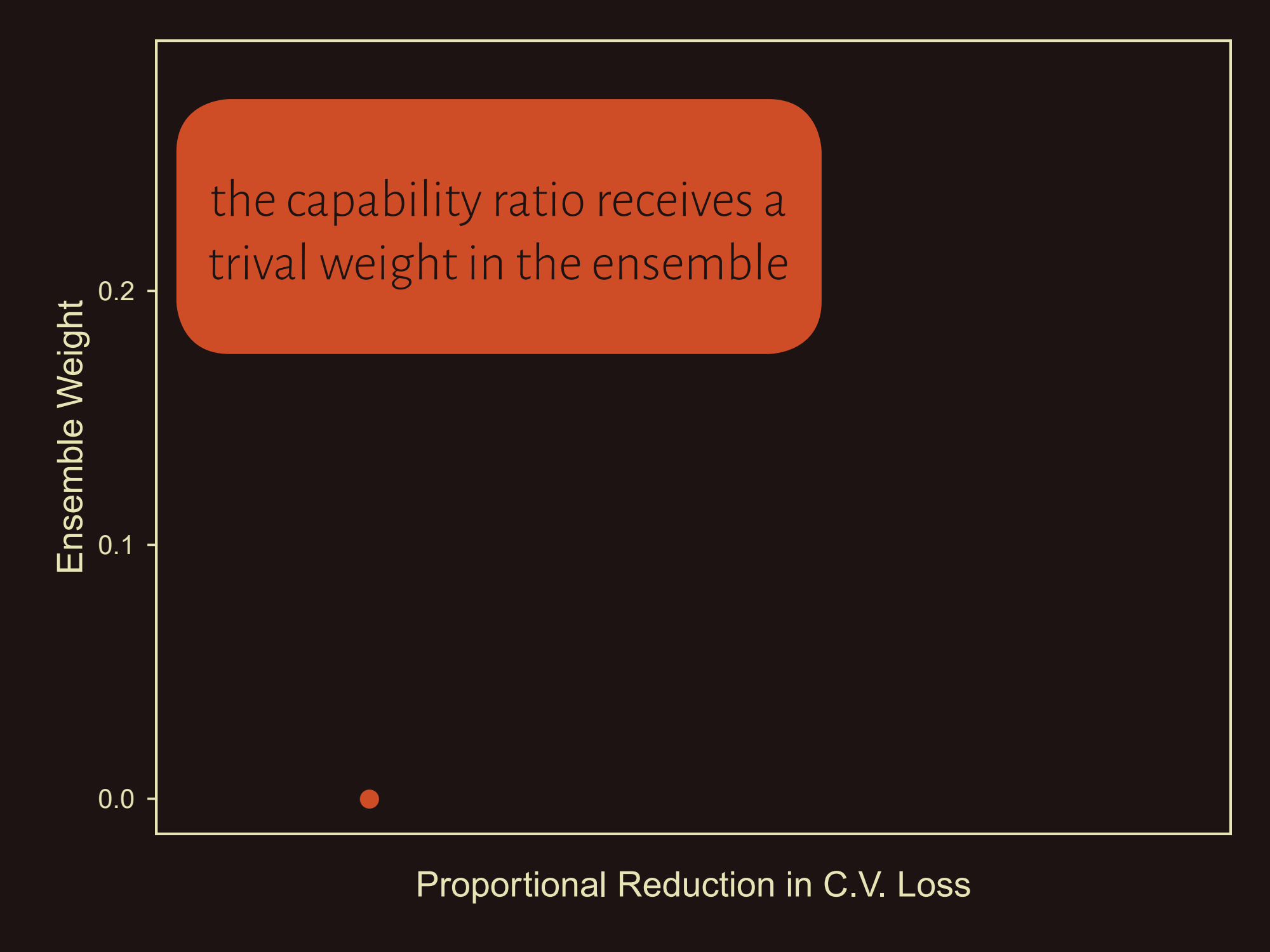


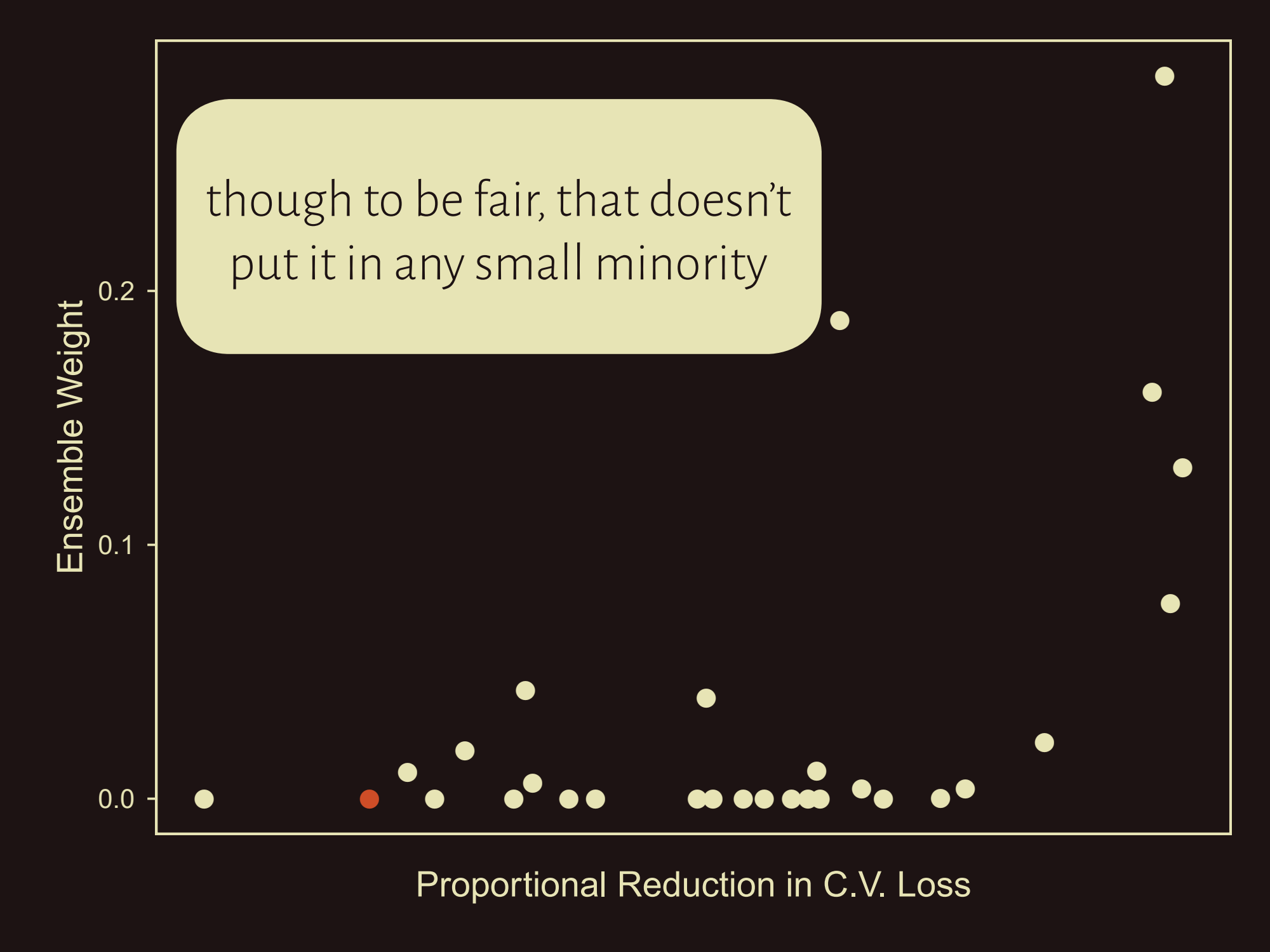

the capability ratio ordered logit does a poor job, with only a **1%** improvement
Proportional Reduction in C.V. Loss
0.20
0.15
0.10
0.05
0.00
Model


all but one of the other models do better, with most doing far better
Proportional Reduction in C.V. Loss
0.20
0.15
0.10
0.05
0.00
Model

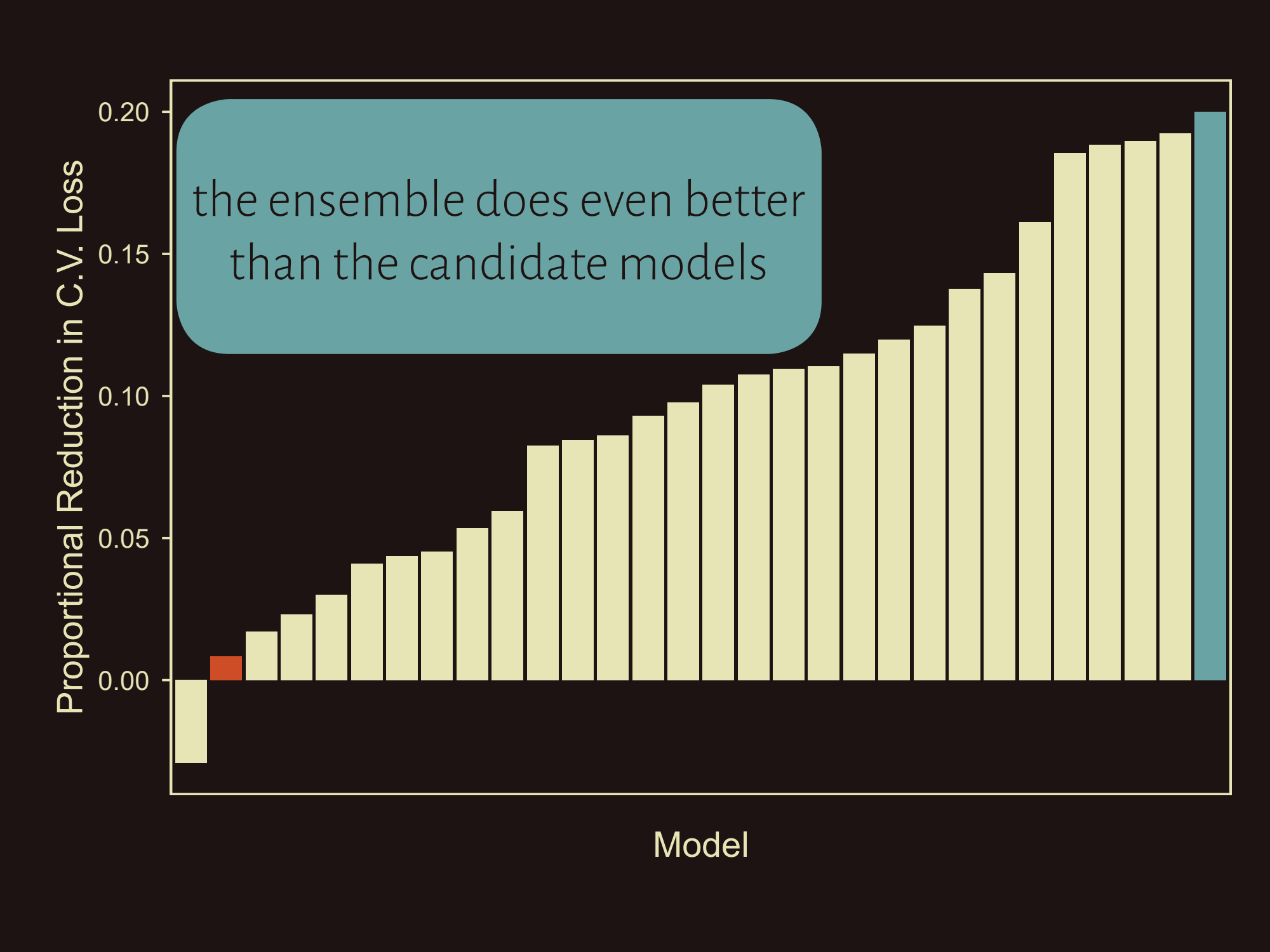
how does this translate into how the ensemble uses the component models?

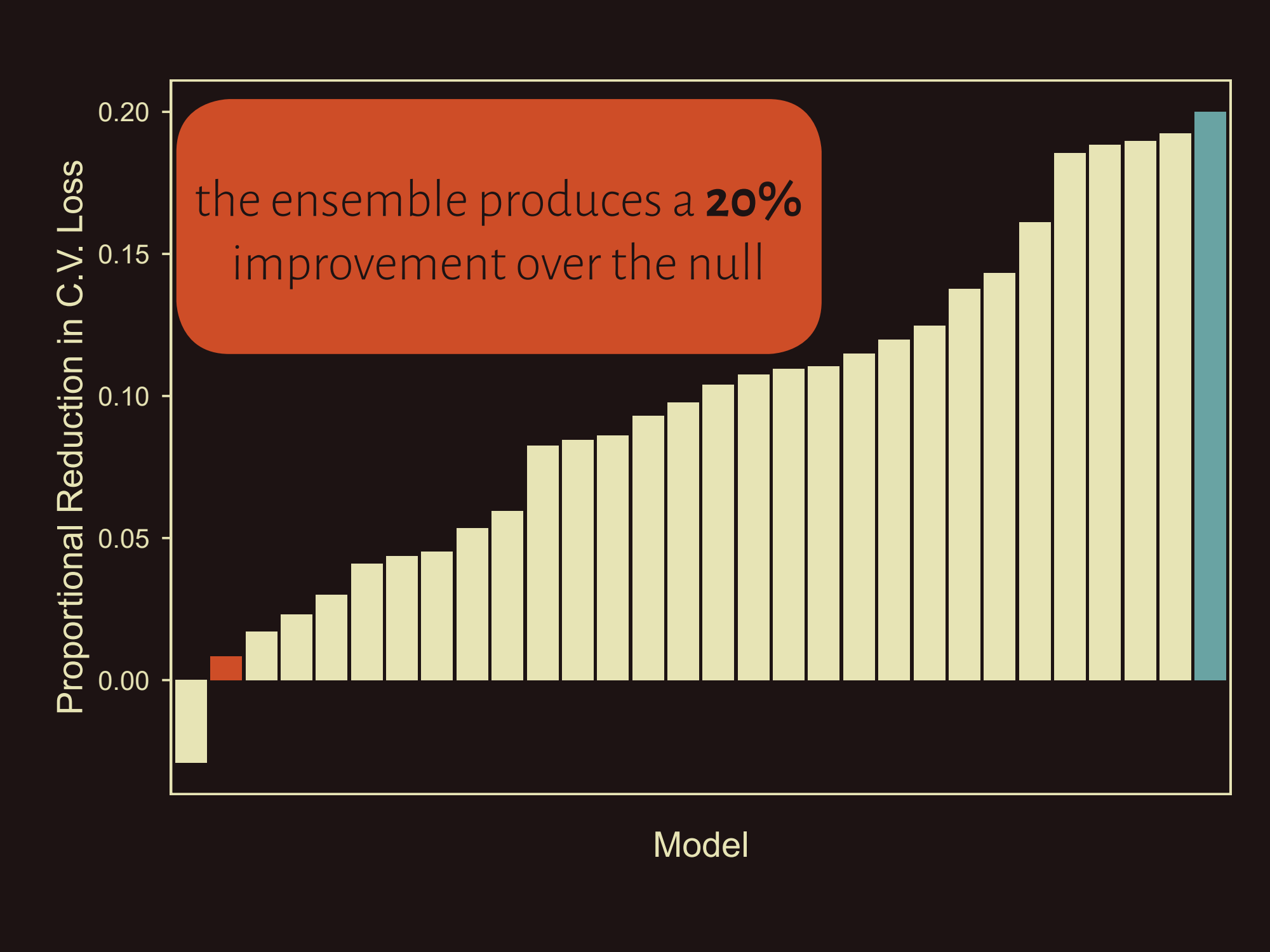

how does this translate into how the ensemble uses the component models?
Ensemble Weight
0.2
0.1
0.0
Proportional Reduction in C.V. Loss


the capability ratio receives a trival weight in the ensemble
Ensemble Weight
0.2
0.1
0.0
Proportional Reduction in C.V. Loss


though to be fair, that doesn't put it in any small minority
Ensemble Weight
0.0
0.1
0.2
Proportional Reduction in C.V. Loss

so what does this all mean for the ensemble's predictive performance?


the ensemble does even better than the candidate models
Proportional Reduction in C.V. Loss
0.20
0.15
0.10
0.05
0.00
Model


the ensemble produces a **20%** improvement over the null
Proportional Reduction in C.V. Loss
0.20
0.15
0.10
0.05
0.00
Model

Punchline One

capability ratios predict nothing

Punchline Two

our method predicts far better

introducing the DOE score

we use our results to generated predicted probabilities for all dyads in all years

we refer to these as Dispute Outcome Expectations, or **DOE scores**

almost any time you could use a capability ratio, so too could you use a DOE score

what our proxy did on its summer vacation

3

it replicated a bunch of results

it replicated a bunch of results

the DOE score predicts outcomes better than the capability ratio

it replicated a bunch of results

the DOE score predicts outcomes better than the capability ratio

but usually we just want a good control for expectations or relative power

it replicated a bunch of results

the DOE score predicts outcomes better than the capability ratio

but usually we just want a good control for expectations or relative power

does DOE improve our regressions?

the DOE score predicts outcomes better than the capability ratio

but usually we just want a good control for expectations or relative power

does DOE improve our regressions?

Colaresi and Thompson, 2005; Grieg, 2005; Senese, 2005; Sobek, 2005; Werner and Yuen, 2005; Allee and Huth, 2006; Bearce, Flanagan and Floros, 2006; Bennett, 2006; James, Park and Choi, 2006; Oneal and Tir, 2006; Pevehouse and Russett, 2006; Rasler and Thompson, 2006; Sobek, Abouharb and Ingram, 2006; Bearce and Bondanella, 2006; Chapman and Roeder, 2007; Clare, 2007; Crescenzi, 2007; Gartzke, 2007; Gent, 2007; Gibler, 2007; Lektzian and Sprecher, 2007; Long, Nordstrom and Baek, 2007; Maoz et al., 2007; Mitchell and Hensel, 2007; Morrow, 2007; Aydin, 2008; Beardsley, 2008; Braumoeller, 2008; Clark, Nordstrom and Reed, 2008; Fo lehyan, 2008*a*; Salehyan, 200 Kibben, 2009; Kroenig, 2009 nd Peinhardt, 2010; Dreyer 2 ins and Souva, 2010; Mattes a annon, Morey and Boehmke Hensel, 2011; Choi, 2011; D 2011; Zawahri and Mitchell, 2012; Downes and Sechser, 2 12; Kleinberg, Robinson and French, 2012; Mattes, 2012; Oswiak, 2012; Sobek, Foster and Robison 2012; Uzonyi, Souva and Golder, 2012; Weeks, 2012; Clare, 2013; Gartzke and Weisiger, 2013; Gibler and Hutchison, 2013; Huth, Croco and Appel, 2013; Kroenig, 2013; Mousseau, 2013; Oswiak and Rider, 2013; Rider, 2013; Sechser and Fuhrmann, 2013; Whang, McLean and Kuberski, 2013; Williams, 2013; Carnegie, 2014; Dafoe, Oneal and Russett, 2014; Fang, Johnson and Leeds 2014; Fuhrmann and Sechser, 2014; Gartzke and Weisiger, 2014; Gibler, 2014; Gibler and Tir, 2014; Jung, 2014; Park and Colaresi, 2014; Potter and Baum, 2014; Wolford, 2014

the DOE score predicts outcomes better than the capability ratio

but usually we just want a good control for expectations or relative power

does DOE improve our regressions?

Colaresi and Thompson, 2005; Grieg, 2005; Senese, 2005; Sobek, 2005; Werner and Yuen, 2005; Allee and Huth, 2006; Bearce, Flanagan and Floros, 2006; Bennett, 2006; James, Park and Choi, 2006; Oneal and Tir, 2006; Pevehouse and Russett, 2006; Rasler and Thompson, 2006; Sobek, Abouharb and Ingram, 2006; Bearce and Bondanella, 2006; Chapman and Roeder, 2007; Clare, 2007; Crescenzi, 2007; Gartzke, 2007; Gent, 2007; Gibler, 2007; Lektzian and Sprecher, 2007; Long, Nordstrom and Baek, 2007; Maoz et al., 2007; Mitchell and Hensel, 2007; Morrow, 2007; Aydin, 2008; Beardsley, 2008; Braumoeller, 2008; Clark, Nordstrom and Reed, 2008; Fordham, 2008; Lo, Hashimoto and Reiter, 2008; Mattes, 2008; Salehyan, 2008*a*; Salehyan, 2008*b*; Weeks, 2008; Arena and Palmer, 2009; Bearce, Floros and McKibben, 2009; Kroenig, 2009; Langlois and Langlois, 2009; Maoz, 2009; Rider, 2009; Allee and Peinhardt, 2010; Dreyer 2010; Faussett and Volgy, 2010; Gibler and Tir, 2010; Lektzian, Prins and Souva, 2010; Mattes and Vonnahme, 2010; Reed and Chiba, 2010; Schultz, 2010; Shannon, Morey and Boehmke, 2010; Tir, 2010; Xiang, 2010; Benson, 2011; Brochmann and Hensel, 2011; Choi, 2011; Dafoe, 2011; Huth, Croco and Appel, 2011; Koga, 2011; Morey, 2011; Zawahri and Mitchell, 2011; Allen and Fordham, 2012; Colaresi, 2012; Crescenzi et al., 2012; Downes and Sechser, 2012; Findley, Piazza and Young, 2012; Huth, Croco and Appel, 2012; Kleinberg, Robinson and French, 2012; Mattes, 2012; Owsiak, 2012; Sobek, Foster and Robison 2012; Uzonyi, Souva and Golder, 2012; Weeks, 2012; Clare, 2013; Gartzke and Weisiger, 2013; Gibler and Hutchison, 2013; Huth, Croco and Appel, 2013; Kroenig, 2013; Mousseau, 2013; Oswiak and Rider, 2013; Rider, 2013; Sechser and Fuhrmann, 2013; Whang, McLean and Kuberski, 2013; Williams, 2013; Carnegie, 2014; Dafoe, Oneal and Russett, 2014; Fang, Johnson and Leeds 2014; Fuhrmann and Sechser, 2014; Gartzke and Weisiger, 2014; Gibler, 2014; Gibler and Tir, 2014; Jung, 2014; Park and Colaresi, 2014; Potter and Baum, 2014; Wolford, 2014

Bennett, 2006

Sobek, Abouharb and Ingram, 2006

Gartzke, 2007

Morrow, 2007

Fordham, 2008

Salehyan, 2008*a*; Salehyan, 2008*b*; Weeks, 2008; Arena and Palmer, 2009

Dreyer 2010

Zawahri and Mitchell, 2011

Huth, Croco and Appel, 2012

Owsiak, 2012

Uzonyi, Souva and Golder, 2012; Weeks, 2012

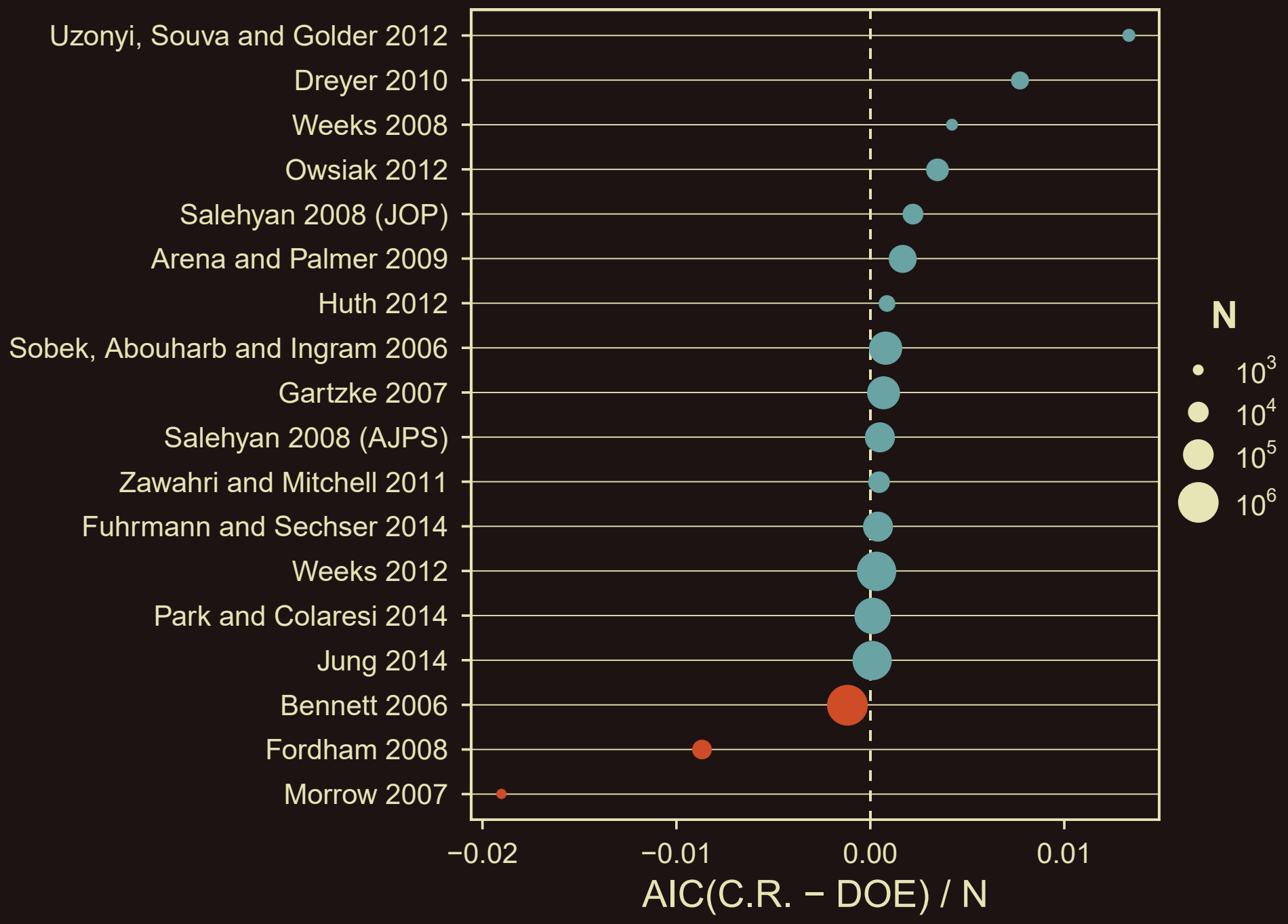
Fuhrmann and Sechser, 2014

Jung, 2014; Park and Colaresi, 2014

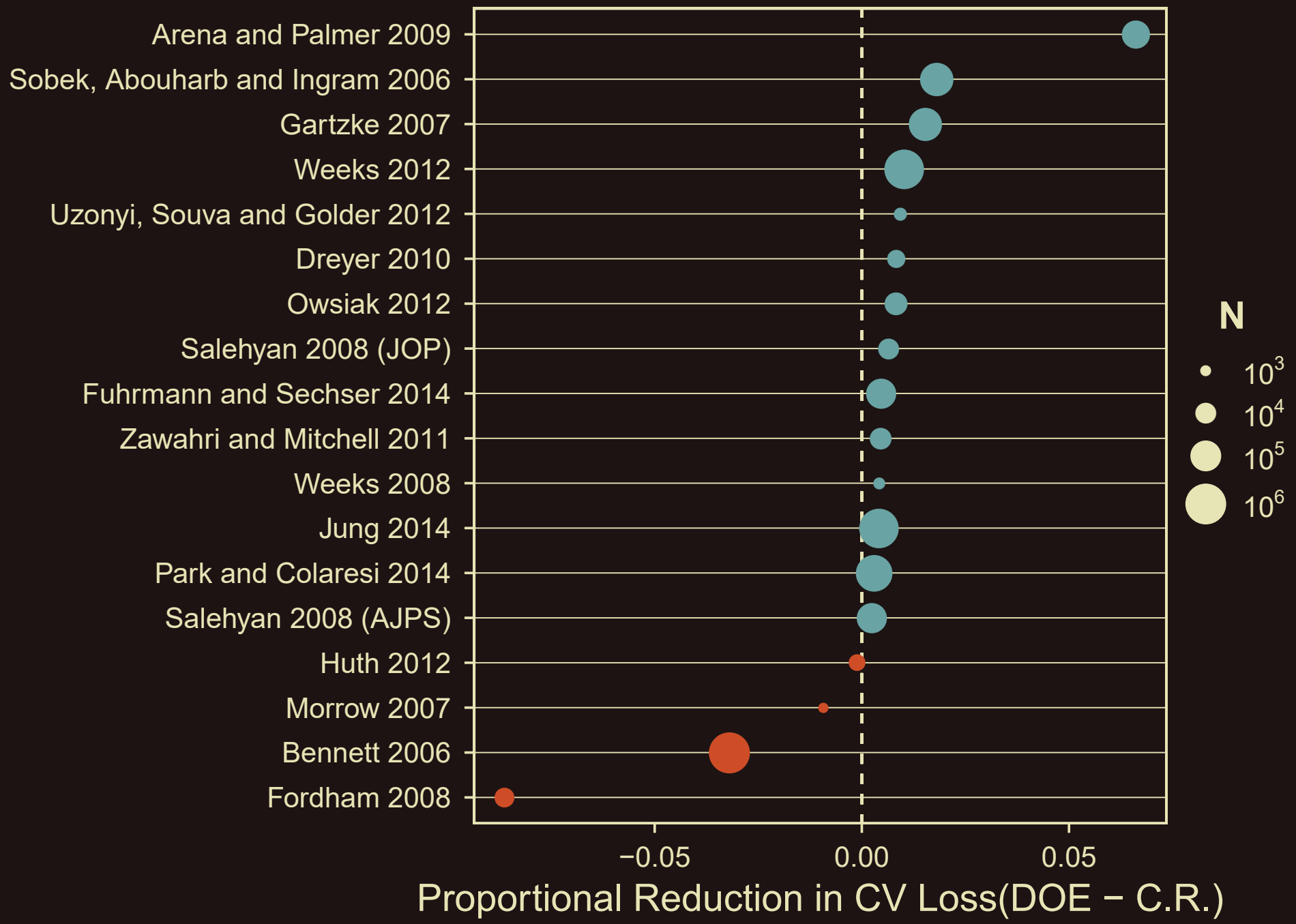
Bennett, 2006
Sobek, Abouharb and Ingram, 2006
Gartzke, 2007
Morrow, 2007
Fordham, 2008
Salehyan, 2008*a*
Salehyan, 2008*b*
Weeks, 2008
Arena and Palmer, 2009
Dreyer 2010
Zawahri and Mitchell, 2011
Huth, Croco and Appel, 2012
Owsiak, 2012
Uzonyi, Souva and Golder, 2012
Weeks, 2012
Fuhrmann and Sechser, 2014
Jung, 2014
Park and Colaresi, 2014

Bennett, 2006
Sobek, Abouharb and Ingram, 2006
Gartzke, 2007
Morrow, 2007
Fordham, 2008
Salehyan, 2008*a*
Salehyan, 2008*b*
Weeks, 2008
Arena and Palmer, 2009
Dreyer 2010
Zawahri and Mitchell, 2011
Huth, Croco and Appel, 2012
Owsiak, 2012
Uzonyi, Souva and Golder, 2012
Weeks, 2012
Fuhrmann and Sechser, 2014
Jung, 2014
Park and Colaresi, 2014

we wonder whether the DOE scores improve in-sample fit and out-of-sample prediction


Uzonyi, Souva and Golder 2012
Dreyer 2010
Weeks 2008
Owsiak 2012
Salehyan 2008 (JOP)
Arena and Palmer 2009
Huth 2012
Sobek, Abouharb and Ingram 2006
Gartzke 2007
Salehyan 2008 (AJPS)
Zawahri and Mitchell 2011
Fuhrmann and Sechser 2014
Weeks 2012
Park and Colaresi 2014
Jung 2014
Bennett 2006
Fordham 2008
Morrow 2007
−0.02
−0.01
0.00
0.01
AIC(C.R. − DOE) / N
N
10^3
10^4
10^5
10^6


Weeks 2012
Sobek, Abouharb and Ingram 2006
Gartzke 2007
Arena and Palmer 2009
Park and Colaresi 2014
Dreyer 2010
Jung 2014
Owsiak 2012
Uzonyi, Souva and Golder 2012
Salehyan 2008 (JOP)
Fuhrmann and Sechser 2014
Zawahri and Mitchell 2011
Salehyan 2008 (AJPS)
Weeks 2008
Huth 2012
Fordham 2008
Morrow 2007
Bennett 2006
−10
−5
0
5
Vuong Statistic
N
10^3
10^4
10^5
10^6


Arena and Palmer 2009
Sobek, Abouharb and Ingram 2006
Gartzke 2007
Weeks 2012
Uzonyi, Souva and Golder 2012
Dreyer 2010
Owsiak 2012
Salehyan 2008 (JOP)
Fuhrmann and Sechser 2014
Zawahri and Mitchell 2011
Weeks 2008
Jung 2014
Park and Colaresi 2014
Salehyan 2008 (AJPS)
Huth 2012
Morrow 2007
Bennett 2006
Fordham 2008
−0.05
0.00
0.05
Proportional Reduction in CV Loss(DOE − C.R.)
N
10^3
10^4
10^5
10^6

Punchline One

capability ratios predict nothing

Punchline Two

our method predicts far better

Punchline Three

use DOE scores!

4

wrap-up and conclusions

summary

we argue that proxies ought to **predict well**

we find that the industry standard predicts terribly

we develop a new way to construct a better proxy, which we find predicts far better

the proxy we construct generally performs better in the models peace scientists care most about

Punchline One

capability ratios predict nothing

Punchline Two

our method predicts far better

Punchline Three

use DOE scores!

most importantly


THE PRESIDENT of the UNITED STATES


thanks so much
for listening!
THE PRESIDENT of the UNITED STATES

thanks so much for listening!

rjcarroll@fsu.edu brenton.kenkel@vanderbilt.edu
http://www.doe-scores.com